

SLEEP AS TEMPORARY BRAIN DISSOCIATION

Y. Bar-Yam

ECS, 44 Cummington St., Boston University, Boston, MA 02215

(currently, New England Complex Systems Institute, 24 Mt. Auburn St., Cambridge, MA 02138)

During sleep the brain is active but is largely isolated from sensory neurons. It is here proposed that during sleep the mind further subdivides into isolated neuron groups. This division decomposes imprinted experience from the waking period into pieces which then form the building blocks for analyzing and responding to future circumstance. The approach is based on the perspective that the central purpose of the brain is not to remember experiences, but rather to obtain from them knowledge that will serve in future circumstances.

Modern theories of sleep suggest it serves a biological restorative function, or that sleep exists because of a survival advantage in removing primitive man from danger.^{1,2} However, the existence of dreams has motivated theories in which sleep plays an important role in psychology or how the brain processes information.^{3,4} One modern framework for describing information processing by the brain is the neural network. In this framework, Crick and Mitchinson^{5,6} suggest that dreams cause selective forgetting of undesirable or parasitic neural network states. In contrast Geszti and Pázmándi⁷ suggest dreams are a form of relearning. In these works, roles in information processing are attributed to Rapid Eye Movement (REM) sleep, or dream sleep. The other parts of sleep, where dreams are infrequent (non-REM sleep), are still generally believed to have a biological role. However, total sleep deprivation causes psycho-functional, not physiological, deterioration in humans and the primary effects occur with loss of non-REM sleep.²

In this article it is proposed that both REM and non-REM sleep play a significant role in information processing by the brain. It is suggested that during sleep brain subdivisions are temporarily isolated from each other. The known specificity of neurotransmitters makes it possible selectively to 'turn-off' synapses or axons that connect between subdivisions. Such selective control

enables isolation of the brain from sensory input even though sleep is a neurologically active state.⁸ The possible role of further internal dissociation of the brain appears to be discussed here for the first time.

It is proposed that the temporary subdivision of the brain during sleep, and a selective relearning process during this time, enables and maintains the distinct roles of different subdivisions of the brain during waking.⁹ The fundamental motivation for subdivision of the brain is the need to generalize experiences by isolating aspects that may reoccur in other contexts. Quite literally the act of combining aspects of prior experience is directly the result of recombining states of partially independent brain subdivisions.

It is further suggested that dissociation in sleep performs an essential restorative function. Neural networks fail catastrophically when overloaded. Selective relearning also results in a selective forgetting of information that prevents overload failure. Experimental observation of psycho-functional failure after sleep deprivation,² which prevents this restorative function, may be directly related to overload failure.

Consequences of this discussion are far reaching both for public policy and progress in understanding brain function and disorders. Public attitudes, and corporate behavior, motivated by or motivating scientific thought often proceeds on the assumption that sleep is a waste of time. In contrast, these proposals suggest that well-balanced sleep schedules are of central importance in human functioning not only for the short term effects of sleep deprivation but for its long term effects. Moreover, the temporary subdivision of the brain and selective relearning may provide a step towards understanding mechanisms of pattern recognition, the 'logic' of human language and the ability of separate parts of the brain to function coherently. Finally, it provides predictions that can be tested by the new techniques currently available for mapping regional brain activity.

This article is arranged as follows. First, the conventional neural model of adaptive learning is reviewed. The existence of brain subdivision is motivated within the neural model. The problem of neural network overload and the impact of selective relearning is discussed. The role of temporary dissociation and the purpose of different levels of sleep are described. The impact of sleep

deprivation and other malfunctions of the sleep process are presented. Finally, it is suggested that the role of sleep provides an intrinsic basis for human individuality.

The neural model

In the conventional model of neural networks, the brain is composed of neurons that can be in an excited or passive state. Neurons affect the activity of other neurons through synaptic connections. The 'state of the mind' at any time is described by the activity pattern of the neurons. This activity pattern evolves in time because the activity of each neuron is determined by the activity of all neurons at an earlier time and the excitatory or inhibitory synapses between them. During waking hours, sensory information directly affects sensory neurons and, therefore, in part determines the state of neuron excitation throughout the brain. Actions are the result of motor neuron activity and, therefore, also reflect the state of neuron excitation. A substantial part of synaptic connections are 'hard-wired' performing functions that are pre-specified. However, synaptic strengths are also affected by the state of neuronal excitation. This influence constitutes a basic step in learning called imprinting originally proposed by Hebb in 1949.

In theoretical models,^{10,11} an imprinted pattern of neural excitations can be recovered if a sufficiently large part of the pattern is re-imposed on the neurons. Evolving the activity pattern then causes the complete original imprinted pattern to be recovered. This is an associative memory which associates the restored pattern to the part-of-it that was imposed. Such neural networks are also considered to be capable of generalization. This capability arises because the region of 'possible experiences' near a particular memory evolves by neural dynamics to the memory. This region, called its basin-of-attraction, becomes a generalization of the memory.

Why brain subdivision?

It is generally accepted that the storage capacity of a neural network increases with the degree of interconnectedness. For a network where each neuron is connected to every other neuron the number of imprints that can be recalled αN is proportional to the number of neurons N with a constant of proportionality α somewhat dependent on the particular imprinting rule. When

additional imprints are added an overload catastrophe causes erasure of all information.¹⁰ Subdivision *inherently* results in a decreased storage capacity. Why then subdivide the brain?

An illustrative example of the subdivided function of the brain is found in vision. Vision begins with a continuous field of light on the retina. However, we describe the image using object attributes: color, shape and movement. In the vision system information that is relevant to each of the attribute categories is mapped to a distinct part of the brain. For example, color is separated from the details of shape and movement and routed to one part of the brain. This part of the brain is then responsible for 'recognizing' colors. In this system the number of distinct images that may be stored is the product of the number of colors, shapes and movements. Considering the total number of neurons N the number of neurons in a subdivision is roughly $N/3$ and the number of distinct images that may be stored is $(N/3)^3$. This is useful because color, shape and motion occur in various combinations because they are largely independent of each other. More generally, the effective storage of information grows exponentially $(N/q)^q$ with q brain subdivisions.

Logic of natural language

The subdivided functioning of the brain may be recognized in the artificial world of 'natural' language. Consider a network large enough to be subdivided into three networks each of which can store three words (coded appropriately). A complete network would then be able to store nine sentences with three words each since the storage capacity grows linearly with size. On the subdivided network we could imprint three sentences and twenty-seven sentences would be recognized (see Fig. 1).

The central difference between the set of sentences that can be remembered by the full network and the subdivided network may be summarized by the notion of 'content' vs. 'grammar.' The complete network knows more full sentences but does not have knowledge of the divisibility of the sentences into parts that can be put together in different ways. The subdivided network knows the parts but has no relationship between them, thus it knows grammar but does not know any context information, like who it is that fell.

Clearly the actual process in the human brain is a combination of the two, where sentences make sense or are 'grammatically correct' if properly put together out of largely interchangeable parts, but an actual event or recalled incident is a specific combination. This can be captured in the network by having a partial interconnection between subnetworks.

Simulations have been performed to test these ideas.¹² For example (see Fig. 2), a network with 100 neurons using conventional Hebbian imprinting can store approximately 12 images. Subdividing it into four equal subdivisions results in a storage capacity of roughly three to four sub-images and this is comparable to the storage capability of the divided network for complete images. However the number of combinations of sub-images is nearly 400. If the strength of interconnecting synapses is set at 0.2 of the strength of sub-network connections, 6 full imprints can be stored and 60 combinations. Thus it is possible to achieve a balance between recall of complete states and independent parts.

Having motivated the existence of subdivision, we must now address the problem of overload.

Mental capacity and overload

An overloaded network fails catastrophically. As the number of imprints on a neural network increases the basins-of-attraction of each decreases. In addition, spurious states composed of pieces of different imprints arise that are also stable and compete with the real memories. Catastrophic failure of the network occurs when the spurious states dominate and the real memories cannot be recalled. A variety of phenomena accompany this transition that is known to be analogous to a 'spin-glass' transition¹³ and which shares some qualitative features with the usual transition of a liquid to a glass. Thus not only do the memories become irretrievable, but the barriers to changes in the state of the neural network become insurmountable -- resulting in a locking of the neural state.

In order to avoid overload failure a process whereby information is erased is necessary. A storage device that selectively retains recent memories is known as a palimpsest. Several models for neural-network palimpsests have been proposed. An approach to selective memory retention better suited for understanding sleep is selective relearning.⁷ Strictly, it is not a palimpsest because the retention of memories is not determined solely by order of imprint. It is important to understand the

tradeoff between the strength of a memory (the size of its basin-of-attraction) and the number of memories that can be stored. The objective is to increase the basin-of-attraction of some memories at the expense of others. To strengthen prior imprints selectively it is necessary to retrieve some of the imprinted activity patterns and reimprint them. Recovery of the imprinted states may be achieved by use of the network itself. Starting from random activity patterns, evolution recovers previously imprinted states. The randomness indicates that the selection of memories has some degree of chance. Moreover this procedure sometimes recovers spurious states. Nevertheless, reimprinting the result is effective in selective reinforcement of memories.

