Neuronal dynamics during the learning of trace conditioning in a CA3 model of hippocampal function

Blake T. Thomas · William B Levy

Received: 17 May 2013 / Revised: 9 September 2013 / Accepted: 25 September 2013 / Published online: 22 October 2013
© Springer Science+Business Media Dordrecht 2013

Abstract The present article develops quantitative behavioral and neurophysiological predictions for rabbits trained on an air-puff version of the trace-interval classical conditioning paradigm. Using a minimal hippocampal model, consisting of 8,000 primary cells sparsely and randomly interconnected as a model of hippocampal region CA-3, the simulations identify conditions which produce a clear split in the number of trials individual animals should need to learn a criterion response. A trace interval that is difficult to learn, but still learnable by half the experimental population, produces a bimodal population of learners: an early learner group and a late learner group. The model predicts that late learners are characterized by two kinds of CA-3 neuronal activity fluctuations that are not seen in the early learners. As is typical in our minimal hippocampal models, the off-rate constant of the *N*-methyl-D-aspartate receptor gives a timescale to the model that leads to a temporally quantifiable behavior, the learnable trace interval.

Keywords Neural network · Trace conditioning · Hippocampus · Sequence learning

Introduction

Our model of the hippocampus (Levy 1989, 1996; Levy et al. 1990) explains the role of this structure in memory formation

via its tendency to encode particular memories along with the context(s) in which they are experienced. In this model, successful context encoding implies an ability to learn sequences because much of any context is formed by the events leading up to the present, as well as the present itself. Such a sequence encoding capability implies that a hippocampus—and a model of the hippocampus—should be able to perform predictions. Because of a temporal compression that is part of our model (Levy 1989), such predictions can be delivered in a timely manner, i.e., before the occurrence of the actual event being predicted. In regard to such predictions, we use the hippocampal-dependent, trace-conditioning paradigm to understand how neural variables affect behaviorally manifested predictions and affect temporal compressions of encoded events (see August and Levy 1996, 1999 for compression in the sleeping animal as opposed to the awake animal simulated here).

Hippocampal lesions ascribe a critical role to the hippocampus in acquiring a conditioned response (CR) in many different training paradigms. Bilateral lesions of the hippocampus destroy the ability to form new memories while past memories remain intact (Scoville and Milner 1957). By varying the time of such a lesion, similar studies (Solomon et al. 1986; Moyer et al. 1990) show that trace conditioning serves as a suitable paradigm for these clinical observations concerning this difference between retrograde versus orthograde memory disturbances.

The trace conditioning paradigm might be one of the simplest sequence learning paradigms. In general, the paradigm takes the following form (see Fig. 1): after an initial stimulus (the conditioned stimulus, CS), there are no external stimuli for an interval (the trace interval), and then a behavior-evoking stimulus (the unconditioned stimulus, US) is delivered. After sufficient training (e.g. Solomon et al. 1986), a rabbit learns to move its nictitating membrane (the

B. T. Thomas
Informed Simplifications, LLC 520 Panorama Road, Earlysville,
VA 22936, USA

W. B. Levy (✉)
Department of Neurosurgery and Psychology,
University of Virginia, Charlottesville, VA 22908, USA
e-mail: wbl@virginia.edu

Fig. 1 The trace conditioning paradigm. **a** Example of CS and US inputs in a 25 timestep (500 ms) trace interval training trial for the 1,000 neuron network. Neurons 1–32 are CS neurons and are activated on timesteps 1–5 (the first 100 ms). Neurons 33–64 are US neurons and are activated on timesteps 31–35 (the last 100 ms). There are no external activations on timesteps 6–30, the 500 ms trace interval. Neurons 65–1,000 are not shown. **b** Example of a test trial for a 25 timestep (500 ms) trace interval. Neurons 1–32 are CS neurons and are activated on timesteps 1–5 (the first 100 ms). Neurons 33–64 are US neurons and are never externally activated during a test trial

conditioned response, CR) before the arrival of US, an air-puff in this case. The stimulus-free period between the CS offset and the US onset is called the trace interval and is an independent experimental variable both in the behavioral studies and in our simulations. In previous studies, we have investigated how parameterization of biophysical variables such as network activity, network connectivity, synaptic failure rate, synaptic modification rate, and the synaptic modification timing rule affect the learnable trace interval (Rodriguez and Levy 2001; Sullivan and Levy 2004; Levy et al. 2005; Howe and Levy 2007). We have also investigated the effect of behavioral variables such as CS length and US length (Wu and Levy 2005).

In this paper, we continue our quantitative studies of trace conditioning and the cell firing dynamics that accompany learning or failure to learn the prediction of the US (Levy and Sederberg 1997). We refine the results of Howe and Levy (2007) that reveal characteristic failure modes when a trace interval is too long to learn. With the current report, we hope (1) to encourage experimentalists to test the viability of our model and (2) to provide further details to computational modelers in order to help them reproduce and extend our results.

Experimental research using rabbits seems to favor a 500 ms trace interval although reports exist that rabbits can

learn traces as long as 750 ms (Smith et al. 1969). Using 100 ms as the off-rate time constant of the *N*-methyl-D-aspartate receptor (NMDA-R), we provide a timescale to our simulations. Previously, we found that trace intervals longer than 600 ms are progressively more difficult to learn, and we documented a variety of unusual and distinctive activity swings that occur in many of the simulations for trace intervals around 700 ms (Howe and Levy 2007). These characteristic activity swings include total activity compared across individual trials as well as activity of US-coding neurons.

Here we concentrate on the 600 ms trace interval for at least two reasons: first, experimentalists will want to study a behavioral situation in which most subjects learn (as opposed to a 700 ms trace interval in which most subjects fail to learn); second, at the 600 ms trace interval, we predict a clear split of the subject population regarding the number of trials before a reliable CR; this clear split should be easy to observe in behavioral experiments. We will refer to these two groups as early and late learners because of the number of trials required for acquisition of the US prediction (the hippocampal precursor of the CR). The neuronal activity patterns of the two “behavioral” groups are quite distinct. The late learners show similar qualitative features of the long trace

interval simulations characterized in Howe and Levy (2007). In particular, the late learners here show a characteristic oscillation in total activity across trials and a certain fluctuation of US-coding neurons before establishing a learned response.

Methods

Network

All but two neurons in the model are McCulloch-Pitts elements with a binary output of one or zero on each timestep (primary neurons, see Fig. 2 for a detailed schematic). Postsynaptic summation of inputs and spike generation occur in a single timestep. Axonal transmission of the spike takes one timestep. Connections ($c_{ij} \in \{0, 1\}$) between primary neurons are sparse and always excitatory. The weights (w_{ij}) for these connections are all initialized to 0.4 and are limited, via their modification equation, to the range (0,1). Excitation of neuron j is calculated by the following equation:

Fig. 2 The CA3 model. A schematic representation of a subset of our CA3 model. Excitatory neurons (*triangles*) are recurrently connected with excitatory synapses (*filled circles*). Each neuron can receive an external input (*filled squares*), which forces neurons to fire. The inhibitory neurons (*ovals*) help maintain network activity. Inhibitory neurons have output on $[0, \text{inf}]$. The interneuron output is divisive, so inhibitory synapses (*unfilled circles*) appear on excitatory neurons (figure and legend reprinted from Fig. 2 Howe and Levy 2007)

$$y_j(t) = \frac{\sum_{i=1}^n Z_i(t-1)C_{ij}W_{ij}(t-1)}{\sum_{i=1}^n Z_i(t-1)C_{ij}W_{ij}(t-1) + K_{FB} \sum_{i=1}^n Z_i(t-1)W_{il}(t-1) + K_{FF} \sum_{i=1}^n x_i(t) + K_0}$$

See Table 1 for parameter settings

Activity of the network is modulated by two inhibitory neurons (see Fig. 2); this modulation appears as the second and third summations in the divisive inhibition of the $y_j(t)$ equation (see Smith et al. 2000). The values for inhibition, K_{FB} , K_{FF} , and K_0 (all greater than zero), are set such that random external activations on timestep zero produce activity near the desired level on the subsequent timesteps of the first trial. The constants K_{FB} and K_{FF} scale the influence of the feedback and feedforward inhibitory neurons, respectively. The feedback inhibitory neuron scales a weighted sum of the network's activity on the previous timestep:

$$K_{FB} \sum_{i=1}^n Z_i(t-1)w_{il}(t-1)$$

The feedforward inhibitory neuron scales a sum of the externally activated neurons ($x_i \in \{0, 1\}$) on the current timestep:

$$K_{FF} \sum_{i=1}^n x_i(t)$$

Each neuron that has excitation greater than inhibition generates an output pulse. In addition, all neurons that

are externally activated fire regardless of inhibition; formally,

$$Z_j(t) = \begin{cases} 1 & \text{if } y_j(t) \geq 0.5 \text{ or } x_j(t) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Table 1 Parameter values

n (number of primary neurons)	1,000	2,000	8,000
CS neurons	32	64	80
US neurons	32	64	80
Fractional connectivity	0.1	0.1	0.1
a (desired fractional activity)	0.1	0.08	0.05
$w_{ij}(t=0)$	0.4	0.4	0.4
$w_{il}(t=0)$	1	1	1
K_0 (initial value)	0.78	1.175	1.05
K_{FB}	0.0477	0.045	0.0467
K_{FF}	0.019	0.0028	0.0064
λ	0.5	0.5	0.5
ε	0.1	0.1	0.1
μ	0.0015	0.002	0.005
α	0.819	0.819	0.819

The weights from primary neurons to the feedback inhibitory neuron are all initialized to a value of one and are modified on every timestep according to the rule:

$$w_{il}(t+1) = w_{il}(t) + \lambda z_i(t+1) \left(\frac{\sum_{k=1}^n z_k(t)}{n} - a \right)$$

where a is the desired activity of the network (Sullivan and Levy 2003). These weights are constrained to the interval $[0, \infty)$. For best activity control, the value of K_0 is modified after every training trial according to the following rule:

$$\Delta K_0 = \varepsilon \left(\frac{\sum_{t=1}^T \sum_{i=1}^n z_i(t)}{nT} - a \right)$$

where T is the total number of timesteps in a trial. Note that time does not appear on the left-hand side of this equation because this modification occurs between training trials. However, the observed range of K_0 modulation is narrow (e.g. 1.035–1.051 in an 8,000 neuron 600 ms trace simulation), and the results presented below obtain in the absence of this modulation.

Weights between primary neurons are modified according to an NMDA-R-like learning rule: $w_{ij}(t+1) = w_{ij}(t-1) + \mu z_j(t)(\bar{z}_i(t-1) - w_{ij}(t-1))$

We use 100 ms as the time-constant for e-fold exponential decay of $\bar{z}_i$ (NMDA-like decay of glutamate synaptic activation):

$$\bar{z}_i(t-1) = \begin{cases} 1 & \text{if } z_i(t-1) = 1 \\ \alpha \bar{z}_i(t-2) & \text{otherwise} \end{cases}$$

The 100 ms is a nominal value used from the literature (Lester et al. 1990; Traynelis et al. 2010). This nominal 100 ms defines time for each timestep of the simulation. Because α is set to $e^{(-1/5)}$, a timestep corresponds to 20 ms.

To find the parameter settings (see Table 1), the following sequential procedures are performed recursively: (1) the rate constants μ , λ , and ε are set to 0, while K_0 , K_{FF} , and K_{FB} are manipulated with the goal of small timestep to timestep activity oscillations around the desired level; (2) determine a value of ε so that average trial activity has small oscillations around the desired level for all 200 trials; (3) determine a value of μ and λ so that average trial activity has small oscillations relative to desired trial activity across 200 trials (this is our typical sequence of 200 training trials; (4) manipulating μ and λ , the 500 ms trace interval task has at least as good performance as any other trace interval, and across all seeds at the 500 ms trace interval, the median trial with the first correct US prediction is as close to trial 120 as possible. The trials to learn are inspired by McEchron and Disterhoft 1997 data (~ 160 trials to learn), not the Solomon et al. 1986 data (~ 500 trials) or the Moyer et al. 1990 data (>700 trials). Apparently, subtle differences between animal paradigms lead to different learning rates.

Network initialization

Three network sizes are studied: $n = 1,000$, 2,000, or 8,000 (primary neurons).

Each simulation of the model is made distinct by its pseudo-randomly generated connectivity matrix constrained to a fixed number of input connections per neuron. There are 200 seeds for the 1,000 and 2,000 neuron simulations; there are 100 seeds for the noise-free 8,000 neuron simulations and 50 for the external noise simulations. One simulation is analogous to one animal of a behavioral study; we will refer to a specific simulation by its trace interval, seed number, and number of neurons, e.g. 700 ms trace interval, seed 7, 8,000 neurons. Because we use the same set of random number seeds for each trace interval, we can make a “within-subjects” comparison across trace intervals.

At the beginning of each simulated trial (nominal timestep zero), a subset of neurons is pseudo-randomly activated. Neurons representing the CS or the US are excluded from this random activation.

Training and testing

A training trial is composed of a sequence of inputs as in Fig. 1a; the CS length is 100 ms (five timesteps); the trace length varies from 300 to 1,000 ms (fifteen to fifty timesteps); the US length is 100 ms (five timesteps). The number of neurons assigned to the CS and US vary along with network size and desired activity (see Table 1). On the first through fifth timestep of each training trial, all the CS neurons are externally activated. On the last five timesteps of each training trial, all the US neurons are externally activated.

A test trial is composed of a sequence of inputs as in Fig. 1b. There are two differences between a training and a test trial. First, a test trial does not have any synaptic modifications of excitatory or inhibitory weights. Second, the US neurons are *not* externally activated on a test trial.

In sum, a full simulation is composed of several steps. First, a network’s connections are pseudo-randomly generated. Second, a pseudo-random stimulus is generated for the start of each of the 200 trials. Third, one training trial and one test trial alternate for a total of 200 trials of each. Note that each pair of training and test trials uses the same random stimulus at the beginning of the trial.

Noise

External noise is introduced in some simulations. All of these simulations have 8,000 neurons. On each timestep during the trace interval, each non-CS/non-US neuron has a pseudo-random $\{0,1\}$ chance of firing. This percentage is determined by the activity level of the network and the desired fraction of activity caused by this noise. Because

noise is injected pseudo-randomly for each neuron on each timestep, some timesteps will have more or less than the average value of random external activation.

Behavioral decoding/visualization

As in our earlier work (Howe and Levy 2007), a successful prediction in a test trial (though somewhat arbitrary) is defined relative to a 140 ms time window (beginning 200 ms before the US onset and ending 60 ms before the same onset). Appropriate firings of US neuron are restricted to this time window. For a correct prediction, at least 30 % of the US neurons must fire on any timestep within this window, while less than 30 % of the US neurons are allowed to fire on any timestep preceding the onset of this window.

As a result there are three failure modes and one success mode. The success mode is defined above. A predict-too-early failure results from greater than 30 % of the US neurons firing on any timestep preceding the acceptable time window (200 ms before the US onset). A predict-too-late failure is defined as at least 30 % of US neurons firing on any timestep after the window ends, but not on any timestep during or before the window. A failure to predict results from fewer than 30 % of US neurons firing on every timestep.

There are multiple ways to visualize the success or failure of a simulation. Here we use a three dimensional display (timestep by test trial by average US activity). We also show US neuron firings of test trial 200. In both such figures, the vertical black bars demarcate the time window that defines a successful US prediction.

Stable correct prediction

A *stable* correct prediction is defined as a correct prediction that lasts until trial 200. The first trial with a stable correct prediction is therefore the earliest trial of these *consecutive* correct predictions.

Results

The main result of this study is the discovery of a trace interval that produces a bimodal time-to-learn, illustrated in Fig. 9; accompanying this result are dynamics specific to each of the two populations. In order to appreciate the context of these results, we first present simpler dynamical characterizations and performance evaluations for a range of learnable and not learnable trace intervals. Then, we more closely examine the dynamics of simulations on the edge of learnable trace intervals. Finally, we describe the effect of one type of noise.

Success and failure modes

One of four modes describes the performance of each simulation: (1) predict the US successfully, (2) predict the US too early, (3) predict the US too late, and (4) never predict the US. Because learned performance can be judged solely from the firing pattern of US neurons, the three-dimensional US-neuron activity plot (timestep by trial number by fraction of US neurons firing) completely details the performance of individual simulations.

Successful prediction

A typical simulation that successfully predicts the US is shown in Fig. 3a for an 8,000 neuron simulation learning a 500 ms trace interval. In this example, the US neurons begin their first learned firings late within trial 103; eventually a correct prediction develops on trial 112 with 31.25 % of the US neurons firing on timestep 26, which is within the required time-window for a correct prediction. This correctly timed, net firing pattern is repeated for every trial from 112 to 200. The detailed US neuron firing patterns from this same simulation on test trial 200 show each individual US neuron during the last trial (Fig. 3b). While a few US neurons fire before the window begins, almost every US neuron begins firing one or two timesteps before the end of the window, thus resulting in a successful prediction.

The probability of successful prediction varies as a function of network size and as a function of trace interval duration. Starting with the 1,000 neuron simulations, success rates are 0.205 for the 300 ms trace interval, and 0.885, 0.745, 0.37, 0.255, 0.11, 0.02, and 0.005 for the 400 ms through 1,000 ms (in increments of 100 ms) trace intervals, respectively (Fig. 3c). The 2,000 neuron simulations have success rates of 0.005, 0.915, 1.0, 0.715, 0.675, 0.45, 0.185, and 0.005 for trace lengths of 300–1,000 ms, respectively. The 8,000 neuron simulations has success rates of 1.0, 1.0, 0.93, 0.56, 0.2, 0.01, 0, and 0 for trace lengths of 300 ms to 1,000 ms, respectively.

Predict too early

A typical simulation that predicts the US too early (that is, earlier than 200 ms before the onset of the US) is shown in Fig. 4a for an 8,000 neuron simulation learning a 700 ms trace interval. In this example, the US neurons begin their first learned firings late within trial 83. Eventually and transiently, a correct prediction develops on trial 123 with 33.75 % of the US neurons firing 140 ms before the onset of the US (timestep 34) and less than 30 % firing any time before the acceptable time-window. However, this correctly timed firing pattern only lasts through trial 127.

Fig. 3 Successful learning.

a The activity of US neurons as a function of training trial number (y-axis) and of within trial timestep (x-axis) from an 8,000 neuron, 500 ms trace simulation. The US neurons begin firing near trial 110. This example is a successful simulation because during the last ten trials >30 % of the US neurons are firing on any timestep between timesteps 21 and 27 as demarcated by the black bars. **b** Individual US neuron firings on test trial 200 from **a**. Note that some US neurons are firing ‘too-early’, but this firing does not reach the 30 % threshold. **c** Proportion of successes as a function of network size and trace length. Error bars are from a matched sign test at 95 % confidence

Starting on trial 128 and lasting until trial 200, the US neurons are firing above the 30 % threshold before the time-window for a successful prediction, thus resulting in a predict-too-early failure. Figure 4b shows the detailed US neuron firing pattern of this same simulation on test trial 200. Here, most of the US neurons fire before the window defining a correct prediction.

The probability of a too-early US prediction varies as a function of network size and as a function of trace interval duration. In general, too-early predictions assert themselves at the longer, but not the longest, trace intervals. Starting with the 1,000 neuron simulations, early failure rates are 0 for both the 300 ms and 400 ms trace interval. The predict-too-early failure rates for the 500 ms through 1,000 ms trace intervals are 0.07, 0.26, 0.175, 0.17, 0.075, and 0.025, respectively (Fig. 4c). The 2,000 neuron simulations have predict-too-early failure rates of 0, 0, 0, 0.28, 0.28, 0.45, 0.4, 0.23 for trace lengths of 300 ms to 1,000 ms, respectively. The 8,000 neuron simulations have predict-too-early failure rates of 0, 0, 0.05, 0.37, 0.8, 0.76, 0.27, and 0.04 for trace lengths of 300 ms to 1,000 ms, respectively.

The 1,000 neuron simulations produce the most too-early failures at the 600 ms trace interval. The 2,000 neuron simulations produce the most too-early failures at the 800 and 900 ms trace intervals. The 8,000 neuron simulations have even more too-early failures. They have the most at the 700 ms and 800 ms trace intervals. No

simulations produced too-early failures at the 300 or 400 ms trace intervals.

Predict too late

A typical simulation that predicts the US too late (i.e. after the acceptable window) is shown in Fig. 5a for an 2,000 neuron simulation learning a 300 ms trace interval. In this example, the US neurons begin their first learned firings above the 30 % threshold on trial 147. Beginning on trial 162, and lasting until trial 200, there is a stable firing of US neurons above the 30 % threshold. However, an appropriately timed prediction never develops since the US neuron firings are always after the defined window, thus resulting in a predict-too-late failure. The detailed US neuron firing patterns from this simulation on trial 200 show each individual US neuron during the last test trial (Fig. 5b). Here, we see that some, but not enough, US neurons fire before the window ends; then, after the window ends, more than half of the US neurons begin to fire.

The probability of a too-late US prediction varies as a function of network size and as a function of trace interval duration. In general, it is a common failure mode for the shortest trace interval studied here, and thereafter it is rather rare. Starting with the 1,000 neuron simulation, predict-too-late failure rates are 0.795 for the 300 ms trace interval, and 0.1, 0.03, 0.035, 0.045, 0.065, 0.03, and 0.005 for the 400 ms through 1,000 ms trace intervals,

Fig. 4 Failed learning—Predict-too-soon. **a** The activity of US neurons as a function of training trial number (y-axis) and of within-trial timestep (x-axis) from an 8,000 neuron, 700 ms trace simulation. The US neurons begin firing near trial 90. This example is a predict-too-early failure because during the last ten trials >30 % of the US neurons are firing on any timestep before timestep 31 as demarcated by the leftmost black bar. **b** Individual US neuron firings on test trial 200 from **a**. Note that most US neurons are firing ‘too-early’; that is, before timestep 31. **c** Proportion of predict-too-soon simulations as a function of network size and trace length

respectively (Fig. 5c). The 2,000 neuron simulations have predict-too-late failure rates of 0.995, 0.085, 0, 0, 0, 0.035, 0.02, 0.005, for trace lengths of 300 ms to 1,000 ms, respectively. The 8,000 neuron simulations never produces a too-late prediction failure.

The 1,000 neuron simulations have a majority of early failures at the 300 ms trace interval. Almost every 2,000 neuron simulation produced an early failure at the 300 ms trace interval (199 out of 200).

No prediction

A typical simulation that never predicts the US, correctly or otherwise, is shown in Fig. 6a for an 8,000 neuron simulation learning a 1,000 ms trace interval. In this example, the US neurons never begin any learned firings, and, therefore, never generate any US prediction. Of course the detailed US neuron firing patterns from this simulation on test trial 200 (Fig. 6b) shows no activity.

The probability of the no-prediction mode varies as a function of network size and as a function of trace interval duration. In general, it becomes more and more common as the trace length is increased. Starting with the 1,000 neuron simulations, no-prediction rates are 0 for the 300 ms trace interval, and 0.015, 0.155, 0.335, 0.525, 0.655, 0.875, and 0.965 for the 400 ms through 1,000 ms trace intervals, respectively (Fig. 6c). The 2,000 neuron simulations have no-prediction rates of 0, 0, 0, 0.005, 0.045, 0.065, 0.395,

and 0.76 for trace lengths of 300 ms to 1,000 ms, respectively. The 8,000 neuron simulations have no-prediction rates of 0, 0, 0.02, 0.07, 0, 0.23, 0.73, and 0.96 for trace lengths of 300 ms to 1,000 ms, respectively.

The 1,000 neuron simulations have the most no-prediction failures, with more at longer trace-lengths. The 2,000 neuron simulations have the fewest no-prediction failures and also have a trend for more at the longer trace-lengths. The 8,000 neuron simulations also have a trend for more no-prediction failures at the longer trace-lengths.

The synaptic modification rate constant (μ) affects success and failure rates

As might be expected, μ controls the number of trials to develop learned US firings, but in addition, the success and failure modes of the simulations are affected by this rate constant. Setting μ to maximize the fraction of simulations that create successful predictions is dependent on a variety of network and simulation parameters including number of neurons, average firing rate, fractional connectivity, and the total number of training trials per simulation (Levy et al. 2005, Howe and Levy 2007). In the past we have limited our reports to simulations derived from empirically optimized values of μ . Here however, our report includes a few sub-optimal values. In general, success rate as a function of μ has an inverted U-shape when total trials are limited. For the 8,000 neuron simulations the success rates at a trace

Fig. 5 Failed learning—predict-too-late. **a** The activity of US neurons as a function of training trial number (y-axis) and of within-trial timestep (x-axis) from a 2,000 neuron, 300 ms trace simulation. This example is a too-late failure because during the last ten trials >30 % of the US neurons are not firing on any timestep between timesteps 11 and 17. The US neurons do not reach the 30 % threshold until timestep 20 as demarcated by the rightmost bar, which, in our model, is not early enough for a useful prediction. **b** Individual US neuron firings on test trial 200 from **a** Note that most US neurons are firing ‘too-late’; that is, after timestep 17. **c** Proportion of predict-too-late failures as a function of network size and trace-length. This failure mode only occurs in the 1,000 neuron and 2,000 neuron networks

Fig. 6 Failed learning—no-prediction. **a** The activity of US neurons as a function of training trial number (y-axis) and of within-trial timestep (x-axis) from an 8,000 neuron, 1,000 ms trace simulation. This example is a no-prediction failure because US neurons never reach the 30 % threshold on any timestep. **b** Individual US neuron firings on test trial 200 from **a** Note that the US neurons never fire. **c** Proportion of no-prediction failures as a function of network size and tracelength. The longer trace intervals are most prone to the no-prediction failures

interval of 600 ms are 0 for a μ of 0.0,025, and 0.44, 0.56, 0.48, and 0.02 for μ of 0.004, 0.005, 0.0075, and 0.01, respectively (Fig. 7a). The fraction of simulations that

predict too early monotonically increases with μ and are 0, 0.12, 0.37, 0.52, and 0.98 for μ of 0.0025, 0.004, 0.005, 0.0075, and 0.01, respectively (Fig. 7b). From all

simulations across all values of μ , there is one predict-too-late failure at a μ of 0.0025 (Fig. 7c). Finally, the fraction of simulations that do not produce any prediction at a trace interval of 600 ms monotonically decreases and rapidly approaches zero at μ of 0.005. The no prediction rates are 0.98, 0.44, 0.07, 0, and 0 for μ of 0.0025, 0.004, 0.005, 0.0075, and 0.01, respectively (Fig. 7d).

A μ of 0.005 produces the best performance at the 600 ms trace interval. Setting μ higher increases early prediction failures, while lower μ increases no prediction failures. There are very few too-late failures. It is important to note that μ of 0.0025 and 0.004 may be capable of better performance if the number of trials are allowed to increase. However, increasing the number of trials, from 200 to 250, for a μ of 0.005, does not increase success rates. Based on this optimal performance, μ is fixed at 0.005 in what follows.

Dynamics

Forming a stable and correct prediction

The minimum trial number before the first stable correct prediction (see “Methods”) has some surprising characteristics for the 8000 neuron simulations. There is a U-shaped effect, across trace lengths, in the mean number

of trials until a stable prediction (Fig. 8a). The 300 ms interval takes, on average, 163 trials until a stable, correct prediction is formed. Similarly, 134 trials elapse before a correct prediction at 400 ms, 110 trials elapse at 500 ms, 146 trials elapse at 600 ms, 161 trials elapse at 700 ms, and 186 trials elapse at 800 ms (There are no successful simulations at 900 or 1,000 ms). However, underlying these means, particularly at the longer trace intervals, is an unexpected observation.

The first hint of a higher-order dynamic appears at the 600 ms trace interval; here there is a very large variance of the number of trials to a stable, correct prediction. The standard deviation at 600 ms is exceptionally large with a value of 42.7 trials; the next largest standard deviation, at the 700 ms trace interval, is 5.8 trials; and all the rest are less than 3.7 trials (Fig. 8b). This relatively large standard deviation at 600 ms begs for deeper analyses (the 800 ms data point did not have a variance because there is only one successful prediction).

A histogram of trials with the ‘first stable correct prediction’ (see “Methods”) reveals the underlying details that produce the large variance at the 600 ms trace interval (Fig. 9). Notice that there is one cluster of 25 simulations centered around trial 99 and another cluster of 31 simulations centered around trial 186. Surprisingly, not a single simulation discovers its first stable, correct prediction between trials 110 and 168. To examine more closely why

Fig. 7 The effect of synaptic modification rate (μ) on success and failure modes at the 600 ms trace interval, $n = 8,000$.

a $\mu = 0.005$ produces the highