

**The Formation of Neural Codes in the Hippocampus: Trace Conditioning as a  
Prototypical Paradigm for Studying the Random Recoding Hypothesis**

William B Levy<sup>1</sup>, Aprotim Sanyal<sup>1</sup>, Paul Rodriguez<sup>2</sup>, David W. Sullivan<sup>3</sup>, and Xiangbao Wu<sup>1</sup>

<sup>1</sup>Department of Neurosurgery, University of Virginia, P.O. Box 800420,  
Charlottesville, VA. 22908, USA

<sup>2</sup>Department of Cognitive Science, University of California, Irvine,  
Irvine, CA. 90297, USA

<sup>3</sup>Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New  
Jersey, 197 University Avenue, Newark, New Jersey 07102, USA

Keywords: CA3, random recoder, prediction, classical conditioning, minimal modeling, neural  
network

Correspondence to: Dr. William B Levy  
University of Virginia Health System  
P.O. Box 800420  
Department of Neurosurgery  
Charlottesville, VA 22908-0420  
Phone: (434) 924-9996  
Fax: (434) 982-3829

E-mail: wbl@virginia.edu

**Abstract.** The trace version of classical conditioning is used as a prototypical hippocampal-dependent task to study the recoding sequence prediction theory of hippocampal function. This theory conjectures that the hippocampus is a random recoder of sequences and that, once formed, the neuronal codes are suitable for prediction. As such, a trace conditioning paradigm, which requires a timely prediction, seems by far the simplest of the behaviorally-relevant paradigms for studying hippocampal recoding. Parameters that affect the formation of these random codes include the temporal aspects of the behavioral/cognitive paradigm and certain basic characteristics of hippocampal region CA3 anatomy and physiology such as connectivity and activity. Here we describe some of the dynamics of code formation and describe how biological and paradigmatic parameters affect the neural codes that are formed. In addition to a backward cascade of coding neurons, we point out, for the first time, a higher-order dynamic growing out of the backward cascade – a particular forward and backward stabilization of codes as training progresses. We also observe that there is a performance compromise involved in the setting of activity levels due to the existence of three behavioral failure modes. Each of these behavioral failure modes exists in the computational model and, presumably, natural selection produced the compromise performance observed by psychologists. Thus, examining the parametric sensitivities of the codes and their dynamic formation gives insight into the constraints on natural computation and into the computational compromises ensuing from these constraints.

## 1 Introduction

In servicing the neocortex as the association cortex of last resort, our interpretation of hippocampal function emphasizes random recoding suitable for prediction and forecasting (Levy 1989, 1994; Levy et al. 2005).

The relevance of this information processing perspective to the cognitive functions of the hippocampus comes through the specific cognitive tasks solved by simulations of the model and the resulting recodings. The variety of tasks is appropriate to the variety of cognitive interpretations of normal hippocampal function including contextual, configural, and spatial paradigms (Levy 1996) that require an intact hippocampus. The oddest learning task that requires normal hippocampal function is the escape-air-puff version of trace conditioning (Solomon et al. 1986). The paradigm (see Fig. 1) is odd in that it is neither spatial, contextual, or configural; yet, the hippocampus is needed to learn the trace interval even though the hippocampus is not necessary for long-term retention (Kim et al. 1995), a close analogy of the H.M. hippocampal syndrome and his problems with declarative memory (Milner 1972; Cohen and Squire 1980).

Although odd from the viewpoint of cognitive theory, the escape-air-puff trace conditioning paradigm has several characteristics that recommend it for computational investigations. First, it is extremely simple so it promises to be easy to understand. Second, the escape version of the task is clearly a sequence forecasting problem. Third, the phenomena of the task are well-quantified. Fourth, and somewhat under-appreciated, this is the only hippocampal task which uses the exact same stimuli, with the same functional meaning (i.e., avoiding the air-puff with some kind of blink), across mammalian species. In particular, the stimuli have been made identical for rabbits and humans (compare, e.g., Clark and Squire (1998) versus Solomon et al. (1986)).

Initially we pointed out the qualitative (Levy and Sederberg 1997), and then pointed out the quantitative (Rodriguez and Levy 2001), correspondences between simulations of our computational model and behavioral and neurophysiological observations of air-puff trace conditioning. These similarities include the learnable and unlearnable ranges of the trace interval, the failure modes, the variability in learning, the abrupt nature of escape onset during training, and the generic type of neurons that are seen once successful escape begins (see Rodriguez and Levy (2001) for details and its comparison with McEchron and Disterhoft (1997) and with Gormezano et al. (1983)). These correspondences and the listed characteristics motivate more detailed studies of this paradigm in regard to the recoding phenomena that occurs during training.

The present article concerns the across-training network dynamics that occur during learning including the underlying neuronal codes of the failure modes and some of the biological variables which affect successful versus unsuccessful learning. While making headway on these issues, we also compare the codes formed by three related models of hippocampal region CA3.

## **2 Models and Methods**

In fact, “the” minimal model is a family of models where details can be included as desired. Within this family of models there are several qualitatively shared properties that will be used here:

- 1) Communication between neurons is binary, i.e., single spikes are passed between connected primary neurons.
- 2) The primary neurons are McCulloch-Pitts, i.e., at each neuron, the recurrent excitation is a weighted sum of the active, connected excitatory inputs.
- 3) There is a mechanism that maintains the average activity of the network.

- 4) Positive feedback is the primary source of neuronal firing: here, when CS or US stimuli are active; 70% of the firing is due to recurrent activation; the remaining 30% of the neuronal firings is due to the entorhinal cortex/dentate gyrus inputs; during the trace interval, 100% of the firings is due to the CA3-to-CA3 recurrent connections.
- 5) Recurrent connectivity is fractional, typically 10% or less, and it is random.
- 6) Activity is less than 50%, here always less than 15% and preferably much lower.
- 7) Synaptic modification is both associative (LTP and LTD) and time-spanning in the form observed by Levy & Steward (1979, 1983)

Of special interest here is the role played by two parameters: average activity and fractional connectivity. We will also shed some light on quantitative differences between three versions of the minimal model. For the purposes of this paper, these three are

Model c1: Competitive activity control and a one time-step spanning synaptic modification equation with potentiation and depression. The trace conditioning results using the model can be found in Sullivan and Levy (2004b).

Model  $c\alpha$ : Competitive activity control and a time-spanning synaptic modification rule where the time-span is an exponentially-decaying function which acts like the off-rate time constant of the NMDA-receptor (see Holmes & Levy 1990).

Model  $f\alpha$ : This model is similar to the one used previously to reproduce the quantitative aspects of trace interval learning (Levy and Sederberg 1997; Rodriguez and Levy 2001) and, like model  $c\alpha$ , model  $f\alpha$  also uses the continuous, time-spanning, exponentially-decaying synaptic modification equation. Again as in the earlier work, but differing from models c1 and  $c\alpha$ , activity control is more biological. A fast responding feedback inhibitory interneuron (one time-step delay) and shunting inhibition including a resting  $K^+$ -like conductive shunt helps control activity and fluctuations (see Smith et al. 2000). Finally, what differs in the present model from the earlier publications is the existence

of modifiable excitatory synapses onto this interneuron. Moreover, here this model always has at least one randomization process operating at each training trial.

Table 1 here

Table 2 here

The variables of the models are defined in Tables 1 and 2. The internal excitation of any neuron  $j$ ,  $y_j$ , is in models c1 and c $\alpha$  a simple weighted summation over the active  $i$  afferent synapses to  $j$ , and when we include synaptic failures it is:

$$y_j(t) = \sum_{i=1}^n \phi(Z_i(t-1)) C_{ij} W_{ij}(t-1)$$

In models c1 and c $\alpha$ , the firing variable  $Z_j(t)$  is calculated by a non-biological  $k$ -winners-take-all rule. This rule enforces strict activity control by specifying that the neurons with the  $k$  highest values for internal excitation fire, where  $k$  is set at the desired number of firing neurons,  $a \cdot n$ . Ties at threshold are resolved by uniform randomization. By virtue of the large number of synapses and the chaotic nature of neuronal firing (Minai and Levy 1994), ties are inevitably restricted to the first training trial. External input,  $X_i(t)$ , always fires a neuron.

Model f $\alpha$  does not have the artificial, perfect activity control of the competitive models. Rather, it employs a feedback interneuron to regulate activity. Specifically, internal excitation of a pyramidal neuron in model f $\alpha$  uses the following equation, which includes divisive inhibition,

$$y_j(t) = \frac{\sum_{i=1}^n \phi(Z_i(t-1)) C_{ij} W_{ij}(t-1)}{\sum_{i=1}^n \phi(Z_i(t-1)) C_{ij} W_{ij}(t-1) + K_R \left( \sum_{i=1}^n D_i(t-1) Z_i(t-1) \right) + K_I \sum_{i=1}^n X_i(t) + K_0}$$

Neuronal firing,  $Z_j$ , is calculated as:

$$Z_j = \begin{cases} 1 & \text{if } y_j \geq \frac{1}{2}, \text{ or } x_j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Synaptic failures are included in model  $c\alpha$  and model  $f\alpha$ . The synaptic failure channel of the connection from neuron  $i$  to neuron  $j$  is represented by the function  $\phi$  (Sullivan and Levy 2003, 2004a), where  $\phi(Z_i = 0) = 0$  and a synaptic failure  $\phi(Z_i = 1) = 0$  occurs with probability  $f$ , and  $\phi(Z_i = 1) = 1$  with probability  $(1 - f)$ .

Two dynamically controlled sets of variables  $W_{ij}$  (in models  $c1$ ,  $c\alpha$  and  $f\alpha$ ) and  $D_i$  (in model  $f\alpha$ ), represent synaptic weights. The variable  $W_{ij}$  represents the weight of a CA3 pyramidal cell to pyramidal cell synapse ( $i$  to  $j$ ), while  $D_i$  represents the weight of a synapse from a CA3 pyramidal cell  $i$  onto the feedback interneuron. The modification of these values is governed by the following equations:

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t), \text{ and}$$

$$\Delta W_{ij}(t) = \mu \cdot Z_j(t) \cdot (\bar{Z}_i(t-1) - W_{ij}(t)).$$

The variable  $\bar{Z}_i$  is a short-term memory of presynaptic excitation. In model  $c1$ ,  $\bar{Z}_i(t)$  is exactly equivalent to  $Z_i(t)$ . That is,  $w_{ij}(t+1) = w_{ij}(t) + \mu Z_j(t) (Z_i(t-1) - w_{ij}(t))$  is used in model  $c1$ . However, in models  $c\alpha$  and  $f\alpha$ ,  $\bar{Z}_i(t)$  is calculated using an NMDA receptor-like saturate-and-decay rule, as previously described by Rodriguez & Levy (2001),

$$\bar{Z}_i(t) = \begin{cases} \bar{Z}_i(t-1)\alpha & \text{if } \phi(z_i(t)) = 0, \text{ or} \\ 1 & \text{if } \phi(z_i(t)) = 1. \end{cases}$$

In the denominator of the model  $f\alpha$ ,  $\sum_{i=1}^n D_i(t-1)Z_i(t-1)$  represents the contribution of a global feedback interneuron using an autonomous rule for interneuron afferent synaptic modification (Sullivan and Levy 2003). The weight of a synapse from pyramidal cell  $i$  to the feedback interneuron are updated as

$$D_i(t+1) = D_i(t) + \lambda \cdot Z_i(t) \cdot \left( \frac{\sum_{j=1}^n Z_j(t)}{n} - a \right).$$

Randomization of the initial state,  $Z(0)$ , (see Shon et al. 2002) is part of model  $f\alpha$ . On each trial, both training and testing,  $a \cdot n$  neurons are activated by a uniform random process.

