

Patterns of Functional Damage in Neural Network Models of Associative Memory

Eytan Ruppin and James A. Reggia *
Departments of Computer Science and Neurology
A.V. Williams Bldg.
University of Maryland
College Park, MD 20742
ruppin@cs.umd.edu reggia@cs.umd.edu

December 8, 1994

Abstract

Current understanding of the effects of damage on neural networks is rudimentary, even though such understanding could lead to important insights concerning neurological and psychiatric disorders. Motivated by this consideration, we present a simple analytical framework for estimating the functional damage resulting from focal structural lesions to a neural network model. The effects of focal lesions of varying area, shape and number on the retrieval capacities of a spatially-organized associative memory are quantified, leading to specific scaling laws that may be further examined experimentally. It is predicted that multiple focal lesions will impair performance more than a single lesion of the same size, that slit like lesions are more damaging than rounder lesions, and that the same fraction of damage (relative to the total network size) will result in significantly less performance decrease in larger networks. Our study is clinically motivated by the observation that in multi-infarct dementia, the size of metabolically impaired tissue correlates with the level of cognitive impairment more than the size of structural damage. Our results account for the detrimental effect of the number of infarcts rather than their overall size of structural damage, and for the ‘multiplicative’ interaction between Alzheimer’s disease and multi-infarct dementia.

*This research has been supported in part by a Rothschild Fellowship to Dr. Ruppin and in part by Awards NS29414 and NS16332 from NINDS. Dr. Reggia is also with the Institute of Advanced Computer Studies at the University of Maryland. Correspondence should be addressed to Dr. Reggia.

1 Introduction

Understanding the response of neural nets to structural/functional damage is important for assessing the performance of neural network hardware, and in gaining understanding of the mechanisms underlying neurological and psychiatric disorders. Recently, there has been a growing interest in constructing neural models to study how specific pathological neuroanatomical and neurophysiological changes can result in various clinical manifestations, and to investigate the functional organization of the symptoms that result from specific brain pathologies (reviewed in [Reggia *et al.*, 1994; Ruppin, 1995]). In the area of associative memory models specifically, early computational studies found an increase in memory impairment with increasing lesion severity [Wood, 1978] (in accordance with Lashley’s classical ‘mass action’ principle), and showed that slowly developing lesions can have less pronounced effects than equivalent acute lesions [Anderson, 1983]. More recently, it was shown that the gradual pattern of clinical deterioration manifested in the majority of Alzheimer’s patients can be explained, and that different synaptic compensation rates can account for the observed variation in the severity and progression rate of this disease [Horn *et al.*, 1993; Ruppin & Reggia, 1994].

Previous work, however, is limited in that model elements have no spatial relationships to one another (all elements are conceptually equidistant). Thus, as there is no way to represent focal (localized) damage in such networks, it has not been possible to study the functional effects of focal lesions on memory and to compare them with those caused by diffuse lesions. This paper presents the first computational study of the effect of focal lesions on memory performance with spatially-organized neural networks. It is motivated by the observation that in neural network models, a focal *structural lesion* (that is, the permanent and complete inactivation of some group of adjacent elements) is accompanied by a surrounding *functional lesion* composed of structurally intact but functionally impaired elements. This region of functional impairment occurs due to the loss of innervation from the structurally damaged region. It is the combined effect of both regions that determines the actual extent of performance decrease in the network. From a modeling perspective, this paper presents a simple but general approach to analyzing the functional effects of focal lesions. This approach is used to derive scaling laws that quantify the effects of spatial characteristics of focal lesions such as their number and shape on the performance of network models of associative memory.

Beyond its computational interest, the study of the effects of focal damage on the performance of neural network models can lead to a better understanding of functional impairments accompanying focal brain lesions. In particular, we are interested in *multi-infarct dementia*, a frequent cause of dementia (chronic deterioration of cognitive and memory capacities) characterized by a series of multiple, aggregating focal lesions. The distinction made in the model network considered here between structural and functional lesions has a clinical parallel: ‘structural’ lesions represent regions of infarcted (dead) tissue, as measured by structural imaging methods such as computerized tomography, and ‘functional’ lesions represent regions of metabolically impaired tissue surrounding the infarcted tissue, as measured by functional imaging techniques such as positron emission tomography. Interestingly, in multi-infarct dementia the correlation between the volume of the primary infarct region and the severity of the resulting cognitive deficit is unclear and controversial [Meyer *et al.*, 1988; del Ser *et al.*, 1990; Liu *et al.*, 1990; Tatemichi *et al.*, 1990; Gorelick *et al.*, 1992]. In contrast, there is a strong relationship between the total volume of metabolically impaired tissue measured in the chronic phase and the severity of multi-infarct dementia [Mieke *et al.*, 1992; Heiss *et al.*, 1993b; Heiss *et al.*, 1993a]. This highlights the importance of studying functional impairment after focal lesions.

The reader familiar with the clinical stroke literature should note that the functional lesions modeled in this paper are *not* the ‘penumbra’ perilesion areas of comprised blood supply and acute metabolic changes that surround focal infarcts during the acute post-infarct period. Rather, they are regions of reduced metabolic activity that are observed in chronic multi-infarct dementia patients months after the last infarct episode. The reduced metabolic activity in these areas is probably a result of both residual post-infarct neuropathological damage and the loss of innervation from the primary infarct region [Mies *et al.*, 1983; Heiss *et al.*, 1993b]. Intuitively, it is clear that in large enough lesions the functional damage resulting from loss of innervation should scale proportionally to the lesion circumference. This entails, in turn, that the functional damage should depend on the spatial characteristics of the structural lesion, such as its shape and the number of spatially distinct sub-lesions composing it. This work is devoted to a formal and quantitative study of these dependencies, and to a discussion of their possible clinical implications.

In Section 2, we derive a theoretical framework that characterizes the effects of focal lesions on an associative network’s performance. This framework, which is formulated in very general

terms, is then examined via simulations with a specific associative memory network in Section 3. These simulations show a fair quantitative fit with the theoretical predictions, and are compared with simulations examining performance with diffuse damage. The effects of various parameters characterizing the network’s architecture on post-lesion performance are further investigated in Section 4. Finally, our results are discussed in Section 5, and are evaluated in light of some relevant clinical data.

2 Analytical Scaling Rules

The model network we study consists of a 2-dimensional array of units whose edges are connected, forming a torus to eliminate edge effects. Each unit is connected primarily to its nearby neighbors, as in the cortex [Thomson & Deuchars, 1994], where the probability of a connection existing between two units is a Gaussian density function of the distance between them in the array. The unit of distance here is the distance between two neighboring elements in the array.

Our analysis pertains to the case where, in the pre-damaged network, all units have similar average activation and performance levels ¹. A focal *structural lesion* (anatomical lesion), denoting an area of damage and neuronal death, is modeled by permanently clamping the activity of the lesioned units to zero at the onset of the lesion. As a result of this primary structural lesion, the activity of surrounding units may be decreased, resulting in a secondary *functional lesion*, as illustrated in Figure 1.

We are primarily interested in large focal lesions, where the area s of the lesion is significantly greater than the local neighborhood region from which each unit receives its inputs. Throughout our analysis we shall hold the working assumption that, traversing from the border of the lesion outwards, the activity of units, and with it, the network performance level, gradually rises from zero until it reaches its normal, predamaged levels, at some distance d from the lesion’s border (see Figure 1). We denote d as the *functional impairment span*. This assumption reflects the notion that units which are closer to the lesion border lose more viable inputs than units that are farther away from the lesion. Since s is large relative to each element’s connectivity

¹The analysis presented in this section is general in the sense that it does not rely on any specific connectivity or activation values. Note however that the above statement is true in general for associative memory networks, when the activity of each unit is averaged over a time span sufficiently long for the cueing and retrieval of a few stored patterns.

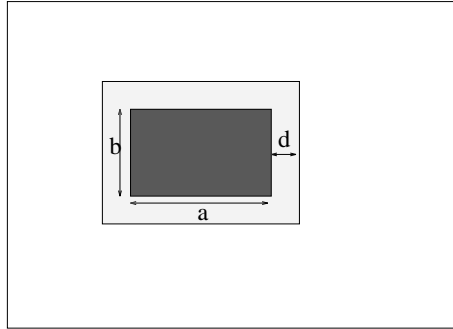


Figure 1: A sketch of a structural (dark shading) and surrounding functional (light shading) rectangular lesion. The a and b values denote the lengths of the rectangle's sides, and d is the functional impairment span.

neighborhood, d is determined primarily by the effect of the inactive regions at the periphery of the lesion. We may therefore assume that the value of d is independent of the lesion size, and depends specifically on the parameters defining the network's connectivity and dynamics. In Section 4 we will use computer simulations to verify that d is invariant over lesion size, and examine its dependence on the network parameters.

Let the intact baseline performance level of the network be denoted as $P(0)$, and let the network area be A . The network's performance, is quantified by some measure P ranging from 0 to 1. For example, if the network is an associative memory (as we study numerically in the next section), P denotes how accurately the network retrieves the correct memorized patterns given a set of input cues (defined formally in equation (17) of Appendix A). In the pre-damaged network all units have an approximately similar level of activity and performance. Then, a structural lesion of area s (dark shading in Figure 1), causing an additional functional lesion of area Δ_s (light shading in Figure 1), results in a performance level of approximately

$$P(s) = \frac{P(0)[A - (s + \Delta_s)] + P_\Delta \Delta_s}{A - s} = P(0) - (\Delta P \Delta_s)/(A - s), \quad (1)$$

where P_Δ denotes the average level of performance over Δ_s and $\Delta P = P(0) - P_\Delta$. $P(s)$ hence reflects the performance level over the remaining viable parts of the network, discarding the structurally damaged region ². Bearing these definitions in mind, the effect of focal lesions on the network's performance level can be characterized by the following rules.

²Alternatively, it is possible to measure the performance over the entire network. This would not affect our findings as long as the same measure is used in both the analysis and simulations, as the mapping between the two performance measures is order preserving.

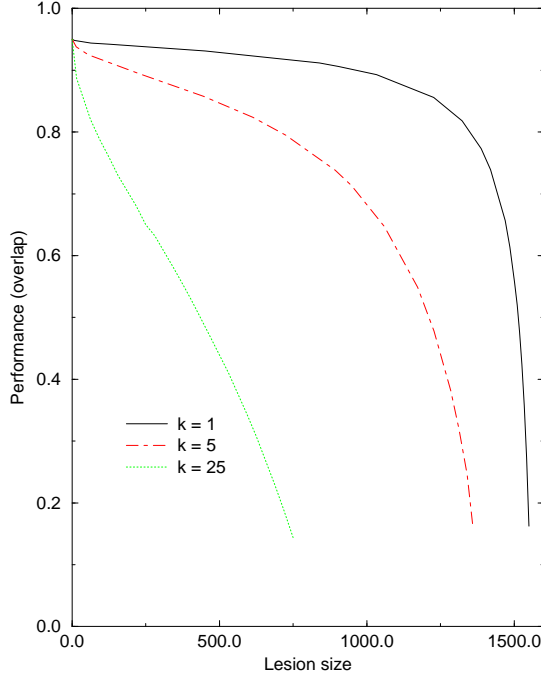


Figure 2: Theoretically predicted network performance as a function of a single focal structural lesion’s size (area): analytic curves obtained for different k values; $A = 1600$.

2.1 A Single Lesion

Consider a symmetric, circular structural lesion of size $s = \pi r^2$. The area of functional damage following such a lesion is $\Delta_s = \pi[(r + d)^2 - r^2] = \pi d^2 + \sqrt{4\pi d}\sqrt{s}$. In networks which operate well below the limit of their capacity and hence have significant functional reserves, the second term dominates since s is assumed to be large relative to d , and therefore

Rule 1:

$$\Delta_s \cong \sqrt{4\pi d}\sqrt{s} , \quad (2)$$

and (substituting the expression for Δ_s in (1))

$$P(s) \cong P(0) - \frac{k\sqrt{s}}{A - s} , \quad (3)$$

for some constant $k = \sqrt{4\pi d}\Delta P$. Thus, the area of functional damage surrounding a single focal structural lesion is proportional to the square root of the structural lesion’s area. Some analytic performance/lesioning curves (for various k values) are illustrated in Figure 2. Note the different qualitative shape of these curves as a function of k . As is evident, the shape of these curves reflects two conflicting tendencies; they are initially concave (in light of rule 1) and then

turn convex (as s increases and the remaining viable area is decreased). Letting $x = s/A$ be the *fraction of structural damage*, we have

Corollary 1:

$$P(x) \cong P(0) - \frac{k\sqrt{x}}{1-x} \frac{1}{\sqrt{A}}, \quad (4)$$

that is, the same *fraction* x of damage results in less performance decrease in larger networks! This surprising result testifies to the possible protective value of having functional ‘modular’ cortical networks of large size. Corollary 1 results from the fact that the functional damage does not scale up linearly with the structural lesion size, but only as the square root of the latter.

2.2 Varying Shape and Number

Expressions 3 and 4 are valid also when the structural lesion has a square shape. The resulting functional lesion of an s -size square structural lesion is $\Delta_s = 4d^2 + 4d\sqrt{s}$. To study the effect of the structural lesion’s shape, we consider the area $\Delta_{s[n]}$ of a functional lesion resulting from a rectangular focal lesion of size $s = a \cdot b$ (see Figure 1), where, without loss of generality, $n = a/b \geq 1$. Then, for large n , (i.e., elongated lesions), the area of functional damage is

$$\Delta_{s[n]} = 2d(a+b) + 4d^2 = (n+1)2d\sqrt{\frac{s}{n}} + 4d^2 \cong 2d\sqrt{n}\sqrt{s} + 4d^2 \leq \sqrt{n/4}\Delta_s, \quad (5)$$

where in the last step we neglect the contribution of the size-invariant term $4d^2$. The functional damage of a rectangular structural lesion of fixed size increases as its shape is more elongated. More quantitatively we have

Rule 2:

$$\Delta_{s[n]} \cong \sqrt{n/4}\Delta_s, \quad (6)$$

and

$$P(s) \cong P(0) - \frac{k\sqrt{ns}}{2(A-s)}. \quad (7)$$

Next, to study the effect of the number of lesions, consider the area Δ_s^m of a functional lesion composed of m focal rectangular structural lesions (with sides $a = n \cdot b$), each of area s/m . Using expression (5), we have

$$\Delta_s^m = m \left[2d(2d + \sqrt{n}\sqrt{s/m}) \right] = \sqrt{m} [2d(2d\sqrt{m} + \sqrt{n}\sqrt{s})] \geq \sqrt{m}\Delta_{s[n]}. \quad (8)$$

The functional damage hence increases as the number of the focal lesions m increases (total structural lesion area held constant), in accordance with

Rule 3:

$$\Delta_s^m \geq \sqrt{m}\Delta_{s[n]} , \quad (9)$$

which is always valid, irrespective of the value of d , and

$$P(s) \cong P(0) - \frac{k\sqrt{mns}}{2(A-s)} . \quad (10)$$

At first glance, the second and third rules seem to indicate that the functional damage caused by varying the shape or by varying the number of focal lesions behaves according to scaling laws of similar order. However, it should be noted that while rule 3 presents a lower bound on the functional damage which may actually be significantly larger, and involves no approximations, rule 2 presents an upper bound on the actual functional damage. As we shall show in the next section, the number of lesions actually affects the network performance significantly more than its precise shape (maintaining the total structural area fixed).

Let $\Delta[x]$ denote the functional damage caused by a single focal square lesion of area x (so $\Delta[s]$ is Δ_s). Since $\sqrt{l}\Delta_s \cong \Delta[s \cdot l]$ (by rule 1), then following rules 2 and 3 we obtain the following corollaries:

Corollary 2:

$$\Delta_{s[n]} \cong \Delta[n/4 \cdot s] \quad (11)$$

That is, the functional damage area following a rectangular structural lesion of area s and sides-ratio n is approximately equal to the functional damage area following a larger single square structural lesion of area $n/4 \cdot s$ (for large n).

Corollary 3:

$$\Delta_s^m \geq \Delta[m \cdot s] \quad (12)$$

In other words, the functional damage following multiple lesions composed of m rectangular focal structural lesions having total area s is greater than the functional damage following a single square lesion of area $m \cdot s$.

As is evident, the analysis presented in this section is based on several simplifying approximations. As such, it cannot be expected to yield an exact match with numerical results from computer simulations. However, as demonstrated in the next section, the scaling rules developed have the same shape as the numerical data, matching quite well at times, testifying to their validity.

3 Numerical Results

We now turn to examine the effect of lesions on the performance of an associative memory network via simulations. The goal of these simulations is twofold. First, to examine how accurately the general but approximate theoretical results presented above describe the actual performance degradation in a specific associative network. Second, to compare the effects of focal lesions to those of diffuse ones, as the effect of diffuse damage cannot be described as a limiting case within the framework of our analysis. Our simulations were performed using a standard Tsodyks-Feigelman attractor neural network [Tsodyks & Feigel'man, 1988]. This is a Hopfield-like network which has several features that make it more biologically plausible [Horn *et al.*, 1993], such as low activity and non-zero thresholds. Spatially-organized attractor networks can function reasonably well as associative memory devices [Karlholm, 1993], and a biologically-inspired realization of attractor networks using cortical columns as its elements has also been proposed [Lansner & Fransen, 1994]. The recent findings of delayed, post-stimulus, sustained activity in memory-related tasks, both in the temporal [Miyashita & Chang, 1988] and frontal [Wilson *et al.*, 1993] cortices, provides support to the plausibility of such attractor networks as a model of associative cortical areas.

A detailed formulation of the network used and simulation parameters is given in Appendix A. Each unit's connectivity is parametrized by σ , where smaller σ values denote a shorter (and more spatially-organized) connectivity range. The network's performance level is quantified by an *overlap* measure m ranging in the interval $[-1, +1]$. The overlap m measures the similarity between the network's end state and the cued memory pattern (which is the desired response), averaged over many trials with different input cues. We now describe the results of simulations examining the scaling rules derived in the previous section.

3.1 Performance Decrease with a Single Lesion

Figure 3 plots the network's performance as a function of the area of a single square-shaped *focal* lesion. As is evident, the spatially-organized connectivity enables the network to maintain its memory retrieval capacities in the face of focal lesions of considerable size. As the connectivity dispersion σ increases, focal lesions become more damaging. Also plotted in Figure 3 is the analytical curve calculated via rule 1 and expression (3) with $k = 5$, which matches well with the

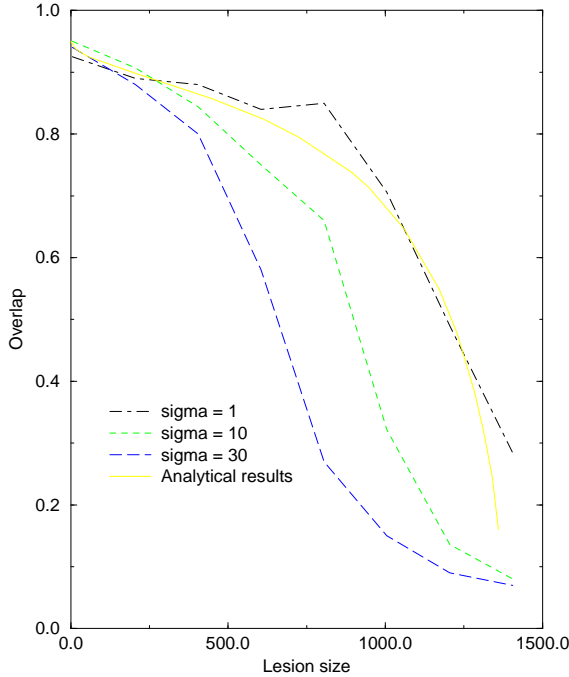


Figure 3: Network performance as a function of focal lesion size. Simulation results obtained in three different networks, each characterized by a distinct distribution of spatially-organized connectivity, and analytic results calculated with $k = 5$ using equation 2.

actual performance of the spatially-connected network parametrized by $\sigma = 1$. Concentrating on the study of focal lesions in a spatially-connected network, we shall adhere to the values $\sigma = 1$ and $k = 5$ hereafter, and compare the analytical and numerical results.

The performance of the network as a function of the *fraction* of the network lesioned, for different network areas A , is displayed in Figure 4. The analytical curves, plotted using equation (4) (with $k = 5$) are qualitatively similar to the numerical results (with $\sigma = 1$). The sparing effect of large networks is marked.

3.2 The Effects of Shape and Number

To examine rule 2, a rectangular structural lesion of area $s = 300$ was induced in the network. As shown in Figure 5a, as the ratio n between the sides is increased while holding the area constant, the network's performance further decreases, but this effect is relatively mild (note values on vertical axis). There is a fair quantitative agreement with the theoretical predictions obtained using (7), and these are also plotted in Figure 5. The effect of varying the lesion number while keeping the overall lesion area fixed stated in rule 3 is demonstrated in

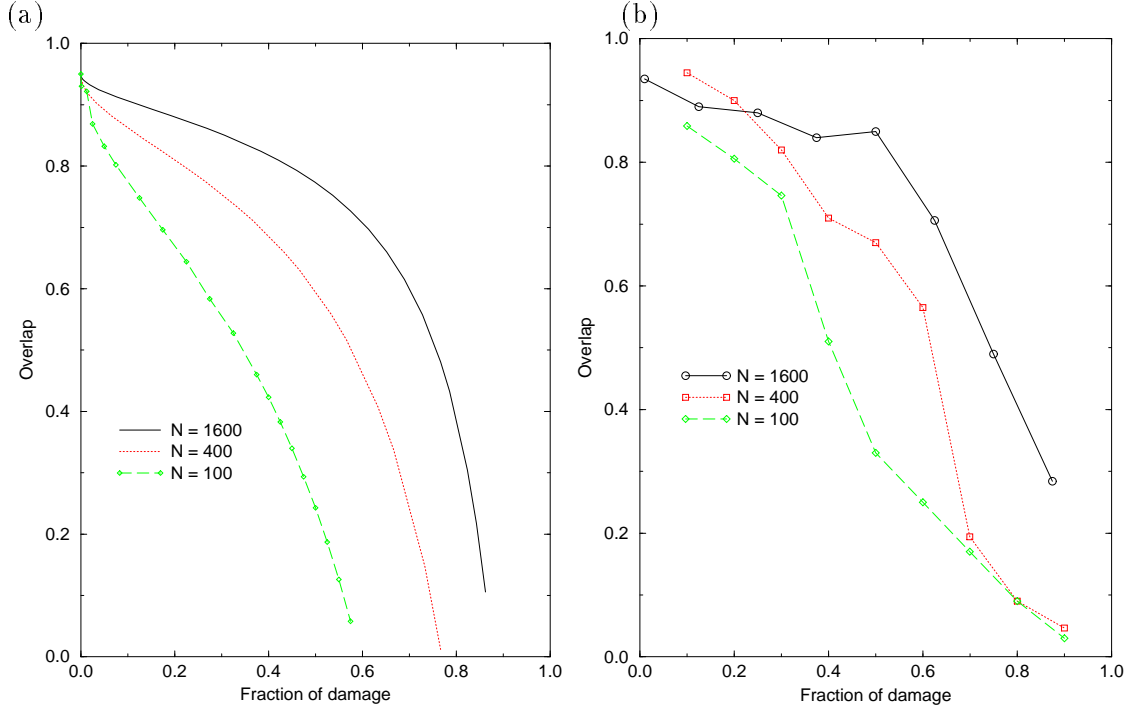


Figure 4: Network performance as a function of the fraction of focal damage, in networks of different sizes. Both analytical (a) and numerical (b) results are displayed.

Figure 5b. This Figure shows the effect of multiple lesions composed of 2, 4, 8 and 16 separate focal lesions. This effect is much stronger than seen with lesion shape (note values on vertical axis). As is also evident, the analytical results computed using (10) correspond quite closely with the numerical ones. However, in both figures, the analytically calculated performance is consistently higher than that actually achieved in simulations, as the d^2 term is omitted in the analytic approximation. Note also (Figure 5b) that as the lesion size is increased the analytic results correspond better to the simulation results, as d becomes smaller in relation to \sqrt{s} .

To compare the effects of focal and diffuse lesions, the performance achieved with a diffuse lesion of similar size is arbitrarily plotted on the 20th x-ordinate. It is interesting to note that a large multiple focal lesion ($s = 512$) can cause a larger performance decrease than a diffuse lesion of similar size. That is, at some point, when the size of each individual focal lesion becomes small in relation to the width of each unit's connectivity, our analysis loses its validity, and rule 3 does not hold any more. Hence, the effect of a diffuse lesion on the network's performance cannot be calculated by viewing it as a 'limiting case' of multiple focal lesions.

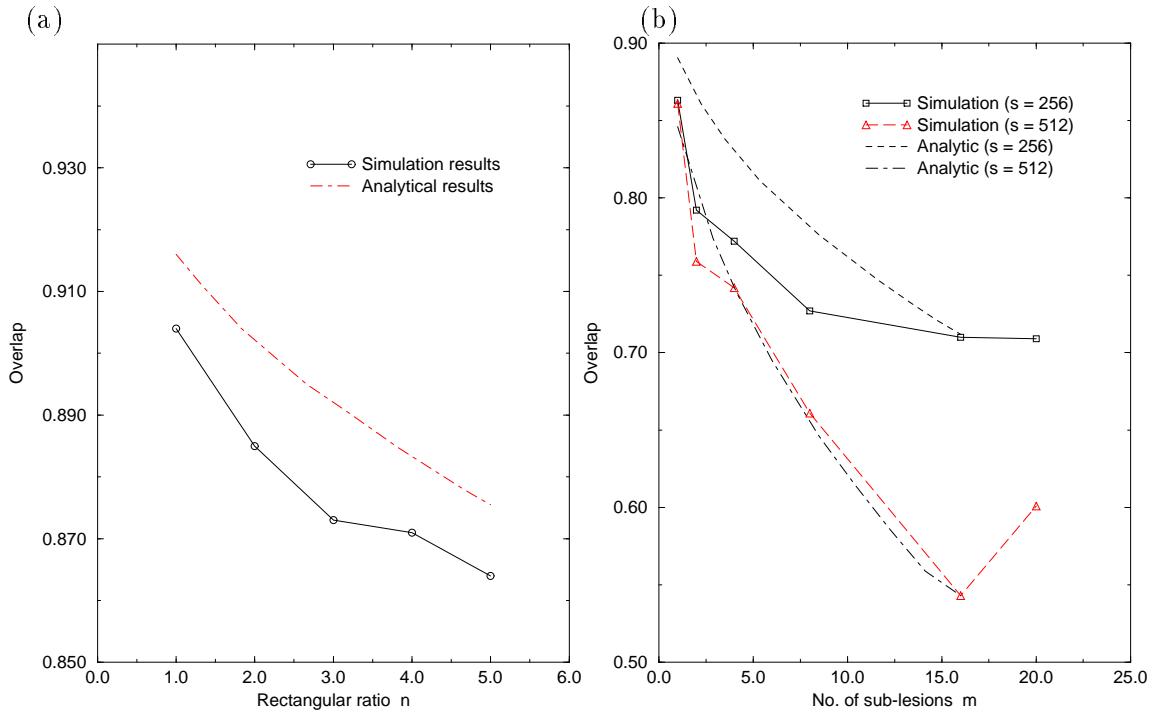


Figure 5: Network performance as a function of structural focal lesion shape (a) and number (b), while keeping total structural lesion area constant. Both numerical and analytical results are displayed. The simulations were performed in a network whose connectivity was generated with $\sigma = 1$. The analytical results are for the corresponding $k = 5$. In Figure 5b, the x-ordinate denotes the number of separate sub-lesions (1,2,4,8,16), and, for comparison, the performance achieved with a diffuse lesion of similar size is plotted arbitrarily on the 20th x-ordinate.

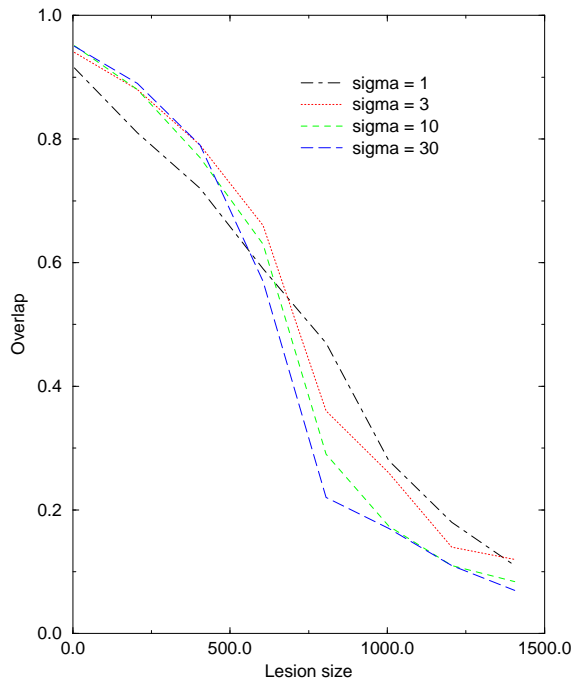


Figure 6: Network performance as a function of diffuse lesion size. Simulation results obtained in four different networks, each characterized by a distinct distribution of spatially-organized connectivity.

3.3 Diffuse Lesions in Spatially-Organized Networks

Figure 6 displays how the performance of the network degrades when *diffuse* structural lesions of increasing size are inflicted upon it by randomly selecting units on the lattice and clamping their activity to zero. While the performance of non-spatially connected networks manifests the classical sharp decline (denoted as ‘catastrophic breakdown’ [Amit, 1989]) at some critical lesion size (Figure 6, $\sigma = 30$), the performance of spatially-connected networks (Figure 6, $\sigma = 1$) degrades in a more gradual manner as the size of the diffuse lesion increases. It is of interest to note that this ‘graceful’ degradation parallels the gradual clinical and cognitive decline observed in the majority of Alzheimer patients [Katzman, 1986; Katzman & et. al., 1988]. A comparison of Figures 3 and 6 demonstrates that diffuse lesions are generally more detrimental than a *single* focal lesion of identical area.

4 The Functional Impairment Span d

The correspondence obtained between the theoretical and simulation results presented in the previous section testifies to the validity of the lesion-invariant impairment span that has been central to our analysis. This assumption is further supported directly by extensive simulations demonstrating that the span d remains practically invariant when the lesion size is varied (for large lesions). We now turn to study the influence of several factors such as the spatial connectivity distribution σ and the noise level T (defined in Appendix A) on the the functional impairment span³. The simulation results described below are compared with analytical results obtained by iterating the overlap system of equations (20) derived in Appendix B. They describe the dependence of an overlap vector whose components are ‘local’ overlaps measured at consecutive distances from the border of the lesion (and hence termed *distance-overlaps*), on various parameters of the network.

Figure 7a displays the distance-overlap span obtained with an almost noiseless ($T = 0.001$) network and with a network with noisy dynamics ($T = 0.020$). As is evident from the analytical results, as the noise level is increased the span d is markedly increased (e.g., from roughly $d = 3$ to $d = 6$ in Figure 7a). In other words, the functional damage is significantly larger for increased noise levels. As one would expect, increasing the noise levels also results in decreased performance levels. As shown in Figure 7b, after decreasing the synaptic connectivity by randomly eliminating some fraction of the synapses of each unit (leaving each unit with 40 instead of 60 incoming synapses), the network’s performance decreases in a similar manner to the noisy dynamics case, but there is only a slight increase in d . In our simplified network the effects of random synaptic deletion are essentially equivalent to that of random neural damage (inactivation of some units), so that Figure 7b also illustrates how diffuse neuronal degeneration in the region of the functional lesion (as observed in the perilesion area after stroke) would effect the severity of multi-infarct dementia. The distance-overlap span calculated via analytic approximation (20) qualitatively agrees with simulation results, but is shifted upward at short distances when compared to the latter. This shift results primarily from approximation in the theoretical derivation: neglecting the effects of the non-cued memory patterns, and from ne-

³Note that the average activity levels and performance have a similar functional impairment span. This follows since, due to the decrease in synaptic inputs to neurons in the functional lesion region, the probability of retrieval errors is mainly due to neurons that should fire (i.e., they belong to the cued pattern) but are silent, and not due to neurons that fire erroneously (i.e., they do not belong to the cued pattern).

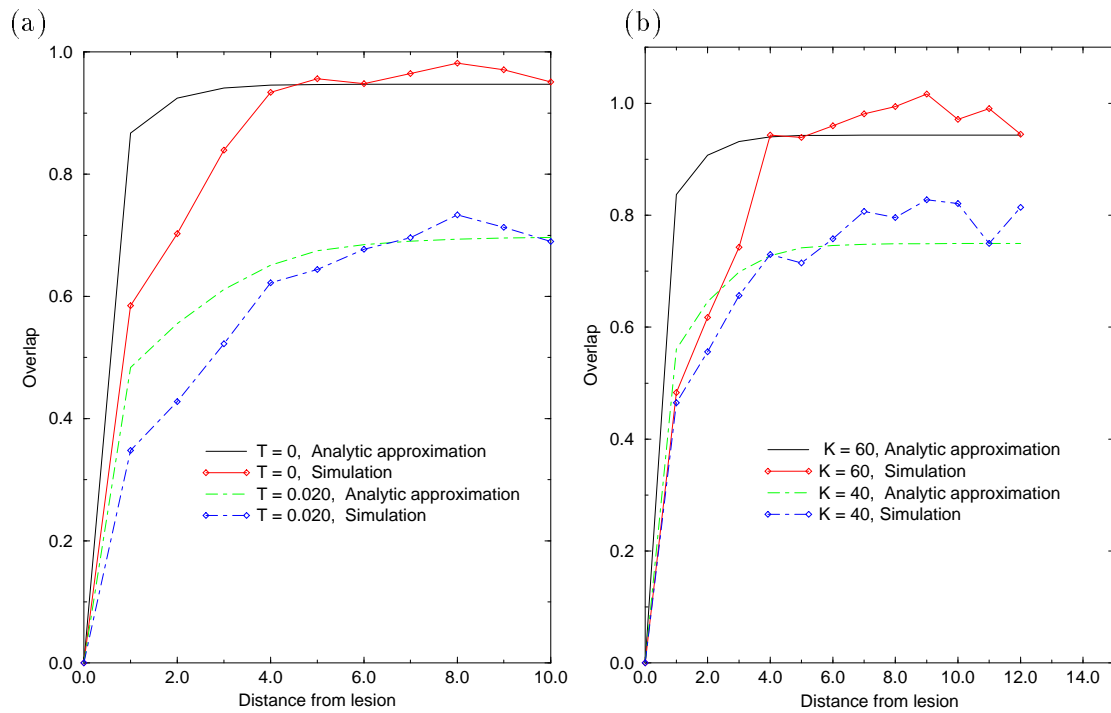


Figure 7: Performance levels as a function of distance from the lesion border. With different (a) Noise levels and (b) connectivity. Both analytical and simulation results are displayed; $s = 200$.

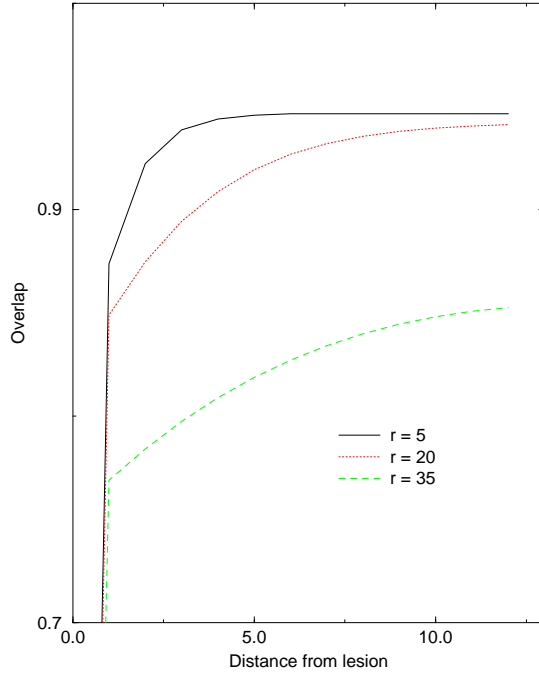


Figure 8: Performance levels as a function of distance from the lesion border, with different r values. Analytical results.

glecting the correlations that evolve between subsequent external inputs to the same unit due to the spatially-organized connectivity.

Studying the dependency of the span d on the connectivity dispersion parameter σ (or its equivalent r in the theoretical expressions (20)) requires larger networks than those which we could practically simulate. Hence, only analytical results are presented in Figure 8. As is evident, increased r levels result in a marked increase in the distance-overlap span, and in a more gradual performance gradient. As one would expect, at sufficiently high levels of r the typical gradient of the distance-overlaps' span vanishes (not shown in Figure 8).

The practically negligible changes in the length of the span d following random synaptic deletion (demonstrated in Figure 7) may then be understood by noticing that, on the one hand, random synaptic deletion tends to increase the noise in network and hence increase d , but on the other hand it tends to shorten the average connectivity span and hence decrease d .

5 Discussion

We have presented a simple analytical framework for studying the effects of focal lesions on the functioning of spatially organized neural networks. The analysis presented is quite gen-

eral and a similar approach could be adopted to investigate the effect of focal lesions in other neural models, such as models of random neural networks [Minai & Levy, 1993] or cortical map organization [Sutton *et al.*, 1994].

Using this analysis, specific scaling rules have been formulated describing the functional effects of structural focal lesions on memory retrieval performance in associative attractor networks. The functional lesion scales as the square root of the size of a single structural lesion, and the form of the resulting performance curve depends on the impairment span d . Surprisingly, the same fraction of damage results in significantly less performance decrease in larger networks, pointing to their relative robustness. As to the effects of shape and number, elongated structural lesions cause more damage than more symmetrical ones. However, the number of sub-lesions is the most critical factor determining the functional damage and performance decrease in the model. Numerical studies show that in some conditions multiple lesions can damage performance more than diffuse damage, even though the amount of lost innervation is always less in a multiple focal lesion than with diffuse damage.

The main parameter determining the relation between structural and functional damage is the length of the impairment span d . This span has been found to increase with the noise level T of the network and the connectivity dispersion parameter σ . It should be noted that when d gets large (in relation to the network dimensions) multiple lesions are likely to ‘interact’ (i.e., their resulting functional lesions are likely to intersect) and increase the overall performance deterioration.

In the introduction we described the parallel between the structural/functional distinction that underlies this study and a similar distinction made with regard to infarcted tissue versus metabolically impaired regions in multi-infarct dementia. What are the clinical implications of this study with respect to the latter disease? Our results indicate a significant role for the number of infarcts in determining the extent of functional damage and dementia in multi-infarct disease. In our model, multiple focal lesions cause a much larger deficit than their simple ‘sum’, i.e., a single lesion of equivalent total size. This is consistent with clinical studies that have suggested the main factors related to the prevalence of dementia after stroke to be the infarct number and site, and not the overall infarct size, which is related to the prevalence of dementia in a significantly weaker manner [Tatemichi *et al.*, 1990; Tatemichi, 1990; del Ser *et al.*, 1990]. As noted by [Hachinski, 1983], “In general, the effect of additional lesions of the brain increases

with the number of lesions already present, so that the deficits in the brain do not add up, they multiply”.

We have found that decreasing the connectivity of each unit, and decreasing the fidelity of network dynamics by increasing the noise level, may not only lead to a decrease in the overall level of performance, but also to an increase in the length of the distance-overlap span in the perilesion area. The degenerative synaptic changes occurring as Alzheimer’s disease progresses are known to lead to a reduction in the number of synapses in a unit volume of the cortex (e.g., [DeKosky & Scheff, 1990]), and the accompanying synaptic compensatory changes increase the level of noise in the system [Horn *et al.*, 1993]. This offers a plausible explanation for the ‘multiplicative’ interaction occurring between co-existing Alzheimer’s and multi-infarct dementia [Tatemichi, 1990], where cortical atrophy contributes as an independent variable to the severity of stroke symptomatology [Levine *et al.*, 1986] and increases the severity of stroke symptomatology in Alzheimer patients.

The loss of innervation from a focal stroke region to its immediate surroundings such as that studied in this paper may be viewed as sort of ‘local diaschisis’. In contrast, global, interhemispheric diaschisis denotes the ‘disconnection’ of neural structures which are far apart in the brain, and may lead to structurally normal regions with reduced metabolism as observed in several neurological disorders [Feeney & Baron, 1986]. Such a metabolic depression of apparently intact structures involving Papez’s circuit and basal anterior regions has been recently observed in human patients suffering from ‘pure’ amnesia [Fazio *et al.*, 1992]. Given more information concerning the patterns of connectivity between these structures, it may be possible in the future to study the functional consequences of interhemispheric diaschisis. Future studies of diaschisis may also address the effects of subcortical infarcts, which frequently accompany cortical lesions in multi-infarct dementia. Interestingly, just very recently it has been shown that, as with cortical infarction, with subcortical infarction the number of infarcts but not the volume of infarction (as measured in computerized tomography scans) are significantly associated with cognitive impairment in stroke patients [Corbett *et al.*, 1994].

References

- [1] D.J. Amit. *Modeling brain function: the world of attractor neural networks*. Cambridge University Press, 1989.
- [2] J.A. Anderson. Cognitive and psychological computation with neural models. *IEEE Trans. on Systems, Man, and Cybernetics*, 13 (5):799–815, 1983.
- [3] A. Corbett, H. Bennett, and S. Kos. Cognitive dysfunction following subcortical infarction. *Arch. Neurol.*, 51:999–1007, 1994.
- [4] 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.
- [5] T. del Ser, F. Bermejo, A. Portera, J.M. Arredondo, C. Bouras, and J. Constantinidis. Vascular dementia: a clinicopathological study. *Journal of Neurological Sciences*, 96:1–17, 1990.
- [6] F. Fazio, D. Perani, M.C. Gilardi, F. Colombo, S.F. Cappa, G. Vallar, V. Bettinardi, E. Paulesu, M. Alberoni, S. Bressi, M. Franceschi, and G.L. Lenzi. Metabolic impairment in human amnesia: A PET study of memory networks. *Journal of cerebral blood flow and metabolism*, 12(3):353–358, 1992.
- [7] D.M. Feeney and J.C. Baron. Diaschisis. *Stroke*, 17:817–830, 1986.
- [8] P.B. Gorelick, A. Chatterjee, D. Patel, G. Flowerdew, W. Dollear, J. Taber, and Y. Harris. Cranial computed tomographic observations in Multi-Infarct Dementia. *Stroke*, 23:804–811, 1992.
- [9] V. Hachinski. Multi-infarct dementia. *Neurol. Clin.*, 1:27–36, 1983.
- [10] W.D. Heiss, H.G. Emunds, and K. Herholz. Cerebral glucose metabolism as a predictor of rehabilitation after ischemic stroke. *Stroke*, 24:1784–1788, 1993.
- [11] W.D. Heiss, J. Kessler, H. Karbe, G.R. Fink, and G. Pawlik. Cerebral glucose metabolism as a predictor of recovery from aphasia in ischemic stroke. *Archives of Neurology*, 50:958–964, 1993.

- [12] D. Horn and E. Ruppin. Synaptic compensation in attractor neural networks: Modeling neuropathological findings in schizophrenia. *Neural Computation*, 1994. To appear.
- [13] 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.
- [14] J.M. Karlholm. Associative memories with short-range higher order couplings. *Neural Networks*, 6:409–421, 1993.
- [15] R. Katzman and et. al. Comparison of rate of annual change of mental status score in four independent studies of patients with alzheimer's disease. *Ann. Neurology*, 24(3):384–389, 1988.
- [16] R. Katzman. Alzheimer's disease. *New England Journal of Medicine*, 314(15):964–973, 1986.
- [17] A. Lansner and E. Fransen. Improving the realism of attractor models by using cortical columns as functional units. In *CNS94*, 1994.
- [18] D.N. Levine, J.D. Walrach, L. Benowitz, and R. Calvino. Left spatial neglect: effects of lesion size on severity and recovery following right cerebral infarction. *Neurology*, 36:362–366, 1986.
- [19] C.K. Liu, B.L. Miller, J.L. Cummings, C.M. Mehringer, M.A. Goldberg, S.L. Howng, and D.F. Benson. A quantitative MRI study of vascular dementia. *Neurology*, 42:138–143, 1990.
- [20] I. Meilijson and E. Ruppin. Optimal signalling in attractor neural networks. *Network*, 5(2):277–293, 1994.
- [21] J.S. Meyer, K.L. McClintic, R.L. Rogers, P. Sims, and K.F. Mortel. Aetiological considerations and risk factors for multi-infarct dementia. *J. Neurol. Neurosurg. Psychiatry*, 51:1489–1497, 1988.
- [22] R. Mieke, K. Herholz, M. Grond, J. Kessler, and W.D. Heiss. Severity of vascular dementia is related to volume of metabolically impaired tissue. *Archives of Neurology*, 49:909–913, 1992.

- [23] G. Mies, L.M. Auer, H. Ebhardt, H. Traupe, and W.D. Heiss. Flow and neuronal density in tissue surrounding chronic infarction. *Stroke*, 14(1):22–27, 1983.
- [24] A.A. Minai and W.B. Levy. Setting the activity level in sparse random networks. *Neural Computation*, 6:85–99, 1993.
- [25] Y. Miyashita and H.S. Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331:68–71, 1988.
- [26] J. Reggia, R. Berndt, and L. D’Autrechy. Connectionist models in neuropsychology. In *Handbook of Neuropsychology*, volume (9), pages 297–333. Elsevier Science, Amsterdam, 1994.
- [27] E. Ruppin and J. Reggia. A neural model of memory impairment in diffuse cerebral atrophy. *Br. Jour. of Psychiatry*, 165, 1994. To appear.
- [28] E. Ruppin. Neural modeling of psychiatric disorders. *Network*, 1995. Invited review paper, to appear.
- [29] G. Sutton, J. Reggia, S. Armentrout, and C. D’Autrechy. Map reorganization as a competitive process. *Neural Computation*, 6:1–13, 1994.
- [30] T.K. Tatemichi, M.A. Foulkes, J.P. Mohr, J.R. Hewitt, D. B. Hier, T.R. Price, and P.A. Wolf. Dementia in stroke survivors in the stroke data bank cohort. *Stroke*, 21:858–866, 1990.
- [31] T. K. Tatemichi. How acute brain failure becomes chronic: a view of the mechanisms of dementia related to stroke. *Neurology*, 40:1652–1659, 1990.
- [32] A. M. Thomson and J. Deuchars. Temporal and spatial properties of local circuits in the neocortex. *Trends in neuroscience*, 17(3):119–126, 1994.
- [33] 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.
- [34] F.A.W. Wilson, S.P.O. Scalaidhe, and P.S. Goldman-Rakic. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260:1955–1958, 1993.

- [35] C.C. Wood. Variations on a theme by Lashley: lesion experiments on the neural model of Anderson, Silverstien, Ritz and Jones. *Psychological Review*, 85(6):582–591, 1978.

Appendix A: The Numerical Simulations

The attractor network used in this study is composed of N units, where each unit i is described by a binary variable $S_i = \{1, 0\}$ denoting an active (firing) or passive (quiescent) state, respectively. M distributed memory patterns ξ^μ , where superscript μ indicates a pattern index, are stored in the network. The elements of each memory pattern are randomly chosen to be 1 or 0 with probability p or $1 - p$, respectively, with $p \ll 1$.

With each set of parameters characterizing a given network, the behavior of the network is monitored over many trials. In each trial, the initial state of the network $S(0)$ is random, with average activity level $q < p$, reflecting the notion that the network's baseline level of activity is lower than its activity in persistent memory states. Each unit's state is updated stochastically, in accordance with its input. The input (post-synaptic potential) h_i of element i at time t is the sum of internal contributions from other units in the network and external contribution F_i^e , given by

$$h_i(t) = \sum_j w_{ij} S_j(t-1) + F_i^e \quad (13)$$

where $S_j(t)$ is the state of unit j at time t and w_{ij} is the weight on the directed connection to unit i from unit j . The updating rule for unit i at time t is given by

$$S_i(t) = \begin{cases} 1, & \text{with probability } G(h_i(t) - \theta) \\ 0, & \text{otherwise} \end{cases}, \quad (14)$$

where G is the sigmoid function $G(x) = 1/(1 + \exp(-x/T))$, T denotes the noise level, and θ is a uniform threshold which is optimally tuned to guarantee perfect retrieval in the network's premorbid state [Horn & Ruppin, 1994].

The weights of the internal synaptic connections are assigned based on the stored memory patterns ξ^μ using

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^M (\xi_i^\mu - p)(\xi_j^\mu - p), \quad (15)$$

and in each trial the external input component F_i^e is used to present a stored memory pattern (say ξ^1) as an input cue to the network, such that

$$F_i^e = e \cdot \xi_i^1 \quad (16)$$

where $0 < e < 1$ is a network-wide constant. Following the dynamics defined in (13) and (14), the network state evolves until it converges to the vicinity of an attractor stable state.

The network's performance level is measured by the similarity between the network's end state S and the cued memory pattern ξ^μ (which is the desired response), conventionally denoted as the *overlap* m^μ [Tsodyks & Feigel'man, 1988], and defined by

$$m^\mu = \frac{1}{p(1-p)N} \sum_{i=1}^N (\xi_i^\mu - p) S_i . \quad (17)$$

where the sum is taken only over the viable, non-lesioned units. This overlap measure, ranging in the interval $[-1, +1]$, keeps track of the neurons which should correctly fire and also counts with lower negative weighting the erroneously firing ones. In all simulations we report the average overlap achieved over 100 trials.

In all simulations, $M = 20$ sparse random memory patterns (with a fraction $p = 0.1$ of 1's) were stored in a network of $N = 1600$ units, placed on a 2-dimensional lattice. The external input magnitude is $e = 0.035$ and the noise level is $T = 0.005$. Unlike in the original Tsodyks-Feigelman model which is fully-connected, the network in our model has spatially organized connectivity, where each unit has $K = 60$ incoming connections determined randomly with a Gaussian probability $\phi(z) = \sqrt{1/2\pi} \exp(-z^2/2\sigma^2)$, where z is the distance between two units in the array, and σ determines the extent to which each unit's connectivity is concentrated in its surrounding neighborhood. A structural lesion of the network is realized by clamping the activity state of the lesioned units to zero.

Appendix B: Calculating Functional Impairment

We investigate here the functional impairment resulting from a single focal (circular or square) structural lesion. Due to symmetry, it is sufficient to study the performance levels obtained at increasing distances from the lesion border, until distance $(\sqrt{A} - a_s)/2$ from the lesion border, where a_s denotes the diameter (or side, if rectangular) of the structural lesion. That is, instead of measuring the overlap $m^1(t)$ over just the whole network, we are now interested in tracing a series of *distance-specific overlaps* $m_l^1(t)$, $l = -r + 1, \dots, (\sqrt{A} - a_s)/2$, where $m_l^1(t)$ denotes the similarity between the cued memory pattern (say ξ^1) and the current network state over the units which are located at distance l of the structural lesion's border, and r denotes the radius of connectivity of each unit on the lattice.

Assume that, without loss of generality, the network is cued with memory pattern ξ^1 as an

input. As shown in [Horn & Ruppin, 1994], the overlap $m^1(t)$ at iteration t can be estimated by

$$m^1(t) = P(S_i(t) = 1 | \xi_i^1 = 1) - P(S_i(t) = 1 | \xi_i^1 = 0) , \quad (18)$$

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

$$P(S_i = 1 | h_i) = \Phi \left(\frac{\mu}{\sqrt{(1.702T)^2 + \sigma^2}} \right) \quad (19)$$

where Φ is the standard normal distribution function. Using the standard decomposition of the neuron's external input to signal and noise terms and neglecting correlations that may evolve [Amit, 1989; Meilijson & Ruppin, 1994], the external input of neuron i at distance l from the lesion border is conditionally normally distributed (with respect to its correct value ξ_i^1) with means

$$\mu_1 = \frac{p(1-p)^2}{C} \sum_{j=l-r}^{l+r} c_{|l-j|} m_j^1(t) + e - \theta$$

and

$$\mu_0 = -\frac{p^2(1-p)}{C} \sum_{j=l-r}^{l+r} c_{|l-j|} m_j^1(t) - \theta$$

and variance approximated by

$$\sigma^2 \approx \alpha q p^2 c^2 \approx \alpha p^3 c^2 ,$$

where $\alpha = M/N$ denotes the memory load, and the coefficients c_k , $k = 0 \dots r$ compose a weighting kernel which is convolved with the distance-overlap values. This kernel, normalized by $C = c_0 + 2 \sum_{k=1}^r c_k$, reflects the spatial organization of each unit's connectivity. Hence, by (18) and (19) we obtain the iterative distance-overlaps map

$$m_l^1(t+1) = \Phi \left(\frac{\frac{p(1-p)^2}{C} \sum_{j=l-r}^{l+r} c_{|l-j|} m_j^1(t) + e - \theta}{\sqrt{(1.702T)^2 + \alpha p^3}} \right) - \Phi \left(\frac{\frac{-p^2(1-p)}{C} \sum_{j=l-r}^{l+r} c_{|l-j|} m_j^1(t) - \theta}{\sqrt{(1.702T)^2 + \alpha p^3}} \right) . \quad (20)$$

The spatial overlap vector has initial values $m_l^1(0) = m_{intact}$, where m_{intact} is the level of performance of the network in its intact, undamaged, state. A lesion is induced by clamping the values of the r 'leftmost' overlaps to zero, that is $m_l^1(t) = 0$, $l = -r + 1 \dots 0$, $t = 0, 1, \dots$. Thereafter, the evolution of the spatial overlaps upon the network is calculated by iterating (20),

and rescaling the overlaps achieved after each iteration of (20) by multiplying them by m_{intact} . The value of m_{intact} is set to 0.95, similar to the value of $P(0)$ used in the previous section. Due to the monotonicity of the map defined in (20) (see [Horn & Ruppin, 1994]), it is straightforward to show that at a given moment t the distance-overlaps $m_l^1(t)$ increase with increasing distance l , and that due to the monotonicity of both the iterative map and the the latter distance-overlaps' gradient, the distance-overlap vector m_l^1 converges to a fixed state. At some distance d from the lesion, $m_d^1 = m_{intact}$, determining the the span d of the functional lesion.

In this work we use (20) to study the extent and form of the overlaps' gradient under various conditions, and compare these results with the gradient measured in simulations carried with identical parameters. Unless specified otherwise, the analytical approximations and simulations are performed with $N = 1600$, $M = 20$, $p = 0.1$, $e = 0.035$ and $T = 0.005$. Since each unit has $K = 60$ incoming connections and is placed on a 2-D lattice, $r = 4$ and the weighting kernel is taken as $c_i = 5 - i$, $i = 1 \dots 4$.