Simulations demonstrate the effectiveness of this process as a selective filter on memories. In a test case (see Fig. 3), a network with 100 neurons was first imprinted with 8 images. Then four random starting points were evolved to stable states. After imprinting these states it was found that on average two stored memories were strengthened at the expense of the others, which were destabilized. Two memories rather than four are reinforced because duplicate reinforcements of the same memory occur and in other cases spurious states are reinforced. Significantly, the process of recovering imprinted memories is a delicate one and it was found important to include a measured amount of noise that promotes the recovery of imprinted images. This use of noise is analogous to its use in avoiding shallow local minima in a minimization problem.¹⁰

Subdivision and selective relearning

The motivation for brain subdivision and the relearning procedure may now be combined into a model for the phenomenon of sleep as a temporary dissociation with relearning.

To simplify the discussion of temporary dissociation, hard-wired connections may be separated from adaptive synapses. Thus, it is assumed that sensory information is mapped by a hard-wired neural network onto a second network containing adaptive synapses. It is further assumed that there are several mappings which perform different operations on sensory information before affecting the neurons of the adaptive network. For vision, one mapping may highlight edges by performing a second-order difference of intensities, another highlights motion by taking temporal or delayed differences, a third extracts only general color information.⁹

During waking the network is imprinted with multiple images. Then while asleep, the network is temporarily subdivided by diminishing the strength of synapses that run between subnetworks. While the subnetworks are dissociated a selective relearning process occurs. Starting from random excitations the neural evolution recovers previously learned information, which is then re-imprinted. As a direct result other memories are diminished in strength. Finally, the strength of inter-sub-network synapses are restored but at a reduced level.

In contrast to conventional artificial networks, the objective of this model of brain function is not to recall each imprint. Relearning emphasizes (1) the memories with largest basins-of-attraction, and (2) the most different memories. Minor deviations or details in standard elements are forgotten, and connections between imprints on different subdivisions are preferentially weakened. By strengthening the basins of each subnetwork separately, these basins continue to be relevant to the description of the subnetwork state even after it is reconnected to the rest of the network during waking. As shown by the simulations, this enables subnetwork imprints to be put together in new ways. Since the process of dissociation and relearning is incremental, recent imprints (experiences) should be remembered in full, while earlier imprints become progressively decomposed into elements.

Sleep as dissociation

Sleep is known to have several levels identified by different brain electrical signals (EEG). There are five recognized levels, the two deepest being Slow Wave Sleep (SWS) while the shallowest is Rapid Eye Movement (REM) Sleep. Typically, in the first part of sleep, the deepest level is reached rapidly. Then the level of sleep alternates in a pattern of shallower and deeper levels with the average level becoming shallower as sleep progresses.

The stages of sleep are here suggested to correspond to levels of connectedness between subdivisions of the brain. SWS corresponds to the greatest dissociation where small neuron groups function independently of each other. At shallower levels of sleep larger groups of neurons are connected and the waking state is a fully connected state, when sensory and motor neurons are connected.

REM sleep has attracted much interest because of the frequency of recalled dreams upon waking. During REM sleep there is an internal triggering of neuron excitation that may simulate sensory input. If dreams are simply response to simulated sensory information, REM sleep would correspond to an internally connected network weakly connected to sensory and motor neurons.⁹ The often bizarre content of dreams may follow from the triggering that simulates sensory input. However, it has been remarked that higher level critical faculties and a "sense-of-self" are absent from dreams. This is similar to the waking mental functioning of post-lobotomy patients, suggesting that major subdivisions are present in REM sleep.¹⁴ The existence of subdivisions may also explain why most dreams are not remembered, since the connections between neural subdivisions are weakened, and the elements present in each subdivision will be mutually incompatible to the waking brain. For deeper levels of sleep with smaller subdivisions, the waking brain can make no coherent picture of the sleeping mental state.

According to this theory there is a distinct architecture of the brain composed of at least three and up to five levels of subdivision. Sleep is thus a largely genetically programmed mechanism for maintaining a specific subdivided architecture of the brain. In order to be effective, the subdivisions must be consistent between different sleep episodes. Interconnection strengths at different levels are balanced by the duration of dissociation at a particular level. Feedback control may be used to balance the independence and interactions of subdivisions. This control is important to ensure that during waking the brain functions in a coherent fashion so that response is a collective function of the component parts.

Overload and sleep deprivation

While the longer term effects of sleep deprivation directly affect the architecture of the brain, the first manifestation of sleep deprivation should be overload failure. In a subdivided architecture, overload occurs in parts of the network rather than the network as a whole. Key to understanding the phenomenon of overload failure is dynamic locking described earlier -- an inability to shift from an existing state to a properly functioning state when stimulus changes.

Direct experimental verification of this result is found in the well studied 'vigilance' test, one of the tests most sensitive to sleep deprivation.² In this test after a period of similar signals, a slightly different signal requires a very different response. The inability to change behavior is not only of experimental interest but is also understood to be responsible for many early morning accidents. A second example is that of visual illusions such as blurred or wavy objects, e.g. a double floor. This may be understood as due to a locking of parts of the brain responsible for receiving the visual signal causing distortions during movement.

One particularly interesting observation is the presence of so-called 'micro-sleeps' in sleep-deprived individuals.² These consist of short lapses in responsiveness. These may actually not be sleep, but rather temporary suspension of effective brain activity due to locking.

Comparisons have been made (both similarities and differences) between extreme aspects of the behavior of a sleep-deprived individual and a psychotic individual. It is not difficult to believe that disorders in the mechanism of sleep may be responsible for or contribute to psychosis. Experimentally, a correlation has been made between sleep disorders and both schizophrenia and severe depression. The severe lack of SWS sleep in 50% of schizophrenic individuals² may be interpreted as leading to a loss of ability to separate distinct information processing tasks.

While excessive sleep does not lead to a dramatic failure such as neural locking, it may also lead to poor functioning through excessive forgetting and dissociation of parts of the brain. Such excessive dissociation may have both physiologic and psycho-functional manifestations.

Non-Universal Computation

The design of modern computers relies upon models that perform all computational tasks. In contrast, the subdivided architecture of the human mind is a non-universal strategy. There are many possible filters of sensory data and many ways of mapping them on subdivisions of the brain. The subdivisions are connected to each other in varying degrees according to how far apart they are in the sleep determined hierarchy. Weaker connections exist between subdivisions reconnected at shallower levels of sleep from those at deeper levels. It should be apparent that different choices of filters and arrangements will be preferentially suited for different tasks. This non-universal strategy

is consistent with the uniqueness of individuals for which science has found little justification. Sleep controls the architecture rather than the information contained. This suggests a novel approach to the balance of genetic and environmental influence on human behavior.

¹ Webb, W. B., in *Sleep Mechanisms and Functions*, Mayes, A. Ed. (Van Nostrand Reinhold,UK,1983),chap. 1.

² Horne, J., *Why We Sleep*, (Oxford University Press, Oxford, 1988).

³ Fishbein, W. ed., *Sleep, Dreams and Memory*, Advances in Sleep Research Vol. 6 (SP Medical and Scientific, New York, 1981)

⁴ Cohen, D. B., *Sleep and Dreaming: Origins, Nature and Functions*, (Pergamon,Oxford,1979)

⁵ Crick, F. , and Mitchison, G. , Nature 304, 111 (1983)

⁶ Hopfield, J. J. , Feinstein, D. I. , and Palmer, R. G., Nature 304, 158 (1983)

⁷ Geszti, T. , and Pázmándi, F. J. Phys. A20, L1299 (1987); Physica Scripta T25, 152-155 (1989)

⁸ Hobson, J. A., *The Dreaming Brain*, (Basic Books, New York, 1988), p.209-210

⁹ see Scientific American, (September, 1992)

¹⁰ Amit, D. J. *Modeling Brain Function, the world of attractor neural networks*, (Cambridge University Press, Cambridge, 1989)

¹¹ Grossberg, S. and Kuperstein, M., *Neural Dynamics of Adaptive Sensory-Motor Control* (Expanded edition) (Pergamon, New York, 1989)

¹² Attractor networks with symmetric synapses and synchronous updating were used (see Ref. 10). Random unbiased imprints were used and results reflect averages over many samples. A more complete discussion of these and related results will be given elsewhere.

¹³ Mezard, M., Parisi, G., and Virasoro, M. A., *Spin Glass Theory and Beyond*, (World Scientific, Singapore, 1987)

¹⁴ Hartmann, E. L., *The Functions of Sleep*, (Yale University Press, New Haven, 1973) pp. 136-8

Figure Captions:

Figure 1: Illustration of subdivided network concept using natural language example. A fully connected network with enough neurons to store exactly nine sentences shown on the left may be imprinted with and recognize these sentences. If the network is divided into three parts it may be imprinted with only three sentences (center). However, because each subnetwork functions independently, all possible twenty-seven combinations of words shown to the right are recognized. Comparing left and right columns suggests the difference between content and grammar in sentence construction.

Figure 2: Illustration of subdivided network concept using a simulation of Hebbian imprinting. The horizontal axis is the number of imprinted states. The vertical axis is the number of recalled states (note logarithmic scale). Networks tested contained 100 neurons partially subdivided into four equal subdivisions. Distinct curves are for different degrees of subdivision resulting from weakening synapses interconnecting between subdivisions by the factor indicated. The curve marked "0" is for a completely subdivided network while the curve marked "1" is for a fully connected network. The recalled states are all fixed points of the network dynamics composed of combinations of substates imprinted on each subnetwork. As in Fig. 1 subdivision causes more states to be recognized from fewer imprints.

Figure 3: Illustration of the effectiveness of a relearning procedure in causing a selective reinforcement and forgetting of prior imprints. The figure shows the number of imprinted states that have basins-of-attraction larger than the horizontal axis value. Four cases are shown. All cases start with eight imprinted states on a network with 100 neurons shown as (a). (b) shows the result of reimprinting four of the original eight states causing four to be strengthened (larger basins-of-attraction) at the expense of weakening the others. (c) shows the result of starting from a random neural state evolving the neural state and imprinting the result on the network. (d) shows the same as (c) with the addition of a measured (optimized) amount of noise during evolution. In (c) approximately 1.5 of the original states are reinforced while in (d) 2 are reinforced.

Big	Bob	ran.		Big	Bob	ran.	Big	Bob	ran.
Kind	John	ate.		Kind	John	ate.	Big	Bob	ate.
Tall	Susan	fell.		Tall	Susan	fell.	Big	Bob	fell.
Bad	Sam	sat.					Big	John	ran.
Sad	Pat	went.					Big	John	ate.
Small	Tom	jumped.					Big	John	fell.
Happy	Nate	gave.					Big	Susan	ran.
Mad	Dave	took.					Big	Susan	ate.
							Big	Susan	fell.
							Kind	Bob	ran.
							Kind	Bob	ate.
							Kind	Bob	fell.
							Kind	John	ran.
							Kind	John	ate.
							Kind	John	fell.
							Kind	Susan	ran.
							Kind	Susan	ate.
							Kind	Susan	fell.
							Tall	Bob	ran.
							Tall	Bob	ate.
							Tall	Bob	fell.
							Tall	John	ran.
							Tall	John	ate.
							Tall	John	fell.
							Tall	Susan	ran.
							Tall	Susan	ate.
							Tall	Susan	fell.

Figure 1

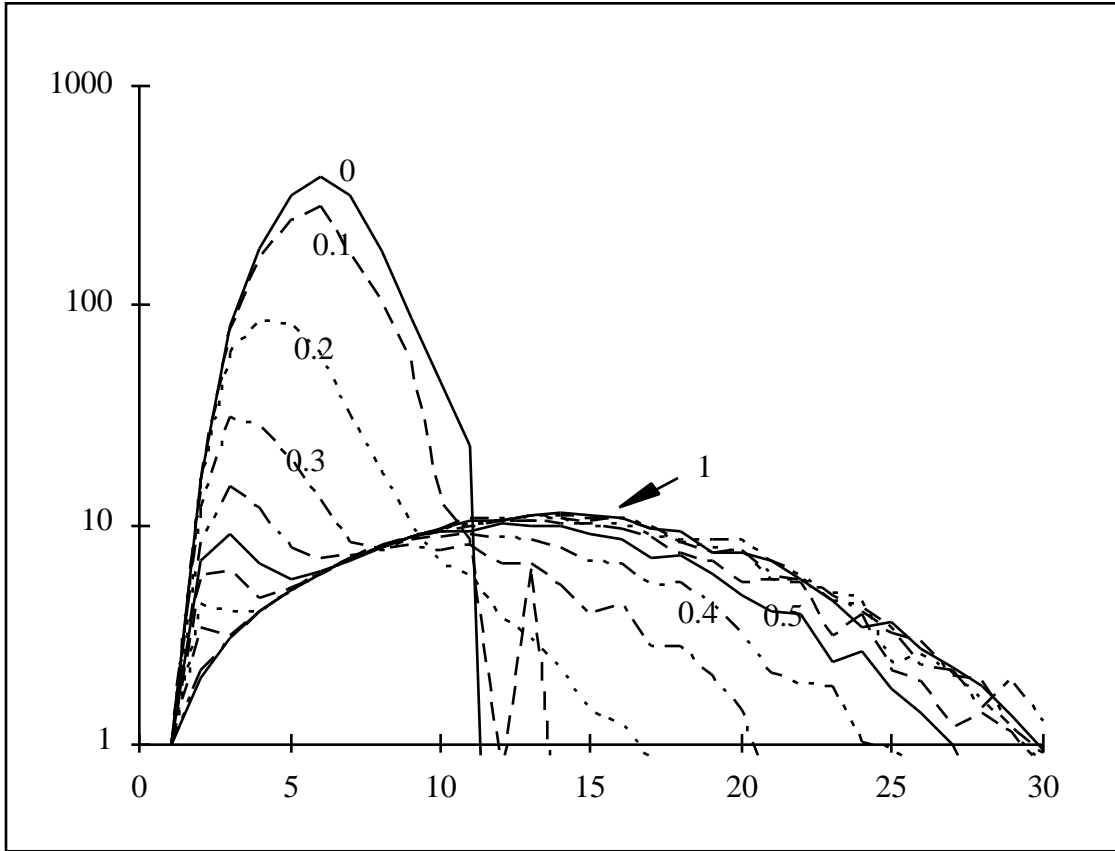


Figure 2

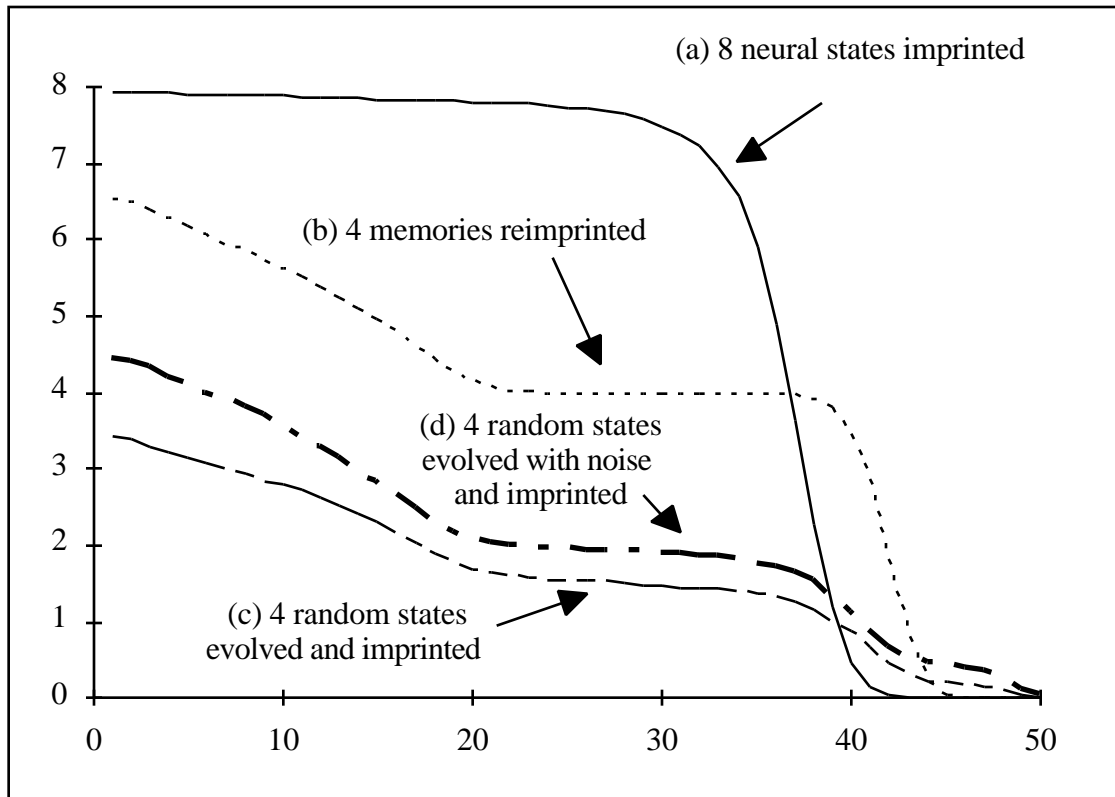


Figure 3