For model  $f\alpha$ , it is necessary to adjust the inhibitory parameters when altering connectivity, failure rate, number of neurons, or activity. Table 3 summarizes the parameterizations used in this paper.

Table 3 here.

The trace paradigm is summarized in Fig. 1. All model  $c\alpha$  simulations use a CS of three time-steps and a US of three time-steps while all model  $f\alpha$  simulations use a CS lasting five time-steps and a US lasting eight time-steps as illustrated in Fig. 1B. These  $f\alpha$  simulations better reflect the published behavioral and neurophysiological studies (compare Rodriguez and Levy 2001 to McEchron and Disterhoft 1997).

### 3 Visualization Methods

### 3.1 Reordering and thinning

It is instructive to visualize individual neuronal firing data generated by the simulations. Unfortunately, because the model behaves as a random recoder, such images are usually too complicated to interpret. Interpretation is further hampered when these diagrams are crowded by virtue of the thousands of neurons in a simulation. To overcome these difficulties, we employ reordering and thinning procedures to make a digestible visualization of the sequential neuronal firing patterns.

To create these diagrams, the CS neurons are placed at the top of the diagram, and just beneath them are the US neurons, both unaltered. However, for visualization purposes, every neuron, except those externally activated, must undergo the reordering and sampling procedure. Since the only meaningful ordering of neurons in a randomly recoding network is the sequence of firing, reordering is based on the time-step at which a neuron first fires during a specified trial. The reordering rule is simple: the earlier a neuron fires, the higher in the reordering it is placed. (Ties are broken pseudo-randomly by original neuronal indices.) To compare firing patterns across training trials, we select a single trial that is the basis for all trials. Therefore, it is important for the reader to note, for any such diagram, on which training trial the ordering depends.

Finally, in order to make the diagrams legible, it is necessary to thin the number of neurons displayed since thousands of neurons will not fit legibly on a single page. To accomplish this, all reordered neurons are proportionately sampled, thus maintaining the same relative frequency of neuron types (including neurons that may have dropped out of the code). For example, in a simulation with  $n = 2048$  where for both CS and US there are 80 neurons, a total of 70 out of 1968 recurrent neurons are displayed, and this is the sampling fraction. Such a display is seen in Figs. 2, 3, 4, 6, 7, 8.

### 3.2 *The within-trial US activity diagram*

For the trace paradigm being modeled, we view the hippocampus as the critical site of US prediction in the brain. Another brain region takes the hippocampus's timely prediction of the US and produces an appropriate motor output. We assume the time delay for transmitting this prediction within the brain is small compared to the time it takes to get the muscles to move the eye (in the case of a nictitating membrane blink) or an eyelid blink (in the case of a human-like eyelid blink). Moreover, both of these times are small compared to the duration of a blink, which can exceed 300 ms.

In any event, the definition of the successful prediction interval in terms of US firing is somewhat arbitrary as is the active fraction of US neurons that are needed for the next brain region to recognize the US is actually being predicted. As a result, this fraction becomes part of the hypothesis that is the model. Fortunately, these issues are not of primary concern here because our interest is in the dynamics and the learned neuronal codes. Here, with a few exceptions as in Fig. 12, we simply measure performance more-or-less continuously. This continuum of US prediction is viewed by the within-trial US activity diagram.

With this diagram, we can visualize the onset times of the US neurons during testing in terms of when and what fraction of US neurons are activated. Added to these two dependent variables, i.e., the fraction of US neurons active as a function of time, we can also plot an independent variable which we place on the y-axis. The independent variable of most interest here is the length of the trace interval (see Figs. 9 and 10). This unconventional format (i.e., independent variable on the y-axis) is justified because the x-axis corresponds to the x-axis of the ordered-and-thinned neuron firing diagrams, which represent within-trial time.

## 4 Results and Observations

To facilitate exposition, we must create some jargon. These terms are listed in Table 4. The most important term is 'code' or phrase 'neural code.'

Table 4 here.

Using the reordered and thinned firing diagrams (see Methods), the dynamics of code formation can be seen for a model  $c\alpha$  simulation (Fig. 2) and a model  $f\alpha$  simulation (Fig. 3). These figures illustrate a sequence of test trials each of which is simulated immediately after the indicated training trial. For the indicated parameter settings, training culminates in a satisfactorily learned US prediction as shown in the last three panels of Fig. 2 and the last two panels in Fig. 3.

- For both simulations, we note five general features of code formation across training trials
- 1) a forward growth or, equivalently, a forward stabilization; i.e., code growth begins at the beginning of the sequence,
  - 2) a predominantly backward cascade of coding neurons from the early firing position, including a striking backward cascade of US neurons,
  - 3) the statistical elimination, in a gradual manner, of firing of many neurons (in Fig. 2, neurons 118-130 and in Fig. 3 ca. neurons 118-151),
  - 4) temporal intermingling of some pure recurrent neurons with the external activation of US neurons,
  - 5) the longevity of local context firings tends to be greater at the end than at the beginning of the trace interval.

Observation (1) will become more precise when we quantify the stabilization process. Processes (2) and (3) go hand-in-hand. That is, as coding neurons are recruited into the bridging sequence, they fire more and more. As a result of their increased firing, the

normalizing inhibition makes other neurons less likely to fire. Observations (4) and (5) are predictions of these two models and can be refuted or confirmed by experimental neurophysiology.

One of the biggest differences between the simulations of the two models is the number of trials needed for learning. However, the learning rates of model  $c\alpha$  vs. model  $f\alpha$  are not comparable since these models differ in several ways. Learning is slower in the model  $f\alpha$  simulation because the synaptic modification rate is smaller ( $\mu = 0.05$  in model  $c\alpha$  vs. and  $\mu = 0.01$  in model  $f\alpha$ ) and because there is a greater trial-to-trial random fluctuation in model  $f\alpha$  simulations. These fluctuations are produced by (1) trial-to-trial  $Z(0)$  randomization, (2) quantal synaptic failures at rate  $f = 0.15$ , and (3) by the inherent activity fluctuations of model  $f\alpha$  simulations that are absent in model  $c\alpha$ .

We now examine the code formation dynamic of model  $f\alpha$  in more detail.

#### *4.1 The backward cascade and the lengthening of local context firing at high resolution*

First, as we have noted and quantified elsewhere, the backward cascade over training dominates the code formation dynamic. However, this observation is an average result of a stochastic process; that is, there is a spectrum of place shifting including some small fraction of neurons that forward cascade across training trials (see also Fig. 3 of Mitman et al. 2003).

It seems impossible to disentangle the development of context cell firing from the backward cascade; moreover, the role played by lengthening of local context in the backward cascade depends on the time within a trial and at what point we look at training. Early in training and early in the trace interval, the backward cascade and repetitive firing by individual recurrent neurons that ultimately survive as coding neurons are closely correlated across trials; i.e., the tendency to backward cascade and the growth of local context neuron firing co-occur. This co-

occurrence does not seem to be the case later in training for the neurons coding the end of this trace interval. Figure 4 shows examples of these phenomena by superimposing two pairs of test trials.

Each panel of Fig. 4 redisplay data from a pair of firing plots of Fig. 3. The upper panel compares trials 5 and 15 and focuses on the subset of recurrent neurons 81-95. Note that for most neurons in the range of 81-90, there is a backward cascade based on earliest, within-trial firing. Specifically, neurons {81, 82, 84, 85, 86, 87, 89, and 90} fire earlier on test trial 15 than they fired on test trial 5. The two exceptions in this range are neurons 83 and 88, while recurrent neurons ranging from neuron 91 and above are still rather unsettled in their firing positions. That is, prior to these neurons settling into their eventual positions late in training and prior to an easily discernible backward cascade, their firing is diffusely localized throughout a trial. Since the upper plot of Fig. 4 illustrates early training trials, the dual observations of (1) the shift by neurons in the range of 81-90 versus (2) the unsettled positioning of neurons with numbers greater than 91 is an example of the forward stabilization of the backward cascade. This observation is supported by the lower panel and, with some qualifications, the data in the next section concerning the late firing neurons.

The lower panel of Fig. 4, also derived from Fig. 3, shows the backward cascade at the end of the sequence which occurs later in training, and also, this panel confirms the earlier forward stabilization claim. This two-trial firing plot compares neurons 81-117 during the test after training trial 50 and the test after training trial 200. Ignoring the neurons that fire only in a single test trial, and comparing to trial 50: a plurality of neurons, 14, fire earlier on trial 200 and this illustrates the backward cascade; one neuron (#107) fires later on trial 200; and 3 neurons fire at the identical time for both tests. Of the remaining neurons, 10 neurons fire on test trial 200, but not on test trial 50; 8 neurons fire on test trial 50 but not on test trial 200; and two neurons do not fire. (Neurons 81, 82, and 97 might seem peculiar, but it must be remembered that the neuron ordering is based on trial 199 and that these simulations have two sources of between

trial randomization—initial state randomization,  $Z(0)$ , and quantal synaptic failures. These same facts also explain the two neurons in the 81-116 range that do not appear on either trial 50 or trial 200.)

Externally-activated neurons are only a little different in their across-trial dynamic. That is, the US neurons backward cascade across training but they are never lost during a training trial, and in a training trial, they can only backward cascade by increasing local context length. The explanation is the same for both phenomena; i.e., on training trials, external neurons are anchored to the end of the sequence by the training procedure.

In conclusion, both parts of Fig. 4 illustrate that the backward cascade dominates the across-trial dynamic of the coding neurons, and together they illustrate that the backward cascade starts at the front of the sequence. Therefore, we describe the dynamic across training as a forward traveling stabilization which is constructed from a backward cascade of coding neurons. Further analysis supports this idea of a forward traveling wave of stabilization.

#### *4.2 Code stabilization*

The concept of a forward cascade of the stabilized code over training can be quantified, and the codes formed by the fast-learning model  $\alpha$  simulations of this section are empirically stable, at least with the selected parameters. As Fig. 5 illustrates, the rate of stabilization depends on (1) activity level, (2) within-trial time-steps that a coding neuron occupies, and to a lesser extent, (3) the trace interval. Consistent with the individual model  $f\alpha$  example of the previous section, coding neurons that sit at earlier time-steps are the first to stabilize. However, the more quantitative approach of this section shows that the stabilization sequence is actually nonmonotonic. Neurons that end up occupying the central time-steps are the last to stabilize in the case of training on successfully predicted trace intervals. For example in Fig. 5, for the 25 time-step trace interval, codewords that occupy time-steps 4, 5, and 6 stabilize in half the

number of trials as the codewords in the center of the sequence. Notably, though, the stabilization of the time-step 28 codeword occurs just one training trial after the stabilization of the time-step 6 codeword.

The existence of an empirical stability that allows this quantification is sensitive to model parameterization. This sensitivity is particularly true for the free-running simulations of model  $\alpha$  when simulating at the upper bound of the learnable trace intervals. Eventually, these simulations tend to collapse. In particular, when trained on a hard to learn, longer trace interval, a simulation can produce satisfactory US prediction for several hundred trials and then deteriorate into a collapse after 500 to 600 training trials. Nevertheless, even if asymptotic stability does not obtain, the large number of training trials required to produce a collapse might be enough to ignore in practical situations; i.e., across training trials, there are many trials of correct prediction, and this run of successes would be biologically useful. Elsewhere we have discussed such an idea, that transient, successful prediction is useful if the hippocampus can be taken off-line (see Monaco and Levy 2003 for other observations and biological interpretations concerning unstable but transiently useful hippocampal encodings).

We now examine the role that model parameters play in the codes that form, but first we must describe our classifications of the codes.

#### *4.3 Failure modes*

From the behavioral viewpoint, there are three failure modes and one success mode: respectively, blink too soon, blink too late, no blink, and a timely blink that blocks the air-puff. As reported in Rodriguez and Levy (2001), the model produces all four modes including the two behaviorally appropriate failure modes for overly long sequences, blink too soon or no blink. Here, only the model  $\alpha$  simulations, which is almost identical to the CA3 model in Rodriguez and Levy (2001), produce each of these modes at all the appropriate trace intervals.

In model  $c\alpha$ , varying the trace interval produces two failure modes and one success mode at qualitatively appropriate trace intervals. In the first panel of Fig. 6, training on a short, six time-step, trace interval leads to a failure to bridge. In the third panel, training on a long, 32 time-step, trace interval leads to a code collapse. The middle panel, trained on a 15 time-step trace interval, successfully predicts the US. See Figure Legend for more details.

Comparison of the failure modes of model  $f\alpha$  to model  $c\alpha$  indicates that model  $c\alpha$  is acceptable for part, but not all of the behavioral phenomenology; i.e., does not produce the failure-to-blink mode at the long time trace interval. (Model  $c1$  and  $c\alpha$  can fail to bridge but only at very short trace length intervals and it is apparent from Fig. 7 that a collapse and a failure to bridge are related across training.)

Figure 8 illustrates a model  $f\alpha$  simulation which fails to bridge a long trace interval. Note that this failure mode includes an attractor of recurrent neurons illustrated by the Fig. 8, trial 200 panel. The observation of a recurrent-based attractor indicates that recurrent neurons can form a very strong attractor, and this attractor is forming without any outside direction. Another interesting observation is the variety of modes displayed at this same parameterization of model  $f\alpha$ . Different simulations of the same model  $f\alpha$  network, train on this one trace interval a in Fig. 8, show a spectrum of neuronal codes, including (1) an appropriate prediction, a failure to bridge as shown in Fig. 8, and a code collapse.

Finally, the typical model  $f\alpha$  (but not necessarily models  $c1$  or  $c\alpha$ ) code collapse is a backward cascade that includes the US. In terms of the development of a model  $f\alpha$  code collapse, the US backward cascade is, during the course of training, typically preceded by end-of-sequence recurrent neuron firing that occupies a relatively large span of the trace interval.

The physiological relevance of the code collapses illustrated here must be tempered by the inappropriately high activity levels which the simulations are using. Lowering activity to physiological levels will be pursued in future research. Moreover, the obvious inclusion of  $Ca^{+2}$ -activated  $K^{+}$ -conductances will prevent the nonphysiological, prolonged firings pictured here.

Nevertheless, we stand by the current model's prediction of the blink-too-soon failure mode that a collapse also predicts.

The next two sections examine the interaction of trace interval with each of three biologically relevant parameters, fractional activity, fractional connectivity, and  $\alpha$ .

#### *4.4 Activity interactions with trace interval in model $f\alpha$*

Using a test trial after the last training trial, we can judge a learned neural code by quantifying the pattern of US neuron firing. This quantification and presentation of US cell firing avoids the arbitrariness faced by both the psychologist (how far must the nictitating membrane move before it is a blink?) and by us (how many US neurons is enough for a prediction of the US itself?). Thus, each phase diagram shows the relative US firing at each time-step of a test trial. The US activity diagrams of Fig. 9 and Fig. 10 are described in the Methods and in the Figure Legends. Note that time within this last test trial is plotted on the x-axis while the y-axis plots an independent variable – the length of the trace period used during training.

As a set of four US activity diagrams, Fig. 9 shows that lower activity helps prevent code collapse brought on by longer trace intervals. Likewise, Fig. 10 shows the same effect by lowering connectivity. That is, either lowering activity or lowering connectivity increases the learnable trace interval. In terms of the activity dependence shown in Fig. 9, one can see that a 12.5% activity setting produces a maximal learnable trace interval of about 30 while at 5% activity a little more than 40 time-steps can be bridged and predicted. In terms of connectivity, Fig. 10 shows that the 10% connectivity simulations are predicting trace intervals of 35-40 while 2.5% connectivity predicts intervals of 50. In both cases, however, there is a penalty to pay for the best upper bound performance. The very short time-span trace intervals are not learned because of a failure to bridge when activity is too low or when connectivity is a relatively small value; i.e., note the white region in the lower right corner of the 5% activity of Fig. 9 phase

diagram and in the 2.5% connectivity of Fig. 10. This white region means a failure to bridge; i.e., US neurons are not activated at all during testing. At 7.5% activity, this region is smaller and at 10% and 12.5% activity this region disappears altogether. The other utility of these diagrams is showing tendencies for code collapses (blink too soon) in these model  $\alpha$  simulations. Clearly, long trace intervals produce this failure mode. On the other hand, one can see the upper boundary of collapses rising as activity or connectivity decreases.

As much detail as Fig. 9 shows, it still does not reveal anything about individual simulations since each point is an average US firing across ten simulations. In particular, there are often two and sometimes three performance modes at any one trace interval/activity pair. With the help of a performance criterion for bridging (specifically, at least 10% US neurons at the first and the second time-step of the US), Fig. 11 helps resolve this uncertainty as it illustrates failures to bridge.

Keeping in mind Fig. 9, Fig. 11 tells us something important about code formation. Let us concentrate our comparisons on the lowest activity, 5.0% versus the highest activity, 12.5%. At 12.5% activity, the trace interval of 40 time-steps is not learnable. From the information in Fig. 9 and Fig. 11, the dominant failure mode at this interval is a too-early prediction of the US. Lowering activity to 5% allows the simulations to succeed in learning this trace interval since (or perhaps because) the backward cascade is reduced. Of course, a simulation can always be defeated by lengthening the trace interval, but in the case of 5% activity, the failures at the upper boundary of learnable intervals are interesting by virtue of differing from the failure modes at this equivalent boundary when simulating at 12.5% activity. As Fig. 11 shows, the failures at the upper boundary of the learnable interval ( $\sim 50$  time-steps) for 5.0% activity simulations is often a failure to bridge in contrast to the too-early prediction of the US seen in simulations run at 12.5% activity and the upper boundary of the learnable trace interval. Thus, lowering activity decreases the backward cascade on two accounts; it allows successful learning at a 40 time-

step trace interval by avoiding a too great backward cascade of US neurons, and it leads to failure to bridge at trace intervals in the vicinity of 50 time-steps.

Similar observations occur in the case of connectivity reduction.

#### *4.5 Connectivity interactions with trace interval in model $\alpha$*

The connectivity effect on the learnable trace interval seems quite substantial so we performed additional model  $\alpha$  simulations examining this variable. Figure 12, a phase diagram, illustrates the regions of consistently successful and consistently unsuccessful US prediction as a function of connectivity and trace interval.

The boundary line of this phase diagram shows the upper boundary of the reliably learnable trace interval. For the five points on the line illustrated in Fig. 12, the number of successful simulations out of 10 are {7, 7, 9, 9, and 10}, for the connectivities of {1%, 1.5%, 2.5%, 5%, and 10%} respectively, using a prediction criteria of 40% of the US neurons at time-step 13. At five time-steps above the upper boundary of successful simulations, simulation performance is unacceptably poor. Specifically, the performance is {3, 4, 3, 2, and 3} out of 10 simulations for the same five connectivities. Five time-steps below the boundary performance is {7, 8, 9, 8, 10} out of 10 simulations of the 1% through 10% connectivities. The lower boundary of this diagram has not yet been explored.

## **5 Discussion**

Motivating this study is a desire to understand the functional range of our random recoder sequence-learning hypothesis and to gain some insight into the workings of this random recoding. In this regard, we have (1) revealed some of the across training dynamics of code formation (particularly Figs. 3 and 4), and we have (2) determined the role played by certain

parameters of behavioral or biological relevance (here trace interval, activity,  $\alpha$ , and connectivity; elsewhere CS and US duration (Wu and Levy 2004)) on the recoding process.

Because a deeper understanding of the code formation dynamic is facilitated by simplifying the  $f\alpha$  model, we juxtapose the results from simulating models  $c1$  and  $c\alpha$  against  $f\alpha$  simulations. Model  $c1$ , which lacks an identifiable time-scale, produces qualitatively similar results where possible except that the local context lengths are homogenous across the code rather than increasing across time-step position (Sullivan and Levy 2004b). In terms of codes, the simplification of model  $c\alpha$  vs. model  $f\alpha$  comes at a cost of losing the equivalent of the no-blink failure mode at overly long trace intervals; this mode is replaced by an overly strong tendency to collapse and to include CS neurons in such a collapse when using the  $c\alpha$  model. Finally, model  $f\alpha$ , model  $c\alpha$ , but less so model  $c1$  codes may show individual neurons firing for overly long periods at the end of the sequence. We consider these cell firings to be merely qualitative or relative observations. That is, we submit the hypothesis that neurons coding for the end of trace interval fire for longer times than the coding neurons early in the sequence but we emphasize that they do not fire at the absolute rates indicated here. Thus, this part of the model is qualitative because the actual frequency of firing is not properly represented. To get firing rates right requires integrate-and-fire neurons and firing-dependent self-inhibitory mechanisms (e.g., voltage and  $Ca^{+2}$ -activated  $K^+$  conductances); thus these models are inaccurate in terms of absolute firing (see Levy et al. (2005) for more insights on this topic and extending model  $f\alpha$ ). Nevertheless, keeping these provisos in mind, we believe these simulations and the following speculations are currently the best explanation for how the hippocampus, using the dentate gyrus and CA3 as a recoder and CA1-subiculum and deep layers of entorhinal cortex as a decoder, is able to learn to predict the US across a learnable trace interval.

We conjecture that the first issue in code formation is bridging, and bridging is a state-space path discovery problem. For successful bridging to occur, enough feasible paths through the restricted activity patterns that define the codeword state-space must exist (e.g., state-space in

model  $\alpha$  is the set of all vectors with  $a \cdot n$  ones and  $n(1 - a)$  zeros). By enough, we mean that there is a reasonable probability of a random discovery of an allowed state-space sequence that ends with externally-activated US neurons while starting with externally-activated CS neurons.

The bridging issue comes to the fore in at least three situations. First, bridging is a particularly acute problem for very short trace intervals (see Fig. 11), second, for long trace intervals in simulations with low activity or low connectivity absence of bridging is also an observed failure mode, and third, there is the dynamic of successful bridging, i.e., the establishment of the final span that completes a bridging sequence. With regard to this third point, one subtle and perhaps critical observation of the code formation dynamic is well illustrated in Fig. 5.

Specifically, the last part of the code to stabilize occupies the later central parts of the trace interval. This last part to stabilize can be thought of as the last span of the state-space bridge; it is added between the end of sequence attractor and the far (i.e., late) end of the sequence transient attractors growing out of the CS. This last span of the bridge cannot be an arbitrary state space vector of activity  $a \cdot n$  because it must also be compatible with the two transient attractors that it bridges.

The possibility of a stable, connection-feasible state-space sequence from one vector to another depends on the number of connections between the firing neurons in the earlier and later codewords that the last span connects. The probability that a random coding process will find a stabilizable bridging for this last span depends on fractional activity, connectivity, and the time-span of associativity (as well as the early and late codewords being bridged). A longer time-span of associativity allows longer bridging segments, and this lengthening increases the number of possible bridging points and therefore number of possible bridgings. Increasing connectivity or activity can be good or bad. Higher connectivity or activity helps create bridging spans by increasing the feasible set of state space patterns. Such parameter increases also increase the probable amount of active neuron overlap between two codewords. However, too much connectivity or activity can also lead to another failure mode.

A final bridging might lead to a code collapse or, equally problematic, a too great backward cascade of the US neurons. Too much activity or connectivity encourages this problem. Metaphorically the collapse issue is something like the over-writing problem of symmetric feedback networks, but the problem of code collapse is more complicated than simple over-writing. Based on the observations here, we can make qualitative conjectures about the typical collapse. First, a collapse grows as a backward cascade of neurons that are extending their local context length. It seems that a critical set of issues is: (i) the net strength of connectivity within a group of end-of-sequence coding neurons which (ii) must be viewed relative to this same net strength property for the preceding set of coding neurons and (iii) that these intra group net connectivity values must be viewed relative to the number of active connections between a slightly earlier set of coding neurons that bridge into both of the later groups.

Let us consider a sequence of state space vectors  $X(T-2)$ ,  $X(T-1)$ , and  $X(T)$  where  $X(T)$  is the end of sequence attractor, and thus it might include many US neurons. The natural tendency of a sparsely connected random recurrent network is to wander through state space, always moving away from its current state space position (Amari 1971; Sompolinsky et al 1988; Minai and Levy 1993). This tendency can be overcome by associative synaptic modification in the form of LTP between a set of interconnected neurons. If the excitatory interconnections among a state space feasible set of neurons is large enough, then this set can be an attractor of some notable permanence. More quantitatively, if activated, the dwell time of the network dynamic, relative to this set, increases as the interconnection strength within this set increases. Building from this single insight, we will use the notation  $\{X(t)\}$  to mean the set of firing neurons at  $t$ .

At some stage in training, it would have to be the case that the net connectivity from  $\{X(T-2)\}$  to  $\{X(T-1)\}$  is on average stronger than the connections from  $\{X(T-2)\}$  to  $\{X(T)\}$ ; otherwise the set of neurons  $\{X(T-2)\}$  would activate  $\{X(T)\}$

rather than  $\{X(T-1)\}$  at this stage of training. Later in training, the backward cascade of  $\{X(T)\}$  is able to infiltrate  $\{X(T-1)\}$  and eliminate, via recurrent inhibition, neurons that were part of  $\{X(T-1)\}$  at an earlier stage in training. A code collapse ensues if the set  $\{X(T)\}$  can be excited by  $\{X(T-2)\}$  without too much weakening (via LTD) of its interconnections. A collapse is facilitated by parameters that increase the probability of strengthening the synapses between the  $\{X(T-2)\}$  and  $\{X(T)\}$  firing neurons and within the  $\{X(T)\}$  back to itself. As made explicit by the results here and in Mitman et al. (2003), these parameters are fractional activity, fractional connectivity, and the time-span of associative potentiation.

In sum, the trace conditioning paradigm is serving as a prototype for understanding the random recoding sequence prediction interpretation of hippocampal information processing. By examining the details of the code formation dynamic in this prototype, we gain insight into the constraints that shaped the biological basis of hippocampal function.

## Acknowledgments

This work was supported by NIH MH63855 to WBL, the Meade Munster Foundation, and by the Department of Neurosurgery. The authors thank Janine M. Vallee for her assistance.

## References

Amari S (1971) Characteristics of randomly connected threshold –element networks and network systems. Proc IEEE 59:35-47.

Clark RE, Squire LR (1998) Classical conditioning and brain systems: The role of awareness. Science 280:77-81.

Cohen NJ, Squire LR (1980) Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science 210:207-210.

Gormezano I, Kehoe EJ, Marshall BS (1983) Twenty Years of Classical Conditioning Research with the Rabbit. In: Progress in Psychobiology and Physiological Psychology. (A. N. Epstein, Ed), New York: Academic Press, pp. 197-267.

Holmes WR, Levy WB (1990) Insights into associative long-term potentiation from computational models of NMDA receptor-mediated calcium influx and intracellular calcium concentration changes. J Neurophysiol 63:1148-1168

Kim JJ, Clark RE, Thompson RF (1995) Hippocampectomy impairs the memory of recently, but not remotely, acquired trace eye-blink conditioned responses. Behav Nsci 109:195-203.

Levy WB (1989) A computational approach to hippocampal function. In: Computational Models of Learning in Simple Neural Systems. (R. D. Hawkins and G. H. Bower, Eds.), New York: Academic Press, pp. 243-305

Levy WB (1994) Unification of hippocampal function via computational considerations. INNS World Congress on Neural Networks IV-661-666.

Levy WB (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. Hippocampus 6:579-590.

Levy WB, Hocking AB, Wu XB (2005) Minimal Models of Hippocampal Structure and Function. Neural Networks, submitted.

Levy WB, Sederberg PB (1997) A neural network model of hippocampally mediated trace conditioning. IEEE International Conference on Neural Networks I-372-376

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. Neurosci 8:791-797

McEchron MD, Disterhoft JF (1997) Sequence of single neuron changes in CA1 hippocampus of rabbits during acquisition of trace eyeblink conditioned responses, J Neurophysiol 78:1030-1044.

Milner B (1972) Disorders of learning and memory after temporal lobe lesions in man. Clinical Neurosurgery 19:421-446.

Minai AA, Levy WB (1993) The dynamics of sparse random networks. Biol Cybern 70:177-187.

Minai AA, Levy W (1994) Setting the activity level in sparse random networks. Neural Comp 6:85-99.

Mitman KE, Laurent PA, Levy WB (2003) Defining time in a minimal hippocampal CA3 model by matching time-span of associative synaptic modification and input pattern duration. International Joint Conference on Neural Networks (IJCNN) 2003 Proceedings, 1631-1636

Monaco JD, Levy WB (2003) T-maze training of a recurrent CA3 model reveals the necessity of novelty-based modulation of LTP in hippocampal region CA3. International Joint Conference on Neural Networks (IJCNN) 2003 Proceedings, 1655-1660

Rodriguez P, Levy WB (2001) A model of hippocampal activity in trace conditioning: Where's the trace? Behav Neurosci 115:1224-1238

Shon AP, Wu XB, Sullivan DW, Levy WB (2002) Initial state randomness improves sequence learning in a model hippocampal network. Phys Rev E 65:031914/1-15.

Smith AC, Wu XB, Levy W (2000) Controlling activity fluctuations in large, sparsely connected random networks. Network 11:63-81

Solomon PR, Vander Schaaf ER, Thompson RF, Weisz DJ (1986) Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. Behav Neurosci 100:729-744.

Sompolinsky H, Crisanti A, Sommers HJ (1988) Chaos in random neural networks. Phys Rev Lett 61:259-262.

Sullivan DW, Levy WB (2003) Synaptic modification of interneuron afferents in a hippocampal CA3 model prevents activity oscillations. International Joint Conference on Neural Networks (IJCNN) 2003 Proceedings, 1625-1630

Sullivan DW, Levy WB (2004a) Quantal synaptic failures enhance performance in a minimal hippocampal model. Network 15:45-67.

Sullivan DW, Levy WB (2004b) Activity affects trace conditioning performance in a minimal hippocampal model. Neurocomputing, in press

Wu XB, Levy WB (2004) Increasing CS and US longevity increases the learnable trace interval. Neurocomputing, in press.

## Figure Legends

**Fig. 1.** Schematic representation of the trace conditioning paradigm. Part A illustrates the version of Pavlovian conditioning discussed in this article. In any one training trial the conditioned stimulus (CS) is presented first. After the CS ends (CS offset), there follows a time interval, called the trace interval; during the trace interval no external stimuli are presented. The onset of the unconditioned stimulus (US), an external stimulus orthogonal to the CS, defines the end of the trace interval and remains active for a brief, fixed time. A test trial consists of the same CS input followed by no external activation. Part B is an input firing diagram showing the trace paradigm inputs as the sequence of a CS, then a trace interval, and finally the cell firing representation of the US. The CS is represented by neurons 1-30 firing at time-steps 1-5; the trace interval is represented by the absence of external input driving for time-steps 6-55; and the US is represented as neurons 31-60 being activated for time-steps 56-63. Because the network is randomly connected, the spatial juxtaposition of two neurons in this schematic is irrelevant to network performance and is used solely for explanatory convenience.

**Fig. 2.** Code formation in a model  $\alpha$  simulation. These reordered and thinned firing diagrams show the formation of a bridging and predicting sequence in a model  $\alpha$  simulation. The neurons in the vicinity of neuron 116 in these graphs display a strong tendency to backwards cascade, forming coalitions which grow backwards over the course of subsequent training trials. Additionally, the US neurons also backwards cascade as training proceeds. Repetitive firing, measured as local context length in earlier publications, is greater among the later coding recurrent neurons than the recurrences coding for early parts of the trace interval. Neurons early in the sequence (starting around neuron 60) show another effect – a forward cascade of stabilization. Each CS-alone test trial is produced after the indicated training trial. The reordering is based on trial 100. The simulation uses 1000 neurons; activity is fixed at 10%;

fractional connectivity is 10%;  $\alpha$  is 0.7165, CS and US  $m_e$  is 30 neurons, and  $\mu = 0.05$ . The trace interval for this simulation is 20 time-steps. The CS (neurons 1-30) and US (neurons 31-59) are held on for 3 time-steps each during each training trial.

**Fig. 3.** Code formation in a model  $f\alpha$  simulation. These reordered and thinned firing diagrams show the formation of a bridging and predicting sequence in a model  $f\alpha$  simulation. In general, the code formation dynamic is similar to the model  $c\alpha$  simulation of Fig. 2. Again we see a strong backward cascade of neurons in the US as well as recurrent neurons that form firing coalitions, i.e., temporally overlapping with the US neurons. The forward stabilization effect is particularly pronounced in model  $f\alpha$  simulations, and it includes the growing context lengths of early firing neurons (compare trial 10 vs. trial 15). Each CS-alone test trial is produced after the indicated training trial. The reordering here is based on trial 199. The trace interval is 25 time-steps. For simulation details see the row of Table 3 with parameters  $\{n = 2048, a = 0.1\}$ . In all 2048 model  $f\alpha$  simulations, the externally-activated CS and US neurons are, each, a total of 40 neurons.

**Fig. 4.** The code formation dynamic. This high resolution overlay display is derived from Fig. 3. Comparing pairs of test trials across training shows that the code formation dynamic is a forward stabilization produced by a backward cascade of coding neurons. The backward cascade is first observable in the early phases of training. The upper panel superimposes trials 5 and 15 for recurrent neurons 81-95; most neurons in the range 81-90 have settled in to their eventual code position while neurons 91-95 appear less committed to their eventual location. Later in training, as shown in the lower panel which superimposes trials 50 and 200 for neurons 81-117, there is a backward cascade and stabilization for the neurons that code later parts of the trace interval code. In both the upper and lower panels, we can also see a lengthening of

local context firing but this effect is much more noticeable in the upper panel. In fact, in the lower panel, many neurons decrease local context length as they backward cascade.

**Fig. 5.** Quantified model  $\alpha$  code stabilizations. This figure displays the mean number of training trials needed to reach code stabilization where stabilization is measured for each time point within a training trial. Note that neurons coding the beginning of the sequence stabilize first, and this is generally true so long as a code does not collapse. However, the neurons coding the end of the sequence are reliably not the last to stabilize. For the longer, 25 time-step, trace interval, the end of trace code stabilizes before the middle stabilizes. Data from model  $\alpha$  simulations are parameterized as in Fig. 2. For each time-step, stabilization is defined as the earliest training trial when the state vector of neurons is approximately unchanged when compared to the state vector of the same time-step on any later trial up to training trial 100. “Approximately unchanged” is defined as 10% or fewer firing neurons in disagreement.

**Fig. 6.** Two of the failure modes compared to a successful prediction. By varying the trace interval, we can take the parameterized simulations of Fig. 2 and produce two failure modes or successful prediction. Specifically, if the trace interval is too short or too long, there is a failure to predict in a timely fashion. In the first panel, training on a short, six time-step, trace interval leads to a failure to bridge (no blink). In the third panel, training on a long, 32 time-step, trace interval leads to a code collapse (blink too soon). The middle panel, trained on a 15 time-step trace interval, successfully predicts the US. The illustrated test trials all occur just after training trial 100. Each panel here is reordered and thinned (see Methods) based on itself.

**Fig. 7.** The dynamic of a model  $\alpha$  code collapse. There is an initial bridging early in training but this bridge then disappears for several trials. Recurrent neurons with extended local context arise early in training at the end of the sequence. As training proceeds, these end-of-sequence

neurons give birth to similar, daughter recurrents that are firing even longer while they, the parents, are lost. Eventually many US neurons are incorporated, or in some cases reincorporated, into the bridging code. In contrast to earlier figures these are training trials. The trace interval is 32. All other parameters are as in Fig. 2. Reordering is based on trial 5. Thinning and reordering as described in Methods.

**Fig. 8.** A model  $f\alpha$  simulation fails to bridge a long trace interval. Model  $f\alpha$  simulations can fail to bridge a long trace interval; simulations of the same network with a different random seed for quantal failures and  $Z(0)$  can also demonstrate a code collapse rather than a failure to bridge. The end of sequence coding neurons show that the recurrent neurons can form a strong attractor without any external direction. This simulation uses 8000 neurons and a 10% connectivity; see Table 3 for more details. Reordering based on trial 199. The gap pictured in the last panel, around time-step 11, is misleading and is due to the thinning process.

**Fig. 9.** Lowering activity postpones code collapse associated with longer trace intervals. However, very short trace intervals lead to a bridging failure, particularly when activity is low. These within-trial-US Activity diagrams are explained in the Methods section. Data points are generated from model  $f\alpha$  simulations with 2048 neurons (see Table 3 for details). Data points are increments of one time-step for both axes with ten simulations per plotted point. The diagonal stripes localize the onset (dashed line) and offset (solid line) of the CS. Time zero is the time of US onset.

**Fig. 10.** Lowering connectivity postpones code collapse associated with long trace intervals. However, very short trace intervals lead to a bridging failure at very low connectivity. The qualitative changes are similar for two different  $\alpha$ 's. Data points are generated from model  $f\alpha$  simulations with 8000 neurons. In contrast to Fig. 9, increments on the y-axis are in time-steps

of 5. 10 simulations per plotted point. See Table 3 for details. The remainder of the figure as in Fig. 9.

**Fig. 11.** Failure to bridge as a function of trace interval and activity in model  $f\alpha$ . Although lowering activity decreases one failure mode, the tendency to collapse (see Fig. 9), lower activity does not necessarily improve prediction because bridging becomes a problem at both the upper and lower limits (all activities but the highest) of the reliably learnable trace interval. Both 7.5% and 10% activities are notable for the presence of a long continuous region of trace lengths, without any failures to bridge. Here, successful bridging is defined as at least 10% of the US neurons activated for at least two time-steps of a test trial. The data comes from the simulations producing Fig. 9.

**Fig. 12.** Lowering connectivity increases the learnable trace interval. This phase diagram illustrates the inverse relationship between the learnable trace interval and percent connectivity. The plotted line is the upper boundary of the simulations that produce a reliably prediction of the US. The criterion for successful prediction is less than 40% of US neurons turn on before time-step 13 from the US onset while more than 40% turn on in the interval of time-steps 7-1 prior to the US onset. For simulation parameters, see Table 3, 8000 neurons.

**Table 1.** Symbols for neuronal and synaptic equations

| <b>Symbol</b> | <b>Domain</b>       | <b>Meaning</b>                                                                                                                                                                                    |
|---------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $y_i(t)$      | $\{0,1\}$           | Internal excitation of neuron $i$ at time $t$                                                                                                                                                     |
| $Z_i(t)$      | $\{0,1\}$           | Binary firing state of neuron $i$ at time $t$ . $Z_i(t) = 1$ when neuron $i$ fires at time $t$ , $Z_i(t) = 0$ otherwise                                                                           |
| $C_{ij}(t)$   | $\{0,1\}$           | Binary variable representing the existence of a connection from neuron $i$ to neuron $j$ . $C_{ij}(t) = 1$ if and only if such a connection exists (this value is static throughout a simulation) |
| $W_{ij}(t)$   | $\{0,1\}$           | Strength of synaptic connection from neuron $i$ to neuron $j$                                                                                                                                     |
| $X_i(t)$      | $\{0,1\}$           | State of the external forcing input of neuron $i$ at time $t$ . When equal to one, $i$ fires regardless of internal excitation                                                                    |
| $\phi(\ )$    | $\{0,1\}$           | The quantal synaptic failure channel (models $c\alpha$ and $f\alpha$ only)                                                                                                                        |
| $D_i(t)$      | $(-\infty, \infty)$ | Weight of excitatory connection from neuron $i$ to the feedback interneuron (model $f\alpha$ only)                                                                                                |
| $i$           | $\{1, \dots, n\}$   | The index of a presynaptic neuron                                                                                                                                                                 |
| $j$           | $\{1, \dots, n\}$   | The index of a postsynaptic neuron                                                                                                                                                                |

**Table 2.** Symbols for network parameters

| <b>Symbol</b> | <b>Meaning</b>                                                                          |
|---------------|-----------------------------------------------------------------------------------------|
| $n$           | Number of neurons in the network                                                        |
| $a$           | Desired activity per time-step, expressed as a fraction of $n$                          |
| $K_r$         | *Feedback inhibition scaling factor                                                     |
| $K_I$         | *Feedforward inhibition scaling factor                                                  |
| $K_0$         | *Shunting inhibition constant                                                           |
| $\mu$         | Synaptic modification rate constant for pyramidal-pyramidal synapses                    |
| $\lambda$     | *Synaptic modification rate constant for pyramidal-interneuron synapses                 |
| $c$           | Fractional connectivity of the network: $n^{-2} \sum_{i=1}^n \sum_{j=1}^n (C_{ij}) = c$ |
| $W_0$         | Initial pyramidal-pyramidal weights                                                     |
| $D_0$         | *Initial pyramidal-interneuron weights 1.0                                              |
| $m_e$         | Number of externals forced on for each external input                                   |
| $f$           | Probability of failure, i.e., $f = P(\phi(1) = 0)$                                      |
| $\alpha$      | Synaptic time constant related to NMDA-receptor off-rate time constant                  |

\*Indicates model  $\alpha$  only.

**Table 3.** Model  $f\alpha$  Parameters

| n    | a     | c     | $\mu$ | $W_0$ | $m_e$ | $K_r$  | $K_i$ | $K_0$  | $\alpha$ | $\lambda$ | $f$  |
|------|-------|-------|-------|-------|-------|--------|-------|--------|----------|-----------|------|
| 2048 | 0.05  | 0.1   | 0.01  | 0.45  | 40    | 0.055  | 0.050 | 0.001  | 0.8187   | 0.5       | 0.15 |
| 2048 | 0.075 | 0.1   | 0.01  | 0.45  | 40    | 0.046  | 0.010 | 0.832  | 0.8187   | 0.5       | 0.15 |
| 2048 | 0.1   | 0.1   | 0.01  | 0.45  | 40    | 0.044  | 0.010 | 0.849  | 0.8187   | 0.5       | 0.15 |
| 2048 | 0.125 | 0.1   | 0.01  | 0.45  | 40    | 0.043  | 0.010 | 0.844  | 0.8187   | 0.5       | 0.15 |
| 8000 | 0.05  | 0.010 | 0.01  | 0.50  | 80    | 0.007  | 0.010 | 0.0686 | 0.82     | 0.5       | 0    |
| 8000 | 0.05  | 0.015 | 0.01  | 0.50  | 80    | 0.010  | 0.010 | 0.261  | 0.82     | 0.5       | 0    |
| 8000 | 0.05  | 0.025 | 0.01  | 0.50  | 80    | 0.0159 | 0.018 | 0.0001 | 0.82     | 0.5       | 0    |
| 8000 | 0.05  | 0.025 | 0.01  | 0.50  | 80    | 0.0159 | 0.018 | 0.0001 | 0.8669   | 0.5       | 0    |
| 8000 | 0.05  | 0.050 | 0.01  | 0.50  | 80    | 0.0298 | 0.018 | 0.463  | 0.82     | 0.5       | 0    |
| 8000 | 0.05  | 0.1   | 0.01  | 0.50  | 80    | 0.0512 | 0.018 | 1.058  | 0.82     | 0.5       | 0    |
| 8000 | 0.05  | 0.1   | 0.01  | 0.50  | 80    | 0.0512 | 0.018 | 1.058  | 0.8669   | 0.5       | 0    |

**Table 4.** Useful Jargon

**code** – the firing sequence of recurrently-activated neurons elicited by activation of the CS after some number of training trials.

**CS neuron** – a recurrently-connected CA3 neuron which is activated externally by the CS.

**US neuron** – a recurrently-connected CA3 neuron which is activated externally by the US.

**recurrent neuron** – any excitatory neuron that is not a CS or US neuron

**collapse or**

**code collapse** – the firing of US neurons too soon after CS offset. This is equivalent to the behavioral failure mode of blink-too-soon.

**bridging or**

**bridging sequence** – beginning with the end of external CS activation, a sequence of neural firings that leads to the activation of a criterion-specified number of US neurons.

**failure to bridge** – absence of a bridging sequence; i.e., insufficient firing of US neurons during a test trial; the equivalent of the behavioral failure mode of no blink to the CS.

**(successful) prediction** – US neurons fire before the US normally turns on but not too soon after the CS offset.

**recall** – bridging without prediction or collapse; that is, the US neurons turn on too late to be a useful prediction; the equivalent of the behavioral failure mode of blink-too-late.

**training trial** – see Fig. 1

**test trial** – the CS is the only external activation and synaptic modification is turned off

**local context neuron** – a neuron that contributes to a code by firing for a, more or less, contiguous subsequence of the code.

**local context length** – the number of time-steps a local context neuron fires.

Fig. 1A.

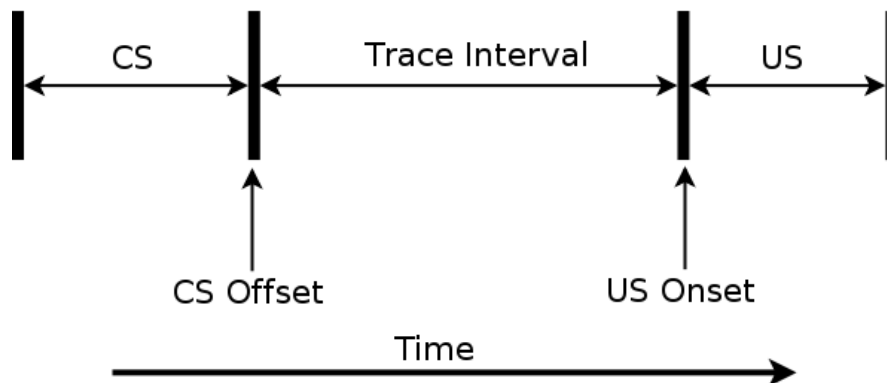
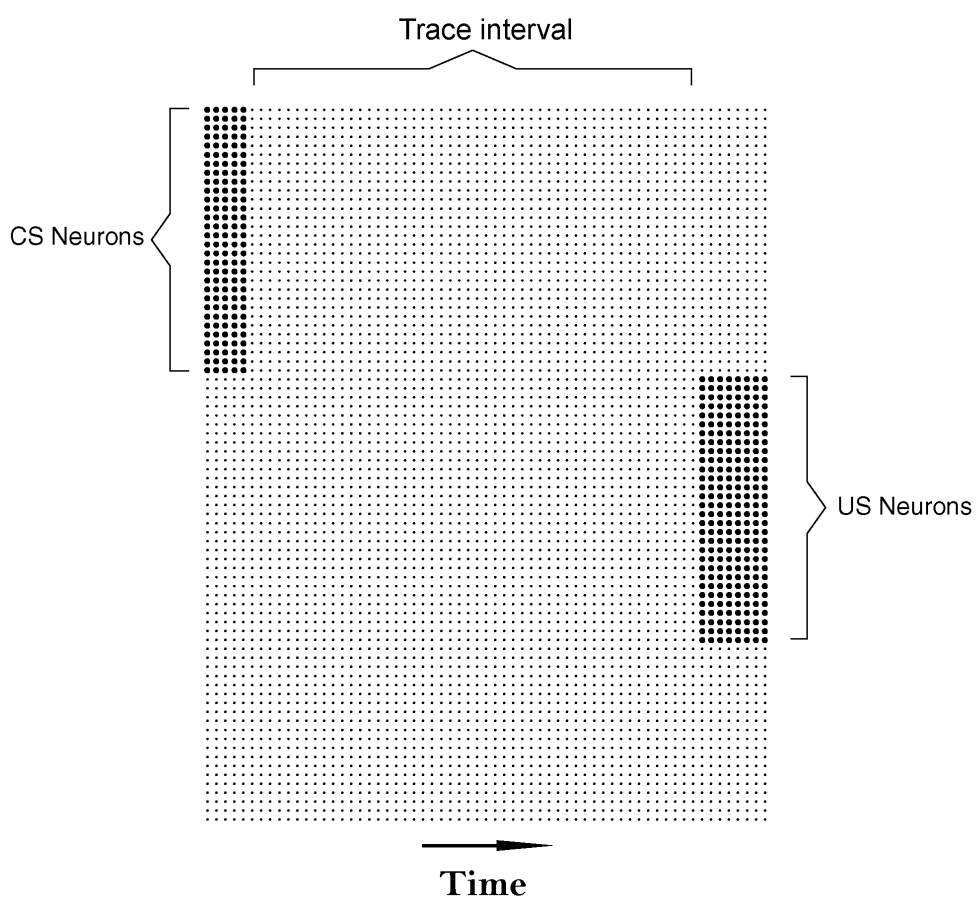
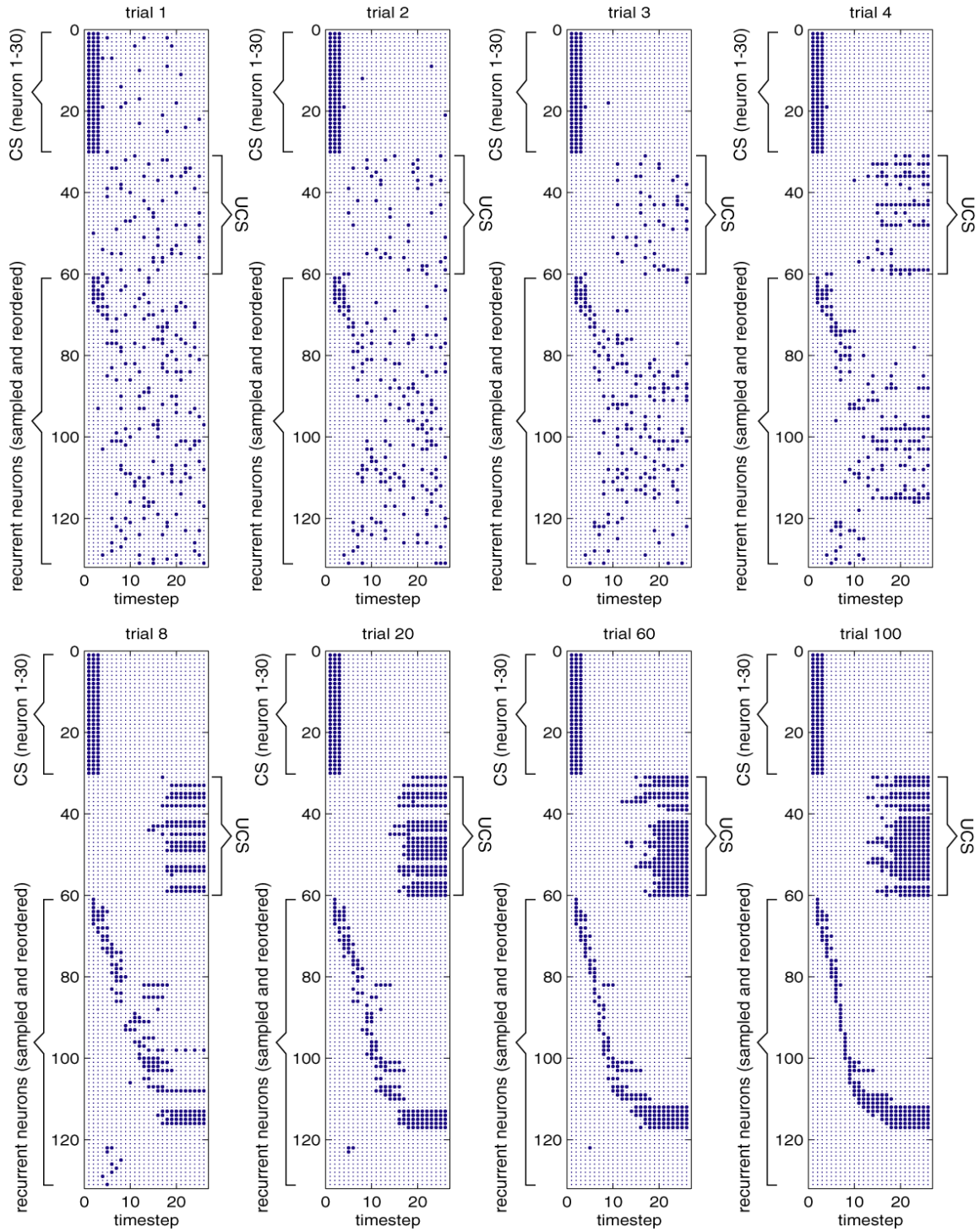


Fig. 1B.



**Fig. 2.**



**Fig. 3.**

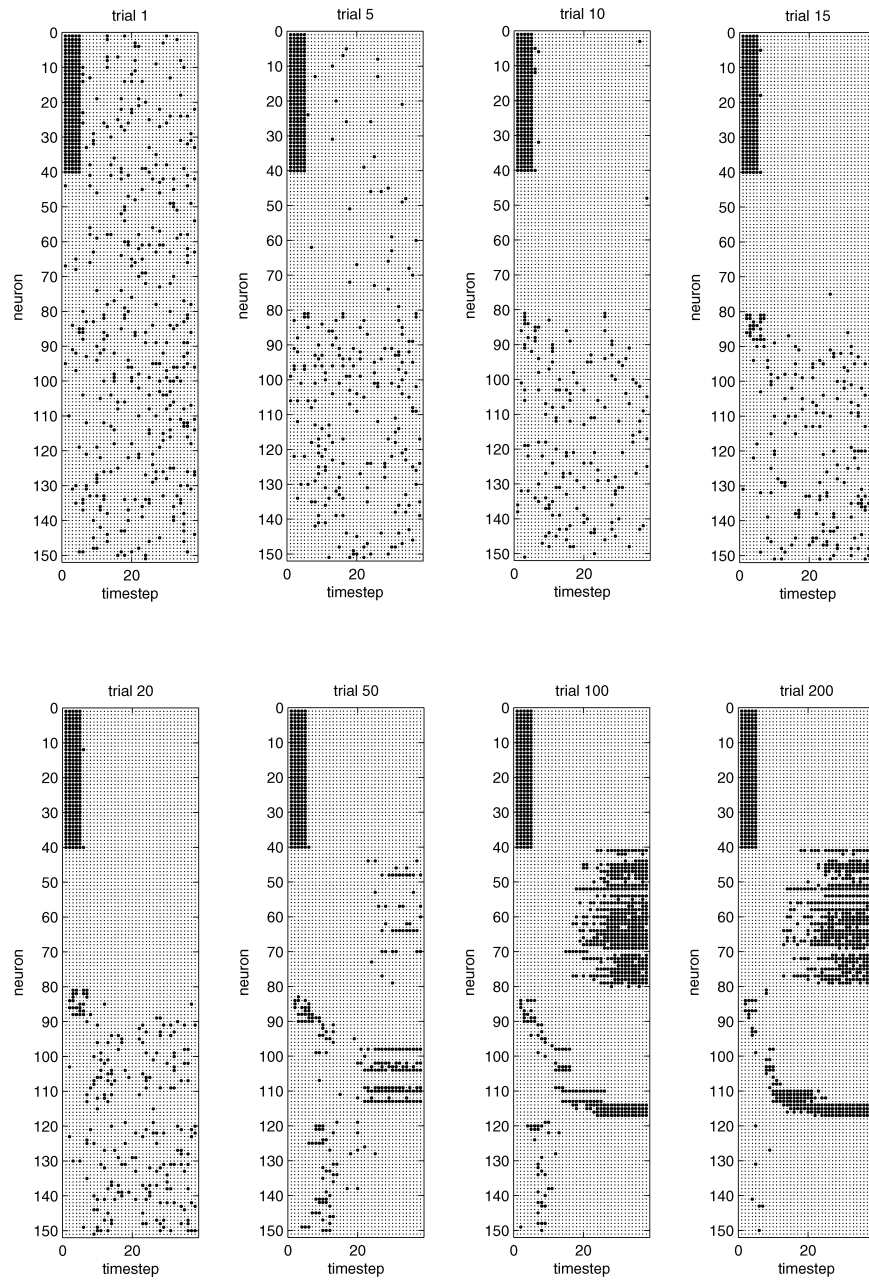


Fig. 4.

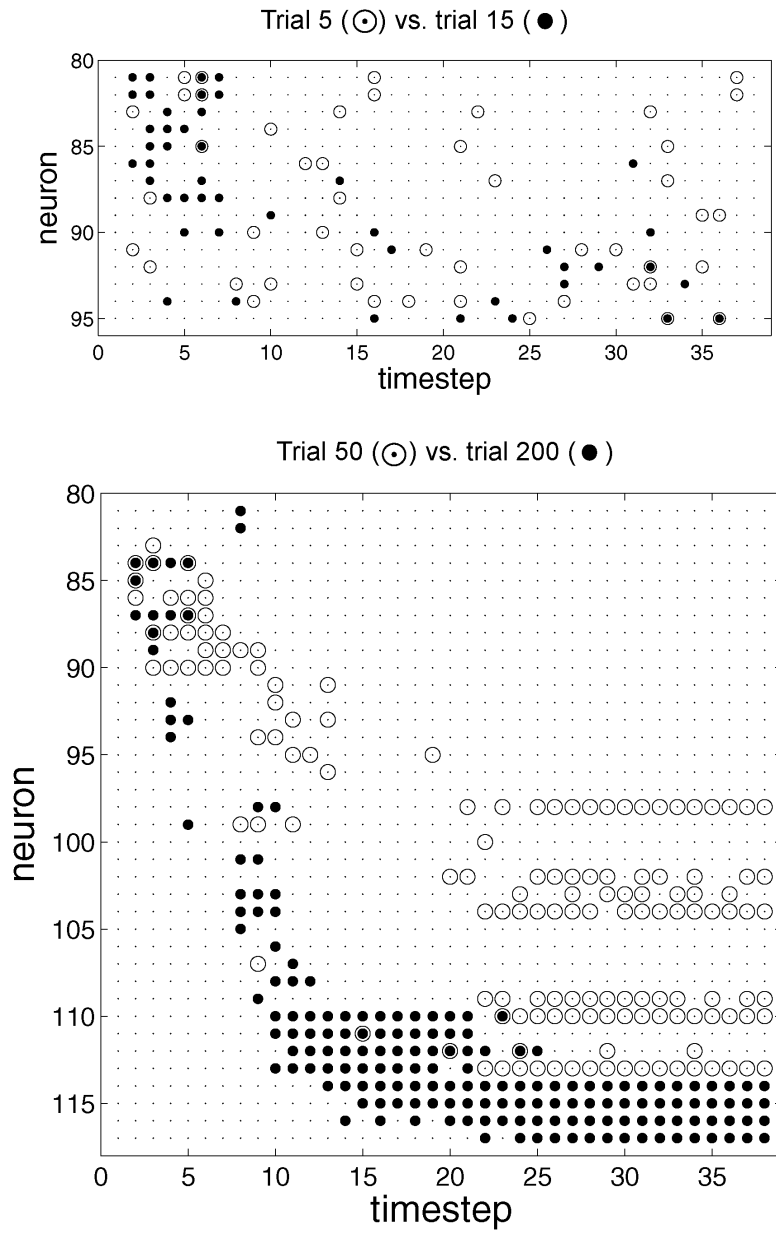


Fig. 5.

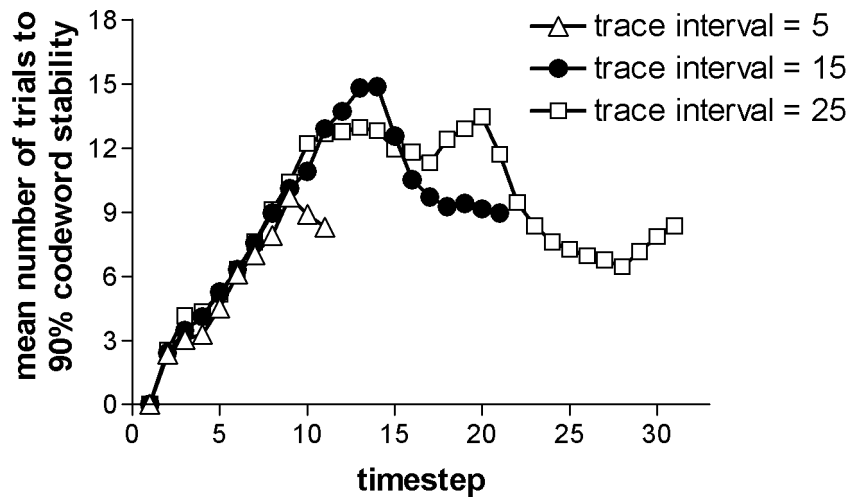
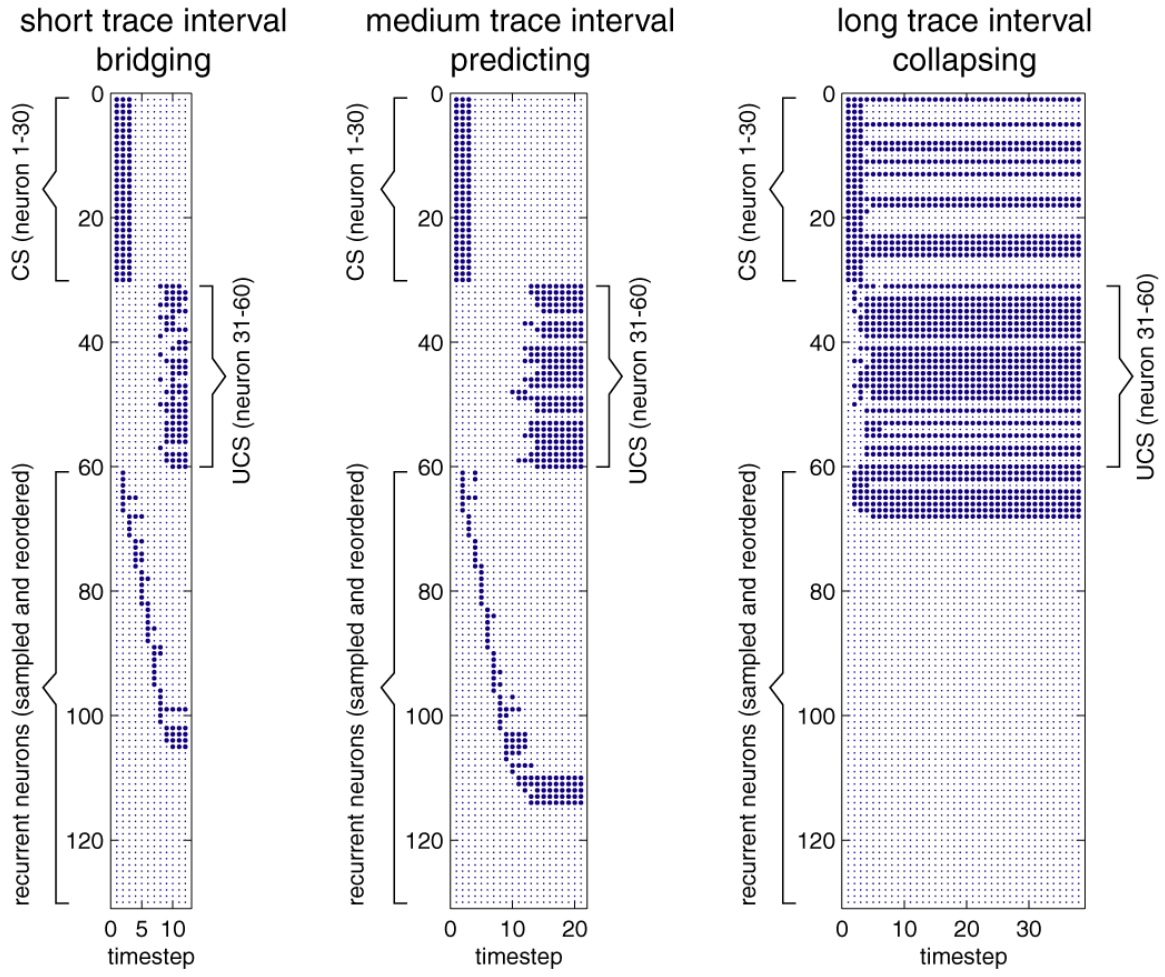


Fig. 6.



**Fig. 7.**

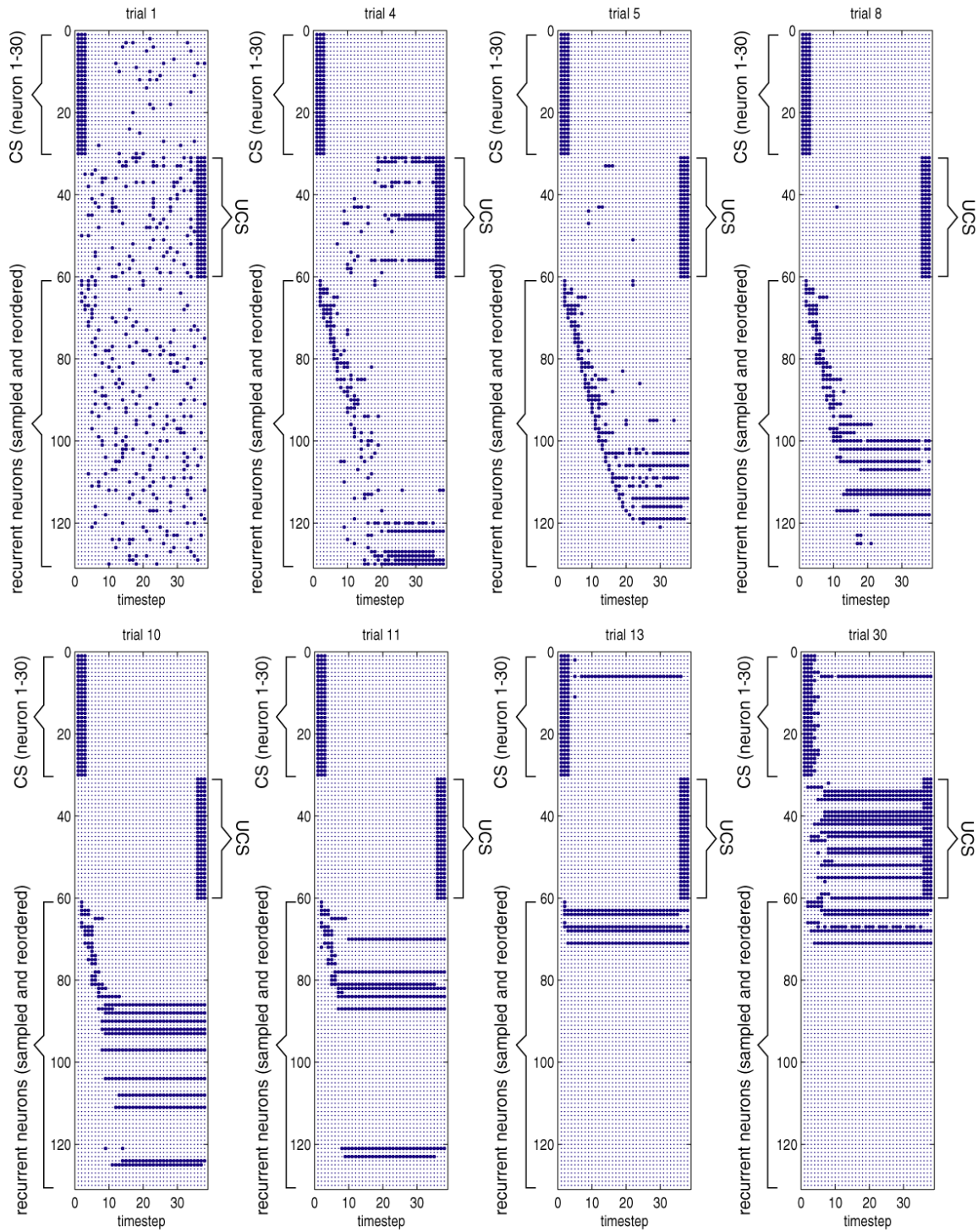
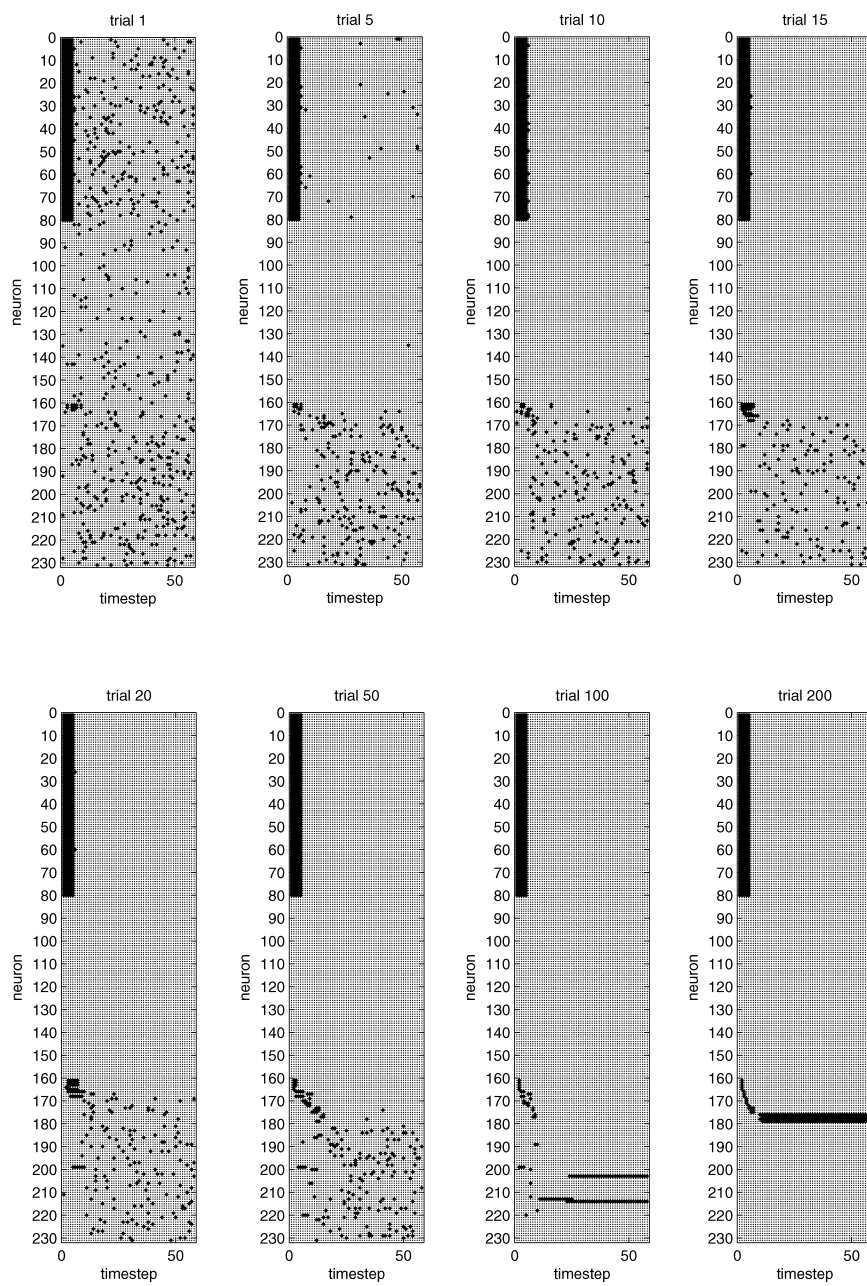
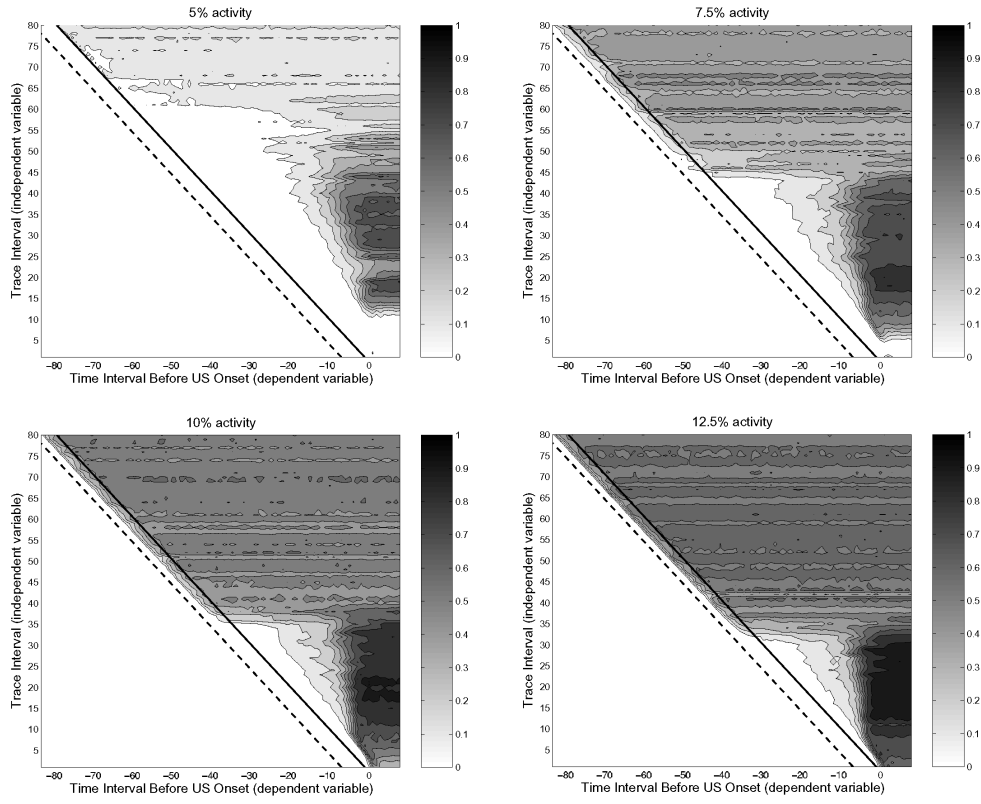


Fig. 8.



**Fig. 9.**



**Fig. 10.**

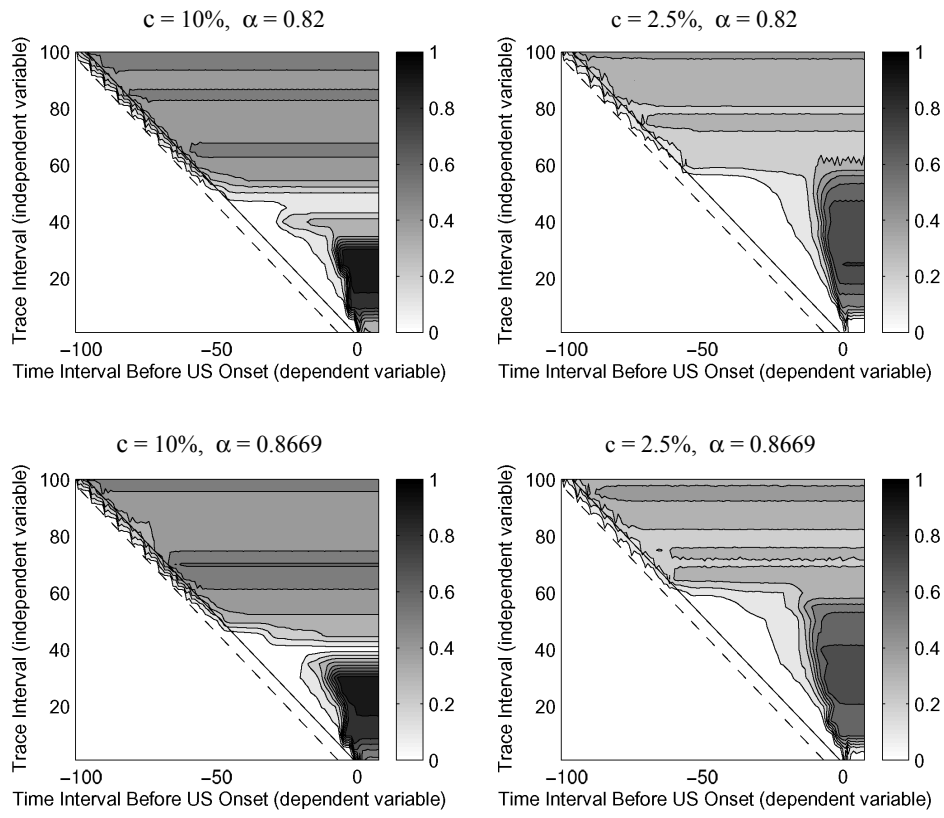


Fig. 11.

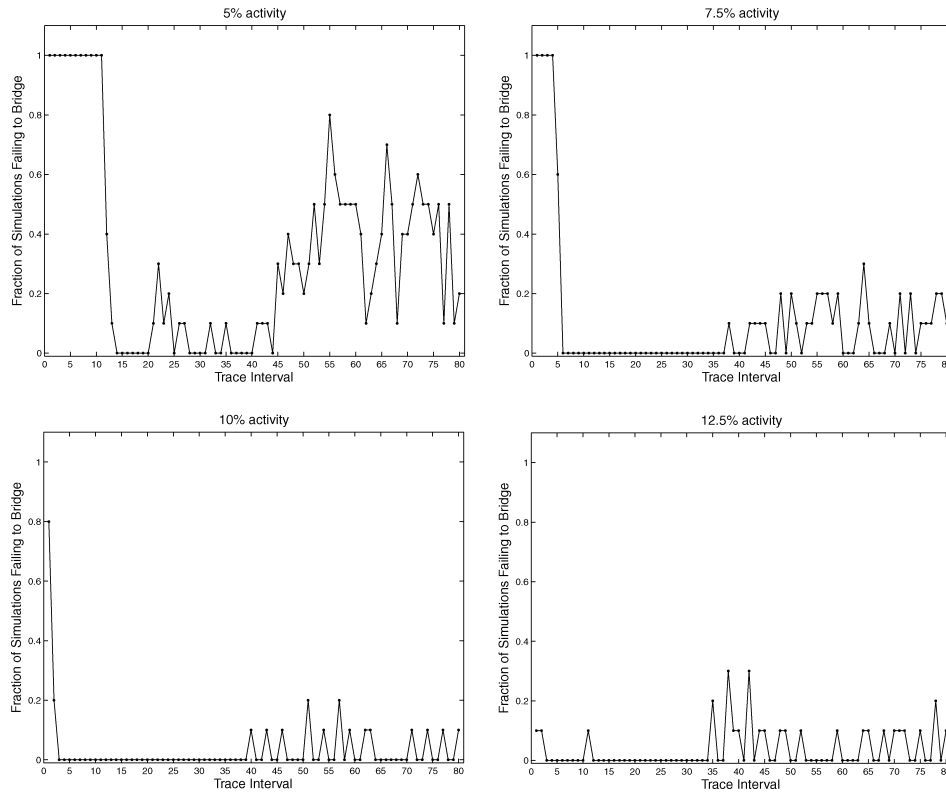


Fig. 12.

