

## UNIFICATION OF HIPPOCAMPAL FUNCTION VIA COMPUTATIONAL/ENCODING CONSIDERATIONS.

William B Levy,\* Xiangbao Wu and Robert A. Baxter  
*Department of Neurological Surgery*  
*University of Virginia Health Sciences Center*  
*Charlottesville, Virginia 22908, USA*

This paper discusses a highly simplified, biologically consistent model of hippocampal region CA3. The model is distinguished by its local, self-supervised creation of a code for temporal context. With such a code, the model solves a sequence disambiguation problem. The similarity between this context code and place cell firing is noted. With such a code, the network also solves two other sequence prediction problems: finding shortcuts and seeking goals. Notably, network dynamics rather than a search strategy produces the solutions.

### 1. Introduction and Motivation.

We are studying models of hippocampal function. As a working hypothesis, we suppose that the same computational and encoding processes that allow a rat hippocampus to create cognitive maps are also used to direct long-term memory storage in the human cerebral cortex<sup>1,2</sup>. The unifying principle in these two problems is the learning of context. To us, context is largely an integration of the past up to the present, and it is such integration that essentially defines episodic memory. In this hypothesis, context defines where memories are stored in the cerebral cortex<sup>3,4,5</sup>, and context-based encodings provide the key for neural network solutions to the cognitive mapping problem.

Here we shall present some of our most recent data showing how an extremely simplified, but biologically motivated, model of hippocampal region CA3 can learn context and can perform context-dependent predictions that are particularly useful for cognitive mapping.

### 2. Representing a Sequence of Patterns.

One atypical aspect of our model is the definition of a learned representation in a recurrent network. Rather than using stable points as the learned states representing externally generated patterns, we define the instantaneous state of the network as its representation of the input pattern occurring just a moment before. Thus, in sequence problems, it is often unnecessary

---

\*Department of Neurological Surgery, Box 420, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, USA.

and undesirable for the network to have attractors. Rather, we look to the astable dynamics of asymmetric networks to produce representations that are sequential in time.

The other atypical aspect of our approach is that the network must find its own temporal context dependent code without teaching by supervisory signals. That is, context codes emerge as our unsupervised network experiences the sequences of patterns in its environment. This emergence results from the network dynamics – local associative synaptic modification, recurrent activation, and sparse random connectivity. Many neurons in the network receive no external (environmental) inputs yet they develop firing patterns that are regularly associated with a particular subsequence. We refer to these neurons as local context neurons because of their selectivity for different temporal contexts. By using these local context neurons and recurrent activation, the network develops predictive representations of successive patterns in a sequence.

To be useful, successive representations created by the CA3 network require a decoding network that matches the inputs with the new representations created by the network. In our previous work we used such a decoding network<sup>6,7</sup>. Here, however, we do not use an explicit decoder. Rather, we shall allow the reader to act as the decoder: just compare the measured similarity between spontaneously generated patterns in the test situation to the representations that result from a full external sequence. By eliminating the decoder and concentrating only on the CA3 representations, the reader will gain greater insight into the fundamental nature of the codes being created.

### 3. The Computational Model.

The network is a sparsely connected recurrent excitatory system. The connectivity of the network is particularly inspired by the CA3 region of the hippocampus. If we consider a small region (ca. 0.5 mm) along the septotemporal axis of the hippocampus, then we might expect connectivity between neurons in this region to be about five to twenty percent. Here we will use a connectivity of ten percent and randomly connect neurons (if  $i$  connects to  $j$   $c_{ij} = 1$ ;  $c_{ij} = 0$  otherwise) with this probability. Such sparse connectivity leads to asymmetric connectivity. That is, the probability of two neurons being reciprocally connected is one in a hundred. The network's feedback connections between primary cells are all excitatory. We model the network as a deterministic, discrete time, dynamical system. There is one inhibitory interneuron. It feeds back, with one time step delay, inhibition ( $K_R$ ) equally to all neurons in proportion to its summed excitation. Thus, we have again been inspired by our knowledge of the hippocampus where the number of excitatory cells far outnumber the number of inhibitory cells. The input excitation ( $x_j$ ) is a small percentage of the total network activation. In most of the examples here, the external input fires about eight neurons during learning whereas the network, via its recurrent excitation elements, ends up firing 50-80 neurons once learning is finished. A feedforward inhibition is proportional to  $K_I$  and the summed input excitation.

The network elements consist of simple McCulloch-Pitts neurons that sum their inputs to fire ( $z_j = 1$ ) or not fire ( $z_j = 0$ ) in a single time step if threshold  $\theta$  is reached. By definition, such neurons have no memory of inputs from one time step to the next. Inhibition in the network is of the shunting type and is thus the denominator of our excitation equation. The full equation for neuronal excitation is:

$$z_j(t) = \begin{cases} 1, & \text{if } \frac{\sum_i c_{ij} z_j(t-1) w_{ij}}{\sum_i c_{ij} z_j(t-1) w_{ij} + K_R \sum z_i(t-1) + K_I \sum x_i(t)} \geq \theta, \text{ or if } x_j(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

There is a one time step propagation delay between neurons, and there is associative synaptic modification which can span one time step. In particular, an association is made between presynaptic activity at time  $t - 1$  and postsynaptic activity at time  $t$ . We have investigated three different associative modification rules<sup>6,7</sup>, but here we just use the following rule<sup>8</sup>.

$$w_{ij}(t+1) = w_{ij}(t) + \epsilon z_j(t) (z_i(t-1) - w_{ij}(t)) \quad (2)$$

where  $i$  is the presynaptic neuron and  $j$  is the postsynaptic neuron. The rate constant  $\epsilon$  is fixed at 0.02.

#### 4. Previous Results.

Computer simulations of this model shows several properties that are useful for explaining the hypothesized functions of the hippocampus. These properties include:

- (i) Sequence completion from a single pattern input<sup>6</sup>.
- (ii) The ability to learn sequences in one trial<sup>6</sup>.
- (iii) The ability to spontaneously rebroadcast parts of learned sequences<sup>6</sup>.
- (iv) Jump ahead prediction which produces a faster than real time sequence prediction<sup>9</sup>.
- (v) Sequence disambiguation<sup>7</sup>.

Sequence completion is the simplest form of testing for memory of a sequence. Here the network is given an opportunity to learn a sequence of length  $n$  by repeated presentations of the external pattern sequence. Then the network is tested by observing the output caused by a presenting smaller number of patterns (e.g. a "probe test" of one input pattern). If the sequence of output patterns stimulated by the probe test is similar enough to the output sequence when all  $n$  input patterns are used, we say that the sequence is learned. One trial learning is important because of the presumed role of the hippocampus in one trial learning problems. Spontaneous rebroadcast is an important attribute of the network because it provides a mechanism for reproducing previously learned encodings. Such a mechanism is necessary if the hippocampus teaches the cerebral cortex (i.e. the transformation from short-term to long-term memory storage). Jump ahead prediction is an attribute that is not often discussed, but one that seems fundamental to solving many real world problems; i.e., it is usually desirable to make predictions at faster than real time rates. In other words, jump ahead prediction is a decided advantage to an organism struggling to survive.

Sequence disambiguation is perhaps the most important of all the phenomena that we have demonstrated. By solving sequence disambiguation problems, the network shows that it does indeed learn context. The beginning of the sequence disambiguation problem is learning to complete sequences, but there is more. Fig. 1 diagrams an example of a sequence disambiguation problem. Note the two separate sequences that the network is to learn for later completion, and note the shared subsequence of five sequential patterns which occurs in the middle of each full sequence. Now consider the sequence completion problem: Suppose I ask you to complete a sequence, and I tell you the starting position is pattern 1. Then you can easily produce the

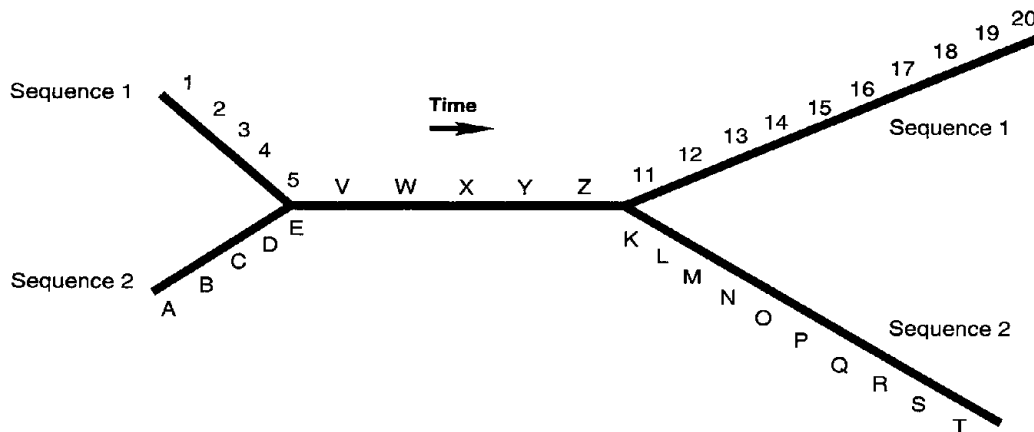


Fig. 1. The sequence disambiguation problem. Here we illustrate two sequences of 20 patterns. The two sequences share the subsequence VWXYZ. The noise free version of a pattern in any one segment of this stick figure is orthogonal to any pattern in another segment.

correct answer as pattern 20. And, likewise, if I give you pattern A, you easily predict that pattern T will occur. However, if I inform you only of pattern Z, then there is no right answer for the end of the sequence because Z precedes both pattern T and pattern 20. It is only by knowing what preceded Z by several time steps, e.g. pattern 5 or pattern E, that it is possible to predict the correct end point of the sequence. Thus, when representation is dynamic, there must be memory that goes back in time to solve this problem. We can consider this memory back in time to be context or, more mathematically, a conditioning variable of a conditional probability. The ability of the model to solve this problem is particularly interesting because the modification rule only associates across one time step, and the conduction delay between neurons is also but a single time step. Therefore, in order to solve this disambiguation problem, the network must create a code for the past that exceeds the time spanning nature of the individual elements. The network is able to do this because of the neurons it recruits (via locally determined associative modification) to code for pattern representations. These recruited neurons are what we call local context neurons. These local context neurons fire for brief and essentially sequential time steps so that they are representations, on a neuron by neuron basis, of small subsequences of the patterns. Fig. 2 is an example of the externally driven neurons during learning, and Fig. 3 shows network firing before and after learning. Note the local context neurons that fire for six or more sequential time steps.

We may also consider the sequence problem to be a spatial problem. That is, the sequence of patterns represents the changing multimodal sensory impressions impinging upon a rat as it explores a maze or an open field. In this case **these local context neurons are, at the very least, analogous to the hippocampal place cells** described in the psychological literature.<sup>10</sup>

Fig. 4 shows network performance in one particular example of the disambiguation problem. The output patterns of the CA3 representations in response to a single probe test are compared to the patterns of cell firing produced by the full input sequence. We are asking you, the reader, to act as a decoder: You look across each row for the most similar code words among the 20 fully driven outputs. In this particular example, we have given a probe test of pattern 1 from sequence 1, and you see that the network goes to very near pattern 20 and very far from pattern

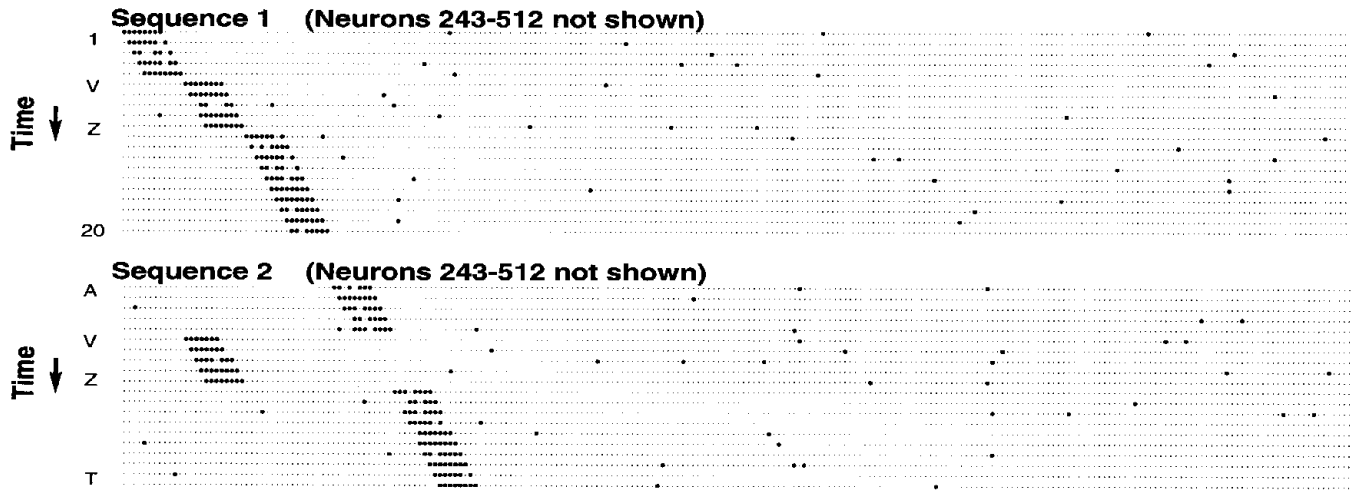


Fig. 2. Two noisy input sequences with overlap. The upper and lower panels illustrate typical input activity patterns. The base patterns have been perturbed by probabilistic noise. On-bits have been complemented with the probability of .1 and off-bits have been complemented with the probability of .01. Time is read as going down, and neurons are read from left to right. Only 243 of the 512 input neurons are illustrated.

T.

## 5. Recent Results.

Having demonstrated that the network can learn context, which we believe is the basis for encoding episodic memories, we wanted to show how useful such codes are for cognitive mapping. Fig. 5 illustrates the looping path problem which is constructed by adding sequence 1 and 2 of Fig. 2 to produce one 40 pattern sequence. The network sequentially receives input patterns 1 to 40 with the proviso that the input pairs for patterns 6 and 26, 7 and 27, 8 and 28, 9 and 29, and 10 and 30 are identical. Again, we remind the reader that the input patterns in the noise free situation merely consist of turning on 8 of 512 neurons. Note also that the external codes for the subsequences that code *i*) the loop, *ii*) the final tail, *iii*) the initial five sequences, and *iv*) the overlapping pairs are orthogonal, one subset compared to any other. As we repetitively present this sequence of 40 external patterns, the network develops its own code for the sequence so that the network can do something like sequence completion. However, the sequence completion it shows is very specialized.

### 5.1. Finding a short cut

There are three characteristics of cognitive mapping, and, in its specialized form of sequence completion, the network produces one of them. In particular, a cognitive map should allow an animal to find and use shortcuts. We start the network by randomizing neural excitation and then give the pattern of position 1 as an input. Then the network spontaneously generates its own sequence of patterns. This sequence is not patterns 1 through 40, instead the sequence of patterns essentially avoids the loop. That is, the sequence generated goes to pattern 40 without traversing the loop. This can be seen in Fig. 6A. As before this figure compares the sequence of patterns produced in response to random network excitation plus the presentation of input

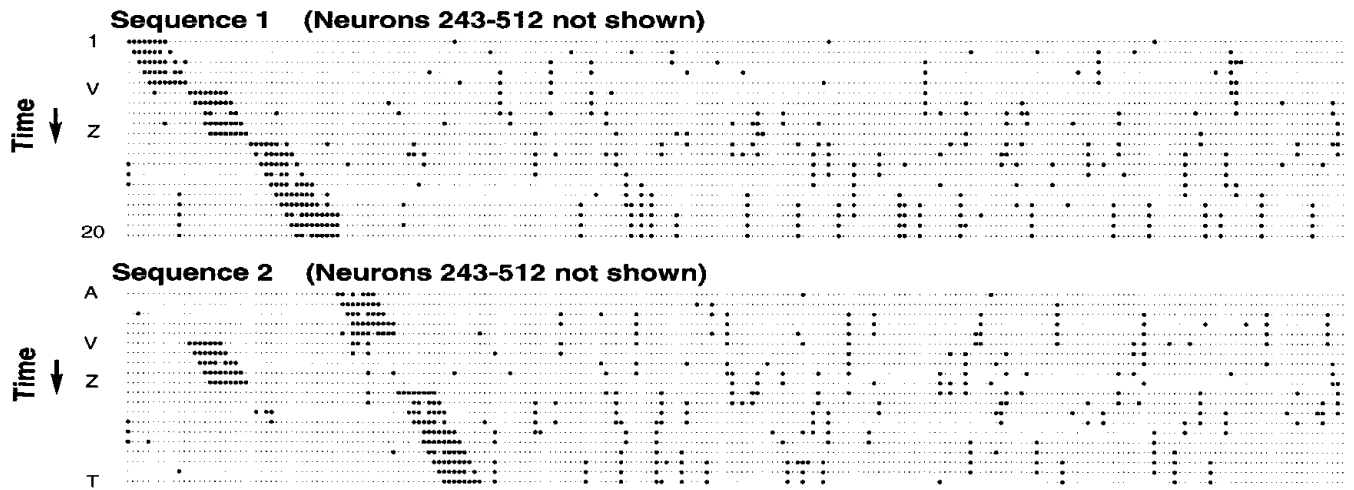


Fig. 3. Fully developed context codes. Here we illustrate the firing patterns of a portion of the network after learning. By comparison to the previous figure, we see the small role played by the externally driven neurons compared to the recurrently activated neurons. Within the two parts of this figure, it should be noted that neurons are firing in short sequences and that once a neuron fires such a sequence, it tends not to fire again. Time goes from top to bottom for each of the two sequences. Neurons go from left to right.

pattern 1 against the fully driven sequence (i.e., when the sequence of all 40 external patterns is presented as inputs). It can be seen that the network goes through the overlapping part of the sequence, appears to waiver a little bit between the tail of the sequence and the beginning of the loop (actually the codes here are similar and the representations must always slightly favor the tail), and then the patterns follow out through the tail. Thus, the network is able to predict shortcuts to get through sequences.

## 5.2. *Goal seeking*

It has been suggested, perhaps as a criticism, that there is an attractor at the fortieth representation, and therefore this shortcut finding behavior is trivial. While we certainly agree that there is an attractor around representation 40, it is not at all clear that such behavior is trivial because the network is able to produce an appropriate sequence (i.e., appropriate to its input experiences), rather than just following any arbitrary sequence to this attractor. Secondly, this attractor is not strong enough to overcome another desirable attribute of the network.

Consider the problem of a rat trying to predict a goal location while sitting at the beginning of a sequence. We place a thirsty rat at position one. We now imagine the looping path to be a maze, but, in fact, we use the exact same inputs and learning as before (therefore, we test the same network). Suppose there is water at pattern 21. Then, by definition, part of the code for position 21 is the input code for water. In analogy to placing a thirsty rat at position one, we provide, as an input pattern, pattern 1 plus the fraction of pattern 21 that corresponds to water. In this case, we turn on four of the external neurons associated with pattern 21 for the entire time. This corresponds to the rat knowing that it is thirsty and thinking about water as it imagines its way down the path.

So, the question is: Can the network produce sensible behavior (e.g. find the water) despite

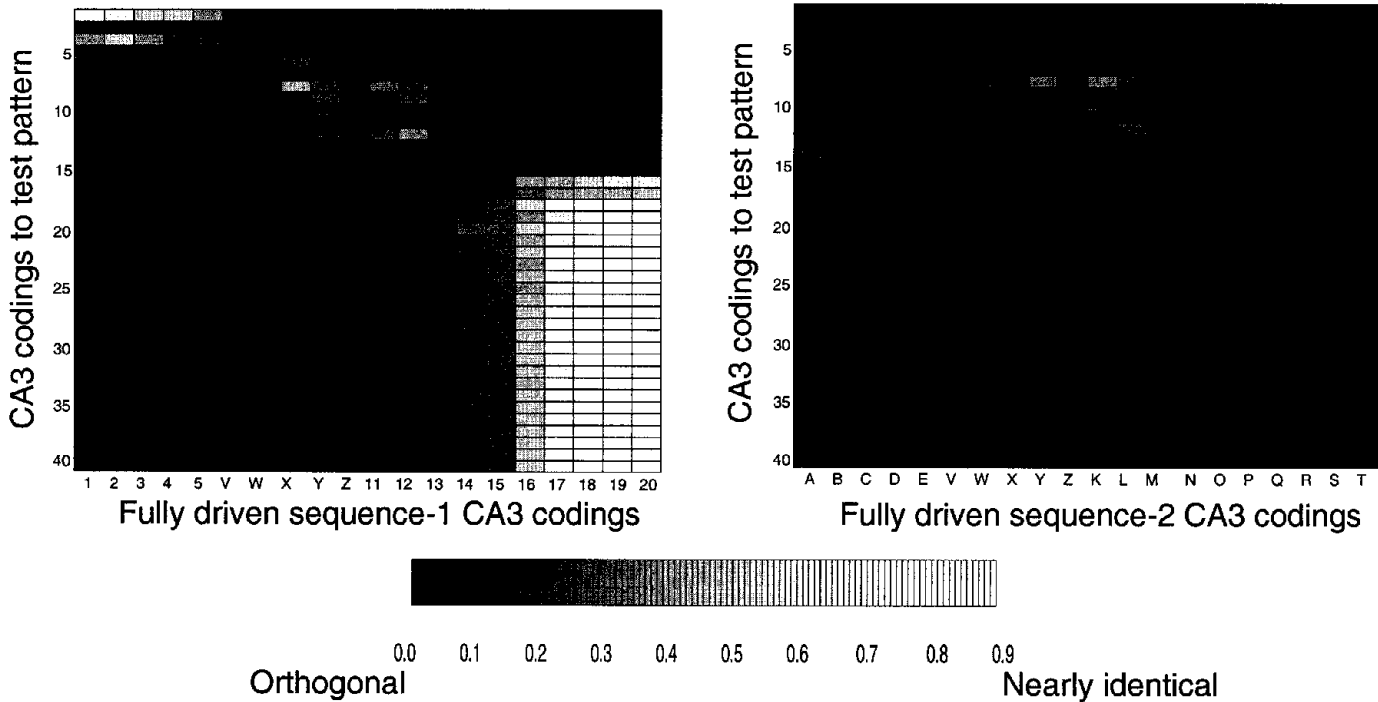


Fig. 4. Successful sequence disambiguation. Here we present two plots of pattern similarity. The left plot compares the output sequence generated in response to a probe test of a noise free version of pattern 1 against the patterns generated by the full 20 patterns of the first input sequence. The graph on the right compares the output sequence generated by the same probe test to the sequential patterns of the second sequence. Probe test generated output patterns are indexed on the ordinate while the fully driven patterns are indexed on the abscissa. Time goes from top to bottom and left to right. The lightness of each rectangle indicates pattern similarity. Nearly identical patterns are white while orthogonal patterns are black. Note that the uppermost left rectangle of the left graph is light and corresponds to a value of 0.82. (This is to be expected because the probe test pattern is exactly like the learned first pattern except for the noise present during learning and the random initializations of the network.) It should be obvious by comparing these two graphs that it is position 20 of the left graph (sequence 1) that is the final representation produced by the probe test pattern. Thus, the network has found the end of sequence 1. Further we note how black column T is on the right graph which indicates how far away from pattern T the final network representations are. In this, and in all following, similarity comparisons, the cosine of the angle between vector pairs is used for similarity comparisons.

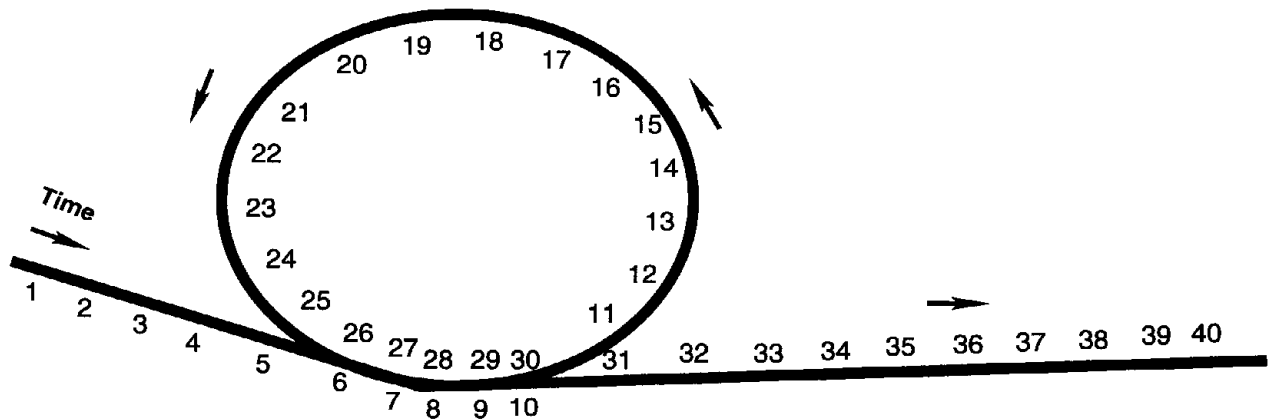


Fig. 5. The looping problem. This figure illustrates a 40 pattern sequence which repeats itself. Specifically, the subsequence 6-10 is identical to the subsequence 26-30 in terms of external neuronal activation.

the attractor that exists around representation 40? Fig. 6B shows that the network can indeed do this. Turning on some of the neurons associated with pattern 21 apparently creates a new attractor that produces the sequential behavior needed for our CA3 model to imagine the location of the water. That is, the network produces nearly the full code for position 21. (In fact, the network actually overshoots by one and ends up oscillating around a pattern most similar to position 22 but still quite similar to its code for position 21.) Once again, the network appears to have trouble right where the tail and the loop split, but this result follows naturally from the similarity of the representations at this point.

Thus, the network is capable of finding shortcuts, and it is capable of goal seeking. It is notable that this goal seeking is accomplished without exhaustive search. That is, the goal becomes an attractor so that the dynamics of the network itself are sufficient to solve the problem.

## 6. Summary.

Many aspects of this network deserve comment, but we shall only make three points. First, it is notable that this network makes its own codes for sequences. That is, there is no back-propagation nor even an error corrector. Nor is there any forcing of neurons receiving zero activity externally, and less than 10% of the active neurons (and therefore the ones subject to synaptic modification) arise from external firing. Thus, local adaptive processes produce the useful context encodings.

Second, the network does something very sensible in creating context codes. As a general principle of neural networks, we recognize the idea: Similar patterns tend to code similarly. But, when we consider a network or an organism that exists in time so that any input pattern is an element of a sequence, we have to change this idea. The simple extension is: Temporal neighbors tend to code similarly. In either case, these are only *a priori* principles. With experience, coding should change. Based on work not shown, the coding changes with experience in these networks so as to approximate the following principle: To the extent one pattern is a good predictor of another pattern, these two patterns are coded similarly. This tendency for similar coding apparently depends on correlatedness. Such similarity between encodings of patterns is expressed by the local context neuron firings and embodies the predictive representations

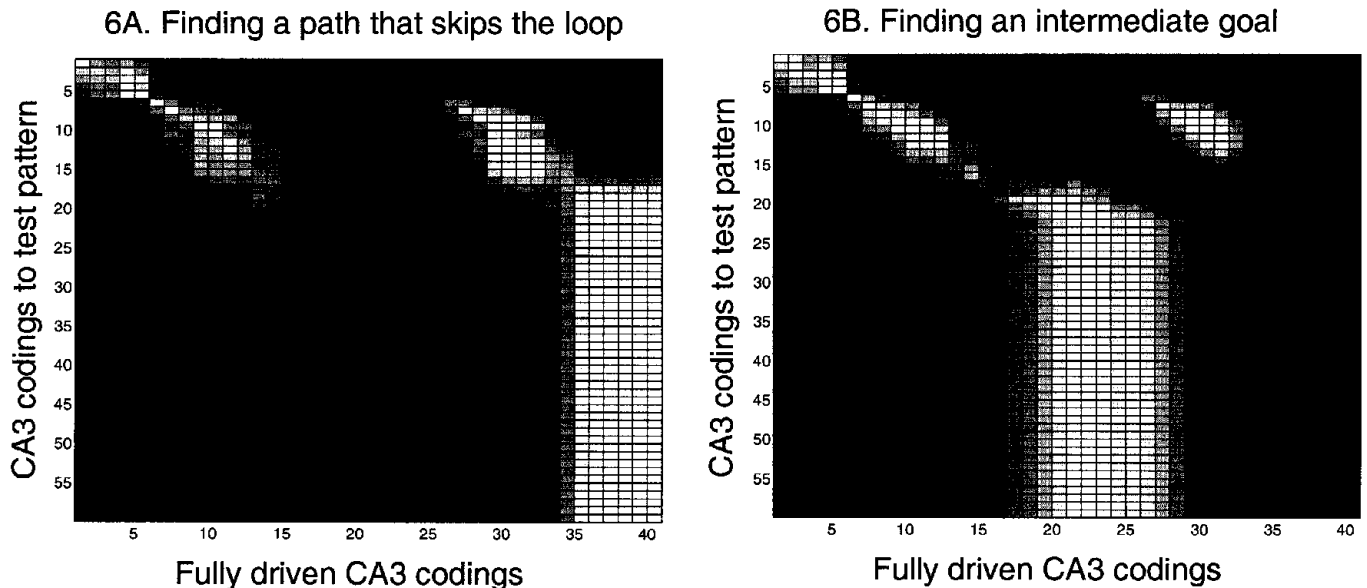


Fig. 6. A: Finding the shortcut. Having repetitively presented the looping sequence and allowed synaptic modification, the network was probed tested with pattern 1 of this sequence and allowed to run freely. This plot shows the similarity of the patterns generated (indexed on the ordinate) in response to the probe test compared to the patterns generated to a fully driven sequence (abscissa). Note how the network largely skips the loop and arrives at pattern 40 within 20 steps.

B: Goal seeking. Here we presented the network with the first pattern and also turned on four of the externally driven neurons of pattern 21 (standing for the location of water). Note how the network avoids the tail of the sequence and goes to pattern 22. The scale and similarity measurement are as in Fig. 4.

discussed previously<sup>1</sup>.

Finally, the simplicity of the model shows that context codes evolve from circuitry and network dynamics. These codes cannot be implicit in the individual elements themselves. That is, we have only endowed the elements of the network with memory across one time step, so the development of local context cell firing must follow from activity patterns of a randomly connected network. Moreover, based on the disambiguation problem, it is clear that these context codes can span at least five time steps. Unfortunately, by making the network so simple, we can only claim an analogy between local context cells and place cells. If we are to make proper predictions of place cell firing, a more biologically realistic network is required. However, there is no obvious reason why greater biological realism will harm the network performance demonstrated here, and such an extension appears relatively straightforward.

### Acknowledgments

This work was supported by NIH MH48161, MH00622, and EPRI RP8030-08 to WBL, and by the Department of Neurosurgery, Dr. John A. Jane, Chairman.

### References

1. W. B. Levy, "A computational approach to hippocampal function," *Computational Models of Learning in Simple Neural Systems*, eds. R. D. Hawkins and G. H. Bower, Academic Press,

New York, 243–305 (1989).

2. W. B Levy, "Unification of hippocampal function via computational considerations," *INNS World Congress on Neural Networks*, **IV**, 661–666 (1994).
3. R. Hirsh, "The hippocampus and contextual retrieval of information from memory," *Behav. Biol.*, **12**, 421–444 (1974).
4. R. P. Kesner, "Learning and memory in rats with an emphasis on the role of the hippocampal formation," *Neurobiology of Comparative Cognition*, eds. R. P. Kesner and D. S. Olton, Lawrence Erlbaum Associates, Hillsdale NJ, 179–204 (1990).
5. H. Eichenbaum, T. Otto and N. J. Cohen, "Two functional components of the hippocampal memory system," *Behav. Brain Sci.*, **17**, 449–518 (1994).
6. A. A. Minai and W. B Levy, "Sequence learning in a single trial," *INNS World Congress on Neural Networks*, **II**, 505–508 (1993).
7. A. A. Minai, G. L. Barrows and W. B Levy, "Disambiguation of pattern sequences with recurrent networks," *INNS World Congress on Neural Networks*, **IV**, 176–181 (1994).
8. W. B Levy, "Associative encoding at synapses," *Proceedings of the Fourth Annual Conference of the Cognitive Science Society*, 135–136 (1982).
9. C. Prepscius and W. B Levy, "Sequence prediction and cognitive mapping by a biologically plausible neural network," *INNS World Congress on Neural Networks*, **IV**, 164–169 (1994).
10. J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map*, Oxford University Press, Oxford, (1978).