

A Sequence Predicting CA3 Is a Flexible Associator That Learns and Uses Context to Solve Hippocampal-Like Tasks

William B Levy

*Departments of Neurological Surgery and of Psychology,
University of Virginia Health Sciences Center,
Charlottesville, Virginia*

ABSTRACT: The model discussed in this paper is, by hypothesis, a minimal, biologically plausible model of hippocampal region CA3. Because cognitive mapping can be viewed as a sequence prediction problem, we qualify this model as a successful sequence predictor. Since the model solves problems which require the use of context, the model is also able to learn and use context. The model also solves configural learning problems of which, at least one, requires a hippocampus. Thus, by solving sequence problems, by solving configural learning problems, and by creating codes for context, this model provides a computational unification of hippocampal functions which are often viewed as disparate.

© 1997 Wiley-Liss, Inc.

KEY WORDS: goal finding, transitive inference, configural learning, cognitive mapping, problem solving, hippocampus

INTRODUCTION

We are studying biologically plausible, neural-like network models of the hippocampus. Our goal is to create and understand the simplest such model that demonstrates the cognitive/behavioral abilities ascribed to the hippocampus. By definition, the simplest network is one with minimal properties, and, for us, all of these properties must be biological. As we begin to understand this minimal model of the hippocampus, we hope to progress to models in which the neurons correspond to biological neurons in ever greater detail. We expect that enhancing the biological "reality" of the neural-like networks will allow us to predict actual neuronal firing patterns. For now, however, we only seek the insights provided by minimal models. Thus, we are presently trying to understand how the connectivity of the hippocampus and its well-known physiology contribute to the roles played by the mammalian hippocampus in learning and memory.

Our candidate minimal models are tested against behavioral paradigms that reflect contemporary cognitive/behavioral theories of the hippocampus: the declarative memory system described by Cohen and Eichenbaum (1993); the context learning systems described by Hirsh (1974), Kesner and Hardy (1983), and Gray (1982); the configural learning system as de-

scribed in a revised theory by Rudy and Sutherland (1995); and the cognitive mapping theory described by O'Keefe and Nadel (1978). Although our work is not yet complete, only a small set of hippocampal properties seems necessary for the minimal model to solve several tasks that are, arguably, paradigmatic for these theories of hippocampal function.

THE MODEL

Gross Features of the Model

We initially proposed (Levy, 1989; Levy et al., 1990) and studied a model with pieces corresponding to hippocampal regions CA1 and CA3. Hippocampal region CA3 and the dentate gyrus were, and are, viewed as recoders of cerebral cortical representations; CA1 (working in concert with the subiculum and entorhinal cortex, I would say today) is viewed as the decoder (see also McClelland and Goddard, 1996; Hasselmo et al., 1996 both this issue). This decoder mediates the adjustments of cerebral cortical representations as directed by the adaptively formed CA3 codes. However even now, modeling such a decoder requires many arbitrary decisions as to the physiologies of CA1, subiculum, and entorhinal cortex. There is always the temptation to fix up a poorly performing CA3 with a very cleverly designed, but biologically questionable, decoder. Moreover, it is the CA3 portion of the model that is really solving the problems described here. As a result, viewing the network output through the CA1 decoder clouds our understanding of how the CA3 model actually solves problems. Thus, we put the CA1-subiculum-entorhinal cortex (EC) computation to one side and just concentrate on studies of the CA3 model (see Fig. 1a).

Instead of constructing a CA1-subiculum-EC neural decoder, we interpret the activity patterns of CA3 using the simplest decoder that we can think of: We compare a CA3 state vector (CA3's activity at one time step during testing) with all the network-created code words (a

Accepted for publication July 15, 1996.

Address correspondence and reprint requests to Dr. William B Levy, Department of Neurosurgery, University of Virginia Health Sciences Center, Box 420, Charlottesville, VA 22908.

code word is a state vector elicited on the last trial of training); the cosine of the angle between a test-elicited output and each driven state vector at the end of learning quantifies the mathematical comparisons. For each successive output during testing, the input pattern that drives the network state producing this maximal cosine value is identified as the decoded output of the network. Thus, the model—in a purely self-supervised fashion—creates a set of code words (firing patterns) during the learning trials in which each code word corresponds to a different input pattern of a sequence.

In terms of the input to the CA3 model, we have functionally combined into a single input what is in fact the dual input of the dentate gyrus and the layer II entorhinal input to CA3. This seems sensible in order to keep the model simple and because the dentate itself is activated by the same layer II inputs that excite CA3 directly. Moreover, we, like others (e.g., Treves and Rolls, 1991), are willing to hypothesize that the sparse connectivity from the dentate to CA3, when combined with the entorhinal input to CA3, leads to extremely sparse but powerful driving of a few CA3 neurons.

Microscopic Features and Functional Characteristics

Table 1 summarizes the fundamental properties of the network with Table 3, making the implementation of these properties precise.

The fundamental microscopic properties—the anatomy of the network and the physiology of the neurons and synapses—were selected based on three ideas: simplicity (e.g., McCulloch-Pitts neurons); observed physiology (e.g., the direction of temporal offset of the associative modification rule, the sparse connectivity of the network, and the low activity levels of the network); and a third idea that might not be so obvious. We wanted to show, in an unambiguous fashion, the critical microscopic properties that allow the network to create codes that are suitable for solving dif-

ficult sequence prediction problems. Therefore, we denied the network certain biologically realistic microscopic properties that should, in fact, be very useful for solving sequence prediction problems (e.g., a long-time-spanning associative modification rule and capacitative neurons). In particular, we wanted to show that the long-time-spanning properties of the system arise as much from the sparse but powerful recurrent circuitry as from the time spans of neural computations and *N*-methyl-D-aspartate (NMDA) receptor events.

Randomness in the naive, inexperienced network was an explicit goal of the network design. We were impressed by the advantages of maximum entropy-based prediction (Levy, 1985, 1989; Levy and Delic, 1993). This philosophy, when combined with the view of the hippocampus as the associator of last resort, implied minimal a priori coding biases as desirable and leading to enhanced network performance (Levy, 1989). Thus, properties 4–7 of Table 1 are seen as enhancing the tendency of the naive network to produce random patterns. And there is another important motivation for random connectivity.

The sparse connectivity of the network (see Fig. 1b) was seen as fundamental for producing sequence codes. Symmetric networks (i.e., the condition where reciprocal synapses are equal, $w_{ij} = w_{ji}$) will tend to form stable attractors (Cohen and Grossberg, 1983; Hopfield, 1982). Because we wanted a dynamic code to reflect the dynamic changes of the input, we saw great advantage in using the sparse connectivity of the actual hippocampus as a design for our network. In particular, randomly made connections in such a sparse network lead to very few reciprocal connections. With few reciprocal connections, a CA3-like model will be largely asymmetric. Therefore, this connectivity biases the network to produce sequences (in contrast to high connectivity and symmetry, which tends to produce stable attractors).

The production of such sequences was confirmed in our early studies (Minai and Levy, 1993a,b, 1994). In these studies we learned how to set the parameters of the network to produce random sequences and to avoid cyclic behavior (oscillations). We also discovered that there are certain technical problems working with such small networks in terms of the variance in synaptic connectivity and in terms of sustainable activity levels. As a result, we enforced a rule in the construction of these networks in which the random connections are made such that each neuron receives its full share of connections and we avoid activity levels that are too low. Presumably, with larger networks and more connections, it would not be necessary to use such an ad hoc rule, and we could go to much lower activity levels.

So that the network could create its own codes in a manner as flexible as possible, we used a local associative synaptic modification rule (Levy, 1982), we used inputs that excite many fewer neurons than recurrent excitation, and we allowed feedback inhibition to regulate activity levels in a rather poor way. That is, a divisive shunting inhibition is appropriately delayed; as a result activity levels are constantly fluctuating (Minai and Levy, 1993a,b), unlike pure competitive networks.

The variables in this minimal model are summarized in Table 3. Recently we have been using 512 primary neurons in these networks and one interneuron. The linear interneuron combines feedforward and feedback excitation to produce a shunting inhi-

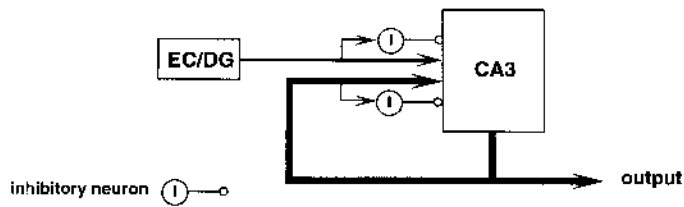
TABLE 1.

Fundamental Network Properties*

1. Neurons are McCulloch-Pitts-type threshold elements (one time step delay, no memory, and excitation is a weighted sum of the input)
2. Synapses modify associatively based on a local Hebbian-type rule and have a time-spanning, associative capability that, at least, equals the neuronal delay
3. Most connections are excitatory
4. Recurrent excitation is sparse and randomly connected
5. Inhibitory neurons approximately control net levels of activity
6. Recurrent excitation more strongly contributes to network activity than do external inputs
7. Activity is low but not too low

*Property 1 derives from a desire for simplicity. Our knowledge of the hippocampus inspires properties 2–7. Properties 4–7 add to the a priori randomness of the network and its sequence coding tendencies.

a. Simplified Hippocampal Model



b. Sparse, Random Recurrent Excitation in CA3

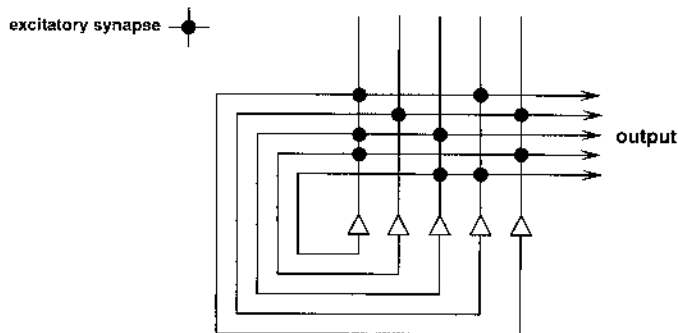


FIGURE 1. The basic model. a: The model consists of an input layer which should be viewed as a combination of the entorhinal cortex and dentate gyrus and which is modeled as an excitation that sparsely activates CA3. Accompanying the feedforward excitation is a proportional feedforward inhibition. The strong input to the network is the recurrent excitation which is also accompanied by a feedback inhibition. The output of the network is the state of the excitatory CA3 cells themselves. b: The recurrent excitatory synapses of CA3 are modeled as being sparse and randomized.

bition [(Furman, 1965; Grossberg, 1976; Rose, 1977); see the denominator of somato-dendritic excitation/inhibition equation of Table 3].

We have tried connectivities from 5% to 20%; the bulk of the work uses a 10% probability of interconnect from one neuron and another neuron. This, by some approximations, might be the degree of interconnection that would occur in a narrow strip of CA3 perhaps .5 mm wide in the septotemporal direction.

Naturally, local associative modification is used (e.g., Levy and Steward, 1979). More than one equation has been tried (Minai and Levy, 1993c), but the bulk of the work uses the postsynaptic rule shown in Table 3.

The time offset of synaptic association (presynaptic activity precedes postsynaptic activity) is qualitatively consistent with experimental observations (Levy and Steward, 1983; Gustafsson and Wigstrom, 1986) and biophysical models (Holmes and Levy, 1990). However, it is arguably not of the appropriate longevity; it all depends on the analogical meaning of one time step. (If a time step is a theta cycle, then the time span of associativity is about right; if a time step is a neuron's RC time constant, then the associative span is too short.) As already mentioned, the rea-

son for choosing such a short time span (i.e., one time step of the simulation) is to show the importance of the neuronal circuitry in creating appropriate codes for solving time-spanning problems that exceed the time span of the synaptic modification rule.

The rate constant (ϵ) of synaptic modification was set to a large value for one trial learning, while in all other experiments the values of .01-.05 were used. Presumably, such a rate constant is controllable by arousal levels in a behaving animal.

In setting up these networks, there are three functional variables (Table 2) to which we pay special attention: activity—both its average value and how much it fluctuates—sequence length memory capacity, and firing patterns called local context neurons (to be described later). In fact, these networks are rather difficult to parameterize. Getting the desired activity levels, memory capacity, and firing patterns takes a great deal of testing and adjusting of parameters.

The desired activity levels that we are able to achieve varies as a function of the number of neurons. As we put more neurons in the network, we can go to lower and lower levels of activity. Currently with 512 neurons, the network typically runs at activity levels between 6% and 15% of neurons firing per time step.

The threshold for cell firing (θ) and the size of the inhibitory constants (K_I and K_R) are the variables most often adjusted to produce the desired activity levels that avoid severe oscillations of

TABLE 2. *Functional and Actual Variables in Our Network Model**

| Variable | Typical desired form |
|---|----------------------|
| A. Functional variables | |
| a. Average activity | Low |
| b. Activity fluctuations | Low |
| c. Sequence length memory capacity | Large |
| d. Average lifetime of local context neurons | Large |
| e. Speed of learning | Fast |
| B. Actual variables of the minimal model | |
| 1. Number of neurons | |
| 2. Percent connectivity | |
| 3. Time span of synaptic associations | |
| 4. Threshold to fire | |
| 5. Feedback inhibition weight constant | |
| 6. Feedforward inhibition weight constant | |
| 7. Rate constant of synaptic modification and number of learning trials | |
| 8. The input code | |

*The functional variables (A) of the network are controlled by the actual variables (B). The number of neurons (1) depends on how much computer power is available; more is better. Connectivity (2) and time span (3) have not been systematically investigated, but small changes have not significantly altered the network's performance. Speed of learning (e), is controlled by actual variable (7). The remaining actual variables 3-6 and 8 control functional variables a through d.

TABLE 3.

*Summary of Network Computations**

Somato-dendritic excitation/inhibition

$$y_j(t) = \frac{\sum_i w_{ij} c_{ij} z_i(t-1)}{\sum_i w_{ij} c_{ij} z_i(t-1) + K_I \sum_i x_i(t) + K_R \sum_i z_i(t-1)}$$

Output

$$z_j(t) = \begin{cases} 1 & \text{if } y_j(t) \geq \theta \text{ or if } x_j(t) = 1; \\ 0 & \text{otherwise} \end{cases}$$

Typical synaptic modification rule

$$w_{ij}(t) - w_{ij}(t-1) + \epsilon z_j(t) [z_i(t-1) - w_{ij}(t-1)]$$

*The variables of the network and the equations relating these variables. Inhibition is incorporated as a divisive, shunting form.

Definitions

y_j = net excitation of j ; x_j = external input to j ; z_i = recurrent input i ; z_j = output to neuron j ; K_I = feedforward inhibition scale constant; K_R = feedback inhibition scale constant; $C_{ij} = \{0,1\}$ variable of connectivity, kept constant once selected; w_{ij} = the excitatory weight, neuron i to j ; ϵ = rate constant of synaptic modification; t = time.

average activity and that allow the network to spontaneously generate activity even in the absence of input.

Why Is This a CA3 Model?

In some sense this recurrent excitatory model applies to many subregions of the hippocampus and of the cerebral cortex; this is not a problem because many of these areas might be viewed as similar from a computational perspective. However, to be very speculative, it is the quantitative combination of two features that distinguishes a CA3 model from models of other regions: 1) the relatively broader, sparser dispersal of recurrent random connections, and to a larger extent than elsewhere; 2) the determination of neuronal firing by recurrent activation much more so than by the external input.

THE PROBLEMS SOLVED TO DATE

A hallmark of our approach is the creation of toy problems which serve as paradigmatic abstractions for real world problems that the hippocampus helps solve. The following problems, listed in Table 4, are solved by the network.

Spontaneous Rebroadcast

Inherent in the idea of the hippocampus serving as a temporary, intermediate storage depot of episodic memory is the requirement for a mechanism by which the hippocampus would teach the cerebral cortex. Spontaneous rebroadcast of a learned sequence is the mechanism that is most obvious to us.

The idea of one portion of cortex teaching another portion of cortex during sleep has been around for over 30 years (for early citations, see Hennevin et al., 1995). More recently the results of Pavlides and Winson (1989) and the results of Wilson and McNaughton (1994) provide strong evidence that the hippocampus would teach the cerebral cortex during sleep. Similarly, Chrobak and Buzsáki (1994) have observed sharp wave-associated cell firing in the hippocampus, and they suggest that this firing is the process by which the hippocampus teaches the entorhinal cortex.

Spontaneous rebroadcast of a learned sequence occurs when the network is allowed to randomly generate patterns. Indeed, this was one of our first results (Minai and Levy, 1993c) which we are continuing to follow up with more detailed biological models (August and Levy, in preparation).

One-Trial Learning

Episodic learning is often a one-trial affair. When the rate of synaptic modification is made large enough (e.g., 0.8), then the system will learn short sequences in a single trial (Minai and Levy, 1993c).

Simple Sequence Completion

The simple sequence completion problem is the easiest test of a network to see if it has learned a sequence. This problem is analogous to the pattern completion problem, but the network's output must be temporally ordered as appropriate to the input used for testing. Most simply, we test by giving the first pattern of a sequence and then look for the rest of the patterns in the proper order.

For example, consider the 12-pattern sequence 1 of Figure 2b. Let the network be given learning trials for just the sequence of patterns ABC α BC β GHJKL. Then, give it just pattern A and see what happens. Perfect sequence completion would see produc-

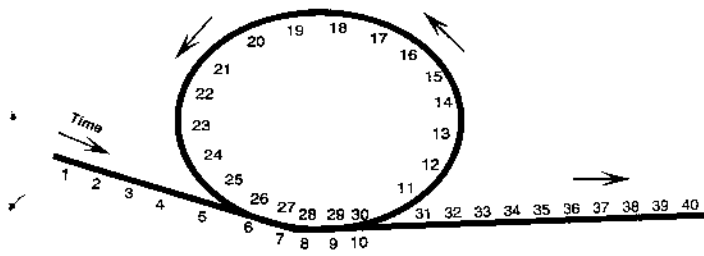
TABLE 4.

*Ten Problems Solved by Our Minimal Network Model**

1. Simple sequence completion
2. Spontaneous rebroadcast
3. One-trial learning
4. Jump-ahead recall (faster than real time)
5. Sequence completion with an ambiguous subsequence (context past)
6. Goal finding without search (context future)
7. Piecing together appropriate subsequences
8. Transverse patterning (context present)
9. Finding a short cut
10. Transitive inference

*We designed input codes that correspond to these ten problems, and the CA3 model solved these problems. Transverse patterning is a configural learning problem as are transitive inference (9) and one version of goal finding (6). Problems 5, 6, and 7 make obvious use of context, but, as indicated, these are different kinds of context.

a.



b.

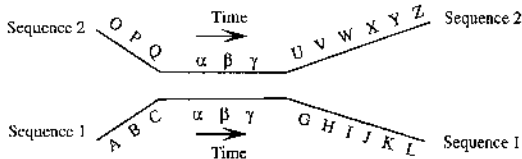


FIGURE 2. Sequences to be learned in two environments. **a:** The looping path input sequence. This sequence of 40 patterns has an identical pair of subsequences (6–10 and 26–30). Learning trials consist of the complete 40-pattern sequence and then resetting the network with a noisy input. **b:** Two overlapping sequences can be learned individually. The two 12-pattern sequences share a common subsequence of three patterns (α, β, γ). During learning, these two sequences were randomly presented with a noisy pattern between the end of one trial and the beginning of the next one.

tion of the remaining 11 patterns in the sequence. The network has shown such behavior under many circumstances (e.g. Minai and Levy, 1993c).

Jump-Ahead Recall

Sometimes (e.g., in cognitive mapping), it might be useful to picture a goal at the end of a sequence while still at the start position. By adjusting the inhibition in the network, sequence completion in response to a test input can be replaced by jump-ahead recall that either skips through the sequence missing many patterns, moves quickly through the sequence, or just skips to the end of the sequence (August and Levy, 1996; Prepscius and Levy, 1994).

Finding a Short Cut

Abbott and Blum (1996) showed that their model of the hippocampus could gradually shorten its path from start to goal. Using the looping path problem (Fig. 2a), our network shows an ability to predict a short cut, but not by virtue of gradual learning (Levy et al., 1995). As in the picture of the looping sequence (Fig. 2a), the training input is a 40-pattern sequence that runs back over itself. When a pattern completion test is given (pattern 1 is the input and then nothing), the network produces a sequence that avoids the loop. That is, the network creates a sequence that goes through the overlapping region and then after a slight hesitation runs out the tail to pattern 40.

Subsequence Disambiguation

Using a nonlocal learning rule and a rather overconnected network, Fukushima (1973) showed that there was a neurally inspired network that could solve the sequence disambiguation problem. This class (problem 5 of Table 4) of sequence completion problems is probably the most important problem for any sequence prediction network to solve. One such problem solved by our model is pictured in Figure 2b: Here there are two sequences to be learned. Each sequence is made up of 12 patterns in which each different pattern is represented by a different letter of the alphabet. A difficulty arises in the two sequence completion tests because there is a shared subsequence of length three (α, β, γ), which is part of both sequence 1 and sequence 2. It is such a shared subsequence that creates the ambiguity of the disambiguation problem. Because neural networks represent patterns sequentially, the disambiguation problem can only be solved if there are different representations for the shared subsequence. To understand the network's difficulties, look at Figure 2b and consider this: If I ask you where you should end up if you start at A, you will say L. And if you start at O, you know the learned sequence's end point is Z. But if I say that you are starting at γ , you should not say anything or you should toss a coin to guess the sequence's end point because there is no right answer. Thus, the sequence completion problem can only be solved by using the context produced by the recent past. Note that this context of the past, which the network does use to solve this problem (Minai et al., 1994; Levy et al., 1995; Wu et al., 1996), arises from more than one time step back in to the past. So the network itself (not just the one time step-spanning associative modification rule) must be producing a code that spans this time gap.

Goal Finding Without Search

To me, a useful property in cognitive mapping is the ability to imagine my way to a goal before actually taking the route that seems the right one. That is, you might have to make a decision quite far away from the goal—to turn left or to turn right—so you would need to know where your left turn or your right turn would end up taking you relative to the possible goals which are much further down the line. The network has successfully solved two versions of this goal-finding problem. One version uses the looping path problem that demonstrates short cut finding, and the other version uses the sequence disambiguation problem with the shared subsequence.

As you recall in the looping path problem (Fig. 2a), it is the natural tendency of the network to go along the beginning, subsequence 1–5, then to carry through sequence 6–10 and recognizing, in some sense, the identity or similarity between sequences 6–10 and 26–30, continuing on into the tail of the sequence to reach pattern 40, the last pattern of the sequence. All the while, the network avoids the looping subsequence of patterns 11–29 that was part of the original learning. This behavior, however, can be changed by giving the network a hint about another goal (Levy et al., 1995).

In both of the goal-finding problems, we view the goal as some type of physical object such as water or food, and we assume that

this goal was present during learning. Therefore, this physical object would be part of the input code given to the network. If we are willing to grant that a thirsty rat can think of water, then we are allowed to input a generic code for water both during learning and during testing. During learning, we would pretend that water would be present somewhere in the loop, for instance at position 21 (which is deep inside the loop). We would use four of the 512 neurons to represent the presence of water at 21 and four other neurons to represent the rest of 21 as an external code. Now when a network trained on the overlapping loop is tested, that is, when pattern 1 is presented during testing, the four neurons representing water are turned on and left on throughout the entire testing period (as if the network is thirsty throughout testing). In this case the network no longer traces the short-cut path to pattern 40. Rather, the path the network traces leads to position 21, and here, or in the near vicinity, it remains. The other goal-finding problem is described in the next section.

Piecing Together Subsequences

The sequence disambiguation problem pictured in Figure 2b can be turned into a much more interesting problem. Taking the same network that has learned how to do sequence disambiguation, we now ask the network to do goal-finding where the goal contradicts the sequence disambiguation results. For example, in the sequence disambiguation problem, the network is given only pattern A, and the correct answer is a sequence that arrives at L as was the case during learning, whereas in the piecing-together/goal-finding problem, the network again gets pattern A as an input, but the correct answer is a sequence leading to Z. Once again, the goal is incompletely defined by turning on a small number of externally activated neurons associated with the goal and then allowing the network to develop a sequence of patterns after being given the first pattern. When a strong enough input is given signifying the goal, typically four out of 512 neurons (Levy and Wu, in preparation; Wu and Levy, 1996), then the network is able to create the appropriate path from the beginning of one sequence to the end of another sequence. As shown in Figure 2b, this path (e.g., ABC α β γ UVWXYZ), which was never experienced by a network during learning, is created by appropriately sewing together subsequences from the two separately learned sequences. Such a path is the appropriate answer to the goal-finding problem.

Transverse Patterning

Transverse patterning (Alvarado and Rudy, 1992) is a configural learning problem in which the meaning of each of three atomic stimuli (e.g., A, B, C) depends on the pairing of these individual stimuli with each other. Thus, the transverse-patterning problem uses context coming from the configuration of the atomic stimuli. This is a context of the present in contrast to earlier problems in which context came from the past (disambiguation) or context comes from the future (goal finding). As shown by Alvarado and Rudy, such problems can be solved by rats when the proper learning procedure is used (Alvarado and Rudy, 1992). Also shown by them is the necessity of a hippocampus to learn the correct re-

sponses in this problem. The stimuli to be learned and their reinforcement are AB where A is the correct answer, BC where B is the correct answer, and AC where C is the correct answer.

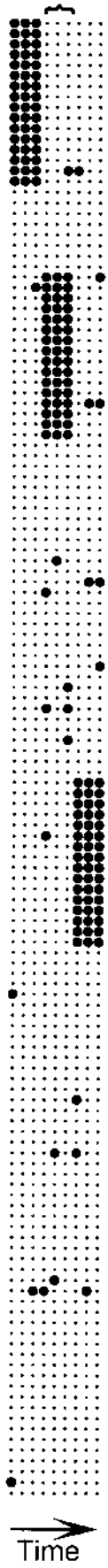
Although not typically viewed as such, such a configural learning task embodies sequence learning, and one such input sequence is shown in Figure 3a-d. Specifically, the sequence is 1) the configured stimulus complex (e.g., AB) followed by 2) a motor decision/response which selects one of the stimuli (e.g., choose A) followed by 3) a reinforcement, (e.g., positive). Once viewed as a sequence, such problems are amenable to solution by our sequence learning model. In fact, if an appropriately coded input is given (as in Fig. 3), the CA3 model can solve the transverse patterning problem (Levy et al., 1996; Wu et al., in preparation). In obtaining this solution, we have again used the method of goal finding by partial designation; this time the goal is the positive reinforcement (neurons 73-88 of Fig. 3). Decoding is produced by comparing the network's output at time periods 4, 5, and 6 to the code for each of the three possible decisions/responses.

Transitive Inference

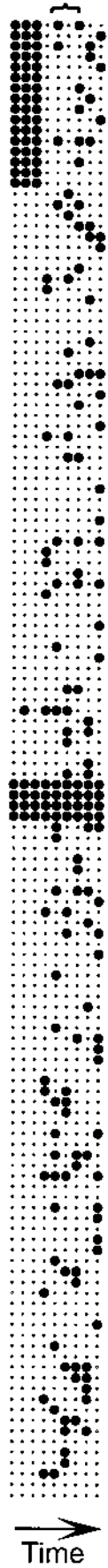
Transitive inference is another configural problem. Again, the atomic stimuli (in this case A, B, C, D, and E) change from being the correct or the incorrect answer as a function of the stimulus with which each one is paired. The learning set in transitive inference consists of four pairs: AB, with A the correct answer; BC, with B the correct answer; CD, with C the correct answer; DE, with D the correct answer. The test for transitive inference is presenting the animal with B and D where the correct answer is B (picture the mathematical relationship $A > B$, $B > C$, $C > D$, $D > E$ then $B > D$). Indeed, this problem proves easier for the network to learn than the transverse patterning problem when using the exact same encoding strategy (Wu and Levy, in preparation).

FIGURE 3. Some neuronal firing patterns in the transverse patterning problem. This figure shows how part of the transverse patterning problem is coded and how learning alters firing patterns. These four vertical strips show a subset of the CA3 neurons (140 out of 512). (The remaining 371 neurons are local context neurons if they fire—not shown due to lack of space.) In each strip, there are nine time steps. Neurons 1-16 mediate the external activation of the AB configuration. Neurons 25-40 mediate the decision/motor response that accompanies the selection of pattern A. Neurons 73-88 mediate the positive reinforcement that is given when the correct atomic stimuli is selected (in this case A of the AB). A comparison of a, i.e., before learning, to c, after learning, shows the altered firing due to learning. Note the additional firings in c, particularly neurons 89 and up. These are the local context firings. By comparing b, i.e., before learning, and d, i.e., after learning, the reader can ascertain the new patterning of firing that occurs during testing. The network is tested by activating the AB input configuration and by activating one-quarter of the positive reinforcement configuration (neurons 73-76). The reader should be able to discern that over time steps 4, 5, and 6 (horizontal brace) the firing patterns in d are more similar to the firing patterns at the same time in c than are patterns 4, 5, and 6 of b compared to c. In performing the decodings, the firings of d over time steps 4, 5, and 6 are compared to the learned neuronal codes of all three possible decisions.

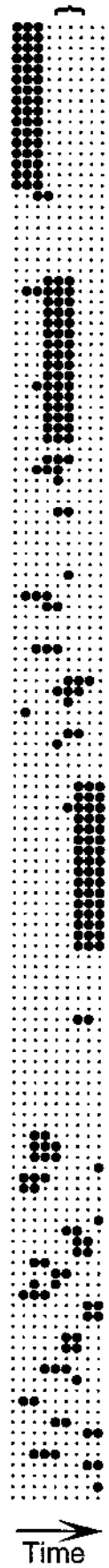
a. Neurons 1-140 (141-512 not shown)



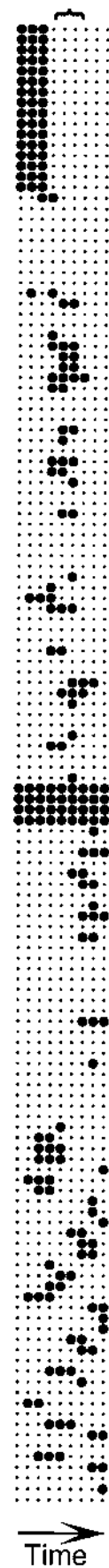
b. Neurons 1-140 (141-512 not shown)



c. Neurons 1-140 (141-512 not shown)



d. Neurons 1-140 (141-512 not shown)



What Other Problems Can the Network Solve?

Can the model solve the water maze problem? Can the model do trace conditioning? It seems likely, but these problems are as yet untried.

WHY THE MODEL WORKS: CODINGS FOR LOCAL CONTEXT AND THE CREATION OF ATTRACTORS

The key to most network success is the formation of local context neurons. These patterns of neuronal firing by individual neurons are best understood by referring to the picture of Figure 4. Here a very simple, slowly shifting sequence of 40 patterns is the input. Each input activates eight neurons; the first pattern activates neurons 1–8; the 40th pattern activates neurons 40–47. On the left is the first learning trial. Note the firing of the neurons not activated by the external inputs (neurons 48 and up), and, in particular, note their somewhat random firing. After a suitable number of learning trials, the firing in the network (shown on the right of Fig. 4) is quite different. Whereas, before learning, neurons might fire once and then fire again some time later, we now can see that neurons have very strict patterns of firing, in which firing does not happen for a period of time, then firing is continuous for a period of time, and then firing ceases for the rest of the sequence.

Why do such firing patterns appear? Most important is the time-spanning associative modification rule. This rule associates neurons across time so repetitive firing is part of the learning repertoire of the neurons. In addition, other variables can be critical, e.g., activity levels and the input code (Levy and Wu, 1996; Wu et al., 1996).

In effect, individual neurons have learned to recognize particular subsequences of the input sequence. In this way, these local context firings are analogs of place cells. That is, a place cell "recognizes" a smaller area within a large area, whereas a local context cell "recognizes" a subsequence of larger sequence.

These local context neurons should be viewed as pattern-recognition devices. Thus, the network has a variety of recognition devices for any one, single pattern, and each neuron might be the recognition device for a slightly different subsequence of patterns. Because these subsequences (i.e., local context neuronal firings) are more or less random, they are interdigitated with one another. As a result of such overlapping and interdigitating local context codes, temporally neighboring patterns and temporally neighboring subsequences will tend to turn on in succession and even simultaneously. By looking at such patterns of neuronal firing and by manipulating them through the variables listed in Table 2B, we conclude that the formation of these local context neurons is critical to the network's ability to solve every problem listed in Table 4, with the exception of simple pattern completion.

But just local context codes are not always enough. Solutions to many of the problems discussed here are correlated with the formation of attractors. For example, in the overlapping loop problems—finding a short cut and finding the goal within the

loop—it is necessary for the network to create an attractor either at the end of the loop or in the vicinity of the goal (Levy et al., 1995; Wu et al., 1996). (Of course, these attractors are not stable point attractors but are more akin to pseudo-chaotic orbits of moderate diameter. That is, once the network is in the region of an attractor, its precise output "randomly" changes without moving very far when looking at network states over a long or short period of time.) Such results—the necessity for local context neurons and attractors—are examples of the advantage of studying simple networks; that is, they allow us to understand the critical codes and the variables necessary for producing successful network performance.

DISCUSSION

What Do We Know Now That We Did Not Know Before?

The discovery of attractors at the ends of sequences was a surprise to us and led to an insight. The network might be able to find goals by being attracted to them. From this followed the demonstration of a network using context future. That is, once we noted that these asymmetric networks could form attractors (although their tendency is not to form such attractors and these are not stable attractors, as already noted above), we were able to hypothesize that the attractors could be moved by virtue of turning on a small number of relevant neurons. This led not only to the demonstration of context future being used for the goal-finding problems, but it also led to versions of the transitive inference and the transverse patterning problems that the network can solve.

Second, although we did know that local context codes could be used to define subsequences (see picture, page 296ff in Levy, 1989), we had no idea how easily they would form. The scheme in Levy (1989) for the formation of local context neurons involving a mismatch detector appears unnecessary. Still the idea that contexts provide the conditioning variable of conditional

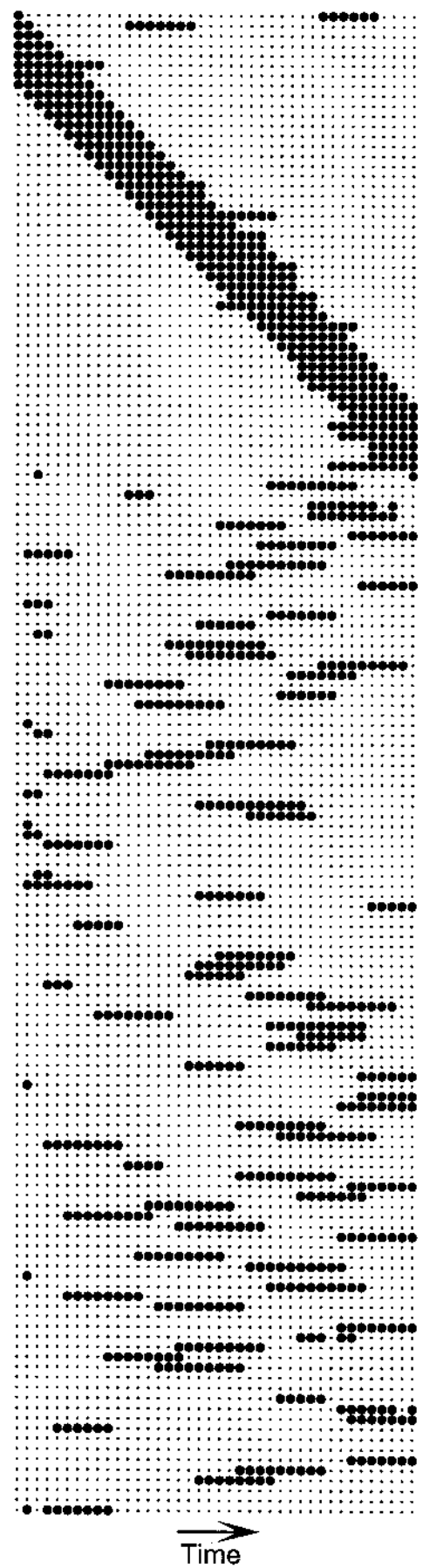
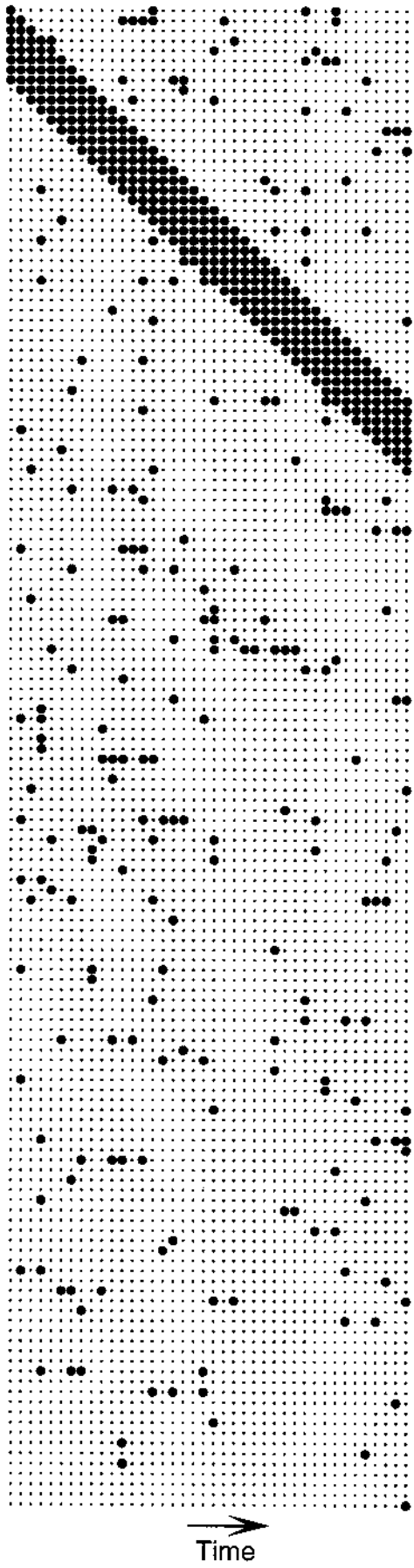
FIGURE 4. Formation of local context neurons. The left- and right-hand panels illustrate a portion of the activity patterns displayed by the network before (left) and after (right) learning. A large dot indicates cell firing, and a small dot indicates nonfiring of a cell. At each successive time step, eight neurons are fired externally. In time step 1, it is neurons 1–8. At time step 2, it is neurons 2–9, and so on until time step 40, when neurons 40–47 are externally activated. In addition to the extrinsically activated neurons, other neurons are fired via recurrent connections. Before learning, recurrently activated firing is somewhat haphazard, although certainly not uniformly random. After learning, however, very special patterns of recurrent firing occur. Individual neurons have learned to identify subsequences of the 40-pattern sequence. A neuron that does fire tends to fire in a consecutive set of firings, with no firings before the beginning of the set or after the end of this set. Such firings are critical to the performance of the network in all the problems described except simple sequence completion. Such local context neuronal firings also occur when the input sequence is composed of more complex patterns that change faster and occasionally make orthogonal jumps between groups of externally active neurons.

CA3 activities before learning

CA3 activities after learning

Neurons 1-150 (Neurons 151-512 not shown)

Neurons 1-150 (Neurons 151-512 not shown)



probabilities connects our current thinking with these earlier ideas.

Finally, the interdigitation of context unit firing was not understood before we did the modeling, and this interdigitation is functionally important. First, interdigitation allows one subsequence to pass its influence on to another subsequence. But there is another, perhaps more fundamental importance of such interdigitated firing patterns when we consider the problem of the hippocampus teaching the cerebral cortex. Such local subsequences that interdigitate lead to a time spanning of patterns that will allow the cerebral cortex to chunk (Miller, 1956) pieces of the sequence. Indeed, when the network is allowed to speed up by lowering inhibition, chunking-like codes become apparent in the network output itself (August and Levy, 1996).

What Did the Model Accomplish That Could Not Have Been Accomplished by Simpler Verbal-Qualitative Reasoning and How Does This Model Relate to Other Noncomputational Theories That Have Been Proposed?

First, the use of local context firings just discussed is not even a part of the cognitive theories so the model adds something immediately.

Second, and more to the point of this section, is the unification of a diversity of theoretical and paradigmatic perspectives.

Table 5 relates some hippocampal theories to the paradigms used here. For us, the disparate nature of the various theories and the different paradigms that have been used to characterize hippocampal function made it difficult to understand their interrelationship.

TABLE 5.

*Summary of the Problems Solved by the Network and Their Relationship to Various Theoretical Perspectives on Hippocampal Function**

| | Intermediate episodic memory store & teach cortex | Sequence learning/ prediction | Cognitive mapping | Learn & use context | Flexible associator |
|-------------------------------|---|----------------------------------|----------------------|------------------------|------------------------|
| Spontaneous rebroadcast | X | x | | | |
| One-trial learning | X | X | | | |
| Simple sequence completion | | X | X | | |
| Jump-ahead recall | | X | X | | |
| Finding a short cut | | X | X | | |
| Goal finding | | X | X | X | |
| Piecing together subsequences | | X | X | X | X |
| Disambiguation | | X | | X | |
| Transverse patterning | | x | | X | X |
| Transitive inference | | x | | | X |

*A large X indicates a straightforward relationship between the paradigm solved by the network and a suggested hippocampal function of a cognitive/behavioral theory. Smaller x's indicate a subtle, or perhaps nonobvious, relationship between sequence prediction and a particular paradigm. The flexible memory theory refers to a subset of the theory advocated by Cohen and Eichenbaum (1993), but not to the entire theory which would presumably claim to encompass all cognitive/behavioral problems within the range of hippocampal function. Likewise, I assume broader claims would come from the modified configural learning theory of Rudy and Sutherland (1995); it could be claimed that a great deal more of these problems are solved via the use of context, but context is explicit and straightforward in only the problems indicated.

The perspective offered by the model provides an interrelationship, if not added harmony, among the theories. That is, the simplest way to understand the interrelationship of the paradigms and theories is via the computational model. To say it another way, the interrelationship grows out of the many characteristics of the unsupervised sequence prediction network described here.

Importance of context

The coding and use of context is one of the most important characteristics of the network. The promiscuous associator of the Cohen-Eichenbaum theory is realized largely by the formation of local context neurons and goal-dependent attractors. The greatest flexibility comes from the use of context (past, present, or future) to distinguish related sequences or to induce the formation of novel sequences. Goal-dependent attractors allow the network to create novel sequences (i.e., piece together subsequences). Of course, context is fundamental to the ideas of Hirsh (1974), Gray (1982), and Kesner and Hardy (1983). Indeed, Kesner understood that sequence learning would be a form of context learning.

What New Experimental Directions Are Suggested by This Modeling, Either (a) To Test Novel Predictions or (b) To Gather More Data Where the Model Indicates That There Is a Need for Such?

The model will not be able to produce predictions with enough details for actual laboratory testing until the model is made more physiological. Generically, however, there are predictions.

One thing we have learned in our modeling is the critical importance of activity levels. Activity levels must not be allowed to fluctuate too drastically or in such a way that would disrupt context codes, and activity levels must be maintained at a moderate level, neither too high or too low. With activity levels too high, there is very little memory capacity in the network (Levy and Wu, 1995; Wu and Levy, 1995; Levy and Wu, 1996), and with activity levels too low, there is no formation of context units. Nature must also worry about the same problems.

We have also learned about the critical nature of the input environment for the network to perform appropriately (Wu et al., 1996; Levy et al., 1996; Wu et al., in preparation). This leads us to speculate about difficult environments: Difficult environments may have to be preprocessed if they are to be solved at all.

Defeating the network

It is possible to find a version of any problem listed in Table 4 that the network cannot solve. That is, inputs can be created that defeat the network by virtue of too much noise, overly long sequences, or sequences that change too drastically and too often. But such failures are not altogether bad. First, the same kind of "extreme" environments will defeat the learning of any mammal asked to solve a similar problem, and thus, such failures are reproducing, at least generically, proper characteristics of cognition. Second, such defeats lead us to hypothesize the nature of the EC-hippocampal signals that lead to success. Specifically, we can predict that the inputs from the EC will not shift too wildly over short periods of time if the hippocampus is to form useful context codes.

In sum, the model was inspired by the idea that hippocampal region CA3 has the anatomy and physiology needed for sequence prediction and that sequence prediction would be useful for solving cognitive mapping problems. The tasks described here are all sequence prediction problems. However, because the themes of the hippocampal function, especially in humans, are not solely concerned with cognitive mapping but also concern declarative memory (flexible associations) and formation of context, we studied paradigmatic examples that would show the learning and use of context and that would also show a flexibility in reconfiguring stimuli as a function of learning context.

Acknowledgments

This work was supported by NIH grants MH48161, MH00622, and EPRI RP8030-08 by Pittsburgh Supercomputing Center grant BNS 950001P to W.B.L., and by the Department of Neurosurgery, Dr. John A. Jane, Chairman. I would like to thank my collaborators on this work who have contributed mightily to these results, including Dr. Xiangbao Wu, Dr. Joanna M. Tyrcha, Dr. Ali A. Minai, Dr. Robert A. Baxter, D.A. August, A.A. Amarasingham, C. Prepscius, and G. Barrows. Special thanks go to Dr. Nancy L. Desmond who provided many suggestions to improve this chapter and to Dr. Xiangbao Wu for his deep, insightful abilities for composing input codes and parameterizing the network.

REFERENCES

- Abbott LF, Blum KI (1996) Functional significance of long-term potentiation for sequence learning and prediction. *Cerebral Cortex* 6:406-416.
- Alvarado MC, Rudy JW (1992) Some properties of configural learning: an investigation of the transverse-patterning problem. *J Exp Psychol [Animal Behav]* 18:145-153.
- August DA, Levy WB (1996) Temporal sequence compression by a hippocampal network model. *INNS World Congress on Neural Networks*, pp 1299-1304.
- Chrobak JJ, Buzsáki G (1994) Selective activation of deep layer (V-VI) retrohippocampal cortical neurons during hippocampal sharp waves in the behaving rat. *J Neurosci* 14:6160-6170.
- Cohen NJ, Eichenbaum H (1993) *Memory, amnesia, and the hippocampal system*. Cambridge: MIT Press.
- Cohen MA, Grossberg S (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans Systems, Man Cybern*, 815-826.
- Fukushima K (1973) A model of associative memory in the brain. *Kybernetik* 12:58-63.
- Furman GG (1965) Comparison of models for subtractive and shunting lateral-inhibition in receptor-neuron fields. *Kybernetik* 2:256-274.
- Gray JA (1982) *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system*. New York: Oxford University Press.
- Grossberg S (1976) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol Cybern* 23:121-134.
- Gustafsson B, Wigstrom H (1986) Hippocampal long-lasting potentiation produced by pairing single volleys and brief conditioning tetani evoked in separate afferents. *J Neurosci* 6:1575-1582.
- Hasselmo M, Wyble B, Wallenstein G (1996) Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus* (This issue.)
- Hennevin E, Hars B, Maho C, Bloch V (1995) Processing of learned information in paradoxical sleep: relevance for memory. *Behav Brain Res* 69:125-135.
- Hirsh R (1974) The hippocampus and contextual retrieval of information from memory. *Behav Biol* 12:421-444.
- Holmes WR, Levy WB (1990) Insights into long-term potentiation from computational models on NMDA receptor-mediated calcium influx and intracellular calcium concentration changes. *J Neurophysiol* 63:1148-1168.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79:2554-2558.
- Kesner RP, Hardy JD (1983) Long-term memory for contextual attributes: dissociation of amygdala and hippocampus. *Behav Brain Res* 8:139-149.
- Levy WB (1982) Associative encoding at synapses. *Proc Fourth Annu Conf Cogn Sci Soc* pp 135-136.
- Levy WB (1985) An information/computation theory of hippocampal function. *Soc Neurosci Abstr* 11:493.
- Levy WB (1989) A computational approach to hippocampal function. In: *Computational models of learning in simple neural systems* (Hawkins RD, Bower GH, eds), pp 243-305. New York: Academic Press.
- Levy WB, Delic H (1993) A generalized theory of maximum entropy prediction by neurons. *INNS World Congress Neural Networks II*: 131-134.
- Levy WB, Steward O (1979) Synapses as associative memory elements in the hippocampal formation. *Brain Res* 175:233-245.
- Levy WB, Steward O (1983) Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience* 8:791-797.

- Levy WB, Wu X (1995) Controlling performance by controlling activity levels in a model of hippocampal region CA3. II. Memory capacity comes at the expense of context cell firing and compressed coding. *INNS World Congress Neural Networks I*:582-586.
- Levy WB, Wu X (1996) The relationship of local context codes to sequence length memory capacity. *Network* 7:371-384.
- Levy WB, Colbert CM, Desmond NL (1990) Elemental adaptive processes of neurons and synapses: a statistical/computational perspective. In: *Neuroscience and connectionist models* (Gluck MA, Rumelhart DE, eds), pp 187-235. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc.
- Levy WB, Wu X, Baxter RA (1995) Unification of hippocampal function via computational/encoding considerations. In: *Proceedings of the third workshop on neural networks: from biology to high energy physics*. *International Journal of Neural Systems*, 6, (Supplementary issue) (Amit DJ, del Giudice P, Denby B, Rolls ET, Treves A, eds) pp 71-80. Singapore: World Scientific Publishing.
- Levy WB, Wu X, Tyrcha JM (1996) Solving the transverse patterning problem by learning context present: a special role for input codes. *INNS World Congress Neural Networks* 96:1305-1309.
- McClelland J, Goddard N (1996) Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus* (This issue.)
- Miller G (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81-97.
- Minai AA, Levy WB (1993a) The dynamics of sparse random networks. *Biol Cybern* 70:177-187.
- Minai AA, Levy WB (1993b) Predicting complex behavior in sparse asymmetric networks. *Neural Information Processing Systems* 5: Natural and Synthetic 556-563.
- Minai AA, Levy WB (1993c) Sequence learning in a single trial. *INNS World Congress Neural Networks II*:505-508.
- Minai AA, Levy WB (1994) Setting the activity level in sparse random networks. *Neural Computation* 6:85-99.
- Minai A, Barrows G, Levy WB (1994) Disambiguation of pattern sequences with recurrent networks. *INNS World Congress Neural Networks IV*:176-181.
- O'Keefe J, Nadel L (1978) *The Hippocampus as a cognitive map*. Oxford: Oxford Univ. Press.
- Pavlidis C, Winson J (1989) Influences of hippocampal place cell firing in the awake state on the activity of those cells during subsequent sleep episodes. *J Neurosci* 9:2907-2918.
- Prepscius C, Levy WB (1994) Sequence prediction and cognitive mapping by a biologically plausible neural network. *INNS World Congress Neural Networks IV*:164-169.
- Rose D (1977) On the arithmetical operation performed by inhibitory synapses onto the neuronal soma. *Exp Brain Res* 28:221-223.
- Rudy JW, Sutherland RJ (1995) Configural association theory and the hippocampal-formation: an appraisal and reconfiguration. *Hippocampus* 5:375-389.
- Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? *Network* 2:371-397.
- Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265:676-679.
- Wu X, Levy WB (1995) Controlling performance by controlling activity levels in a model of hippocampal region CA3. I. Overcoming the effect of noise by adjusting network excitability parameters. *INNS World Congress Neural Networks I*:577-581.
- Wu X, Levy WB (1996) Goal finding in a simple, biologically inspired neural network. *INNS World Congress Neural Networks*: 1279-1282.
- Wu X, Baxter RA, Levy WB (1996) Context codes and the effect of noisy learning on a simplified hippocampal CA3 model. *Biol Cybern* 74:159-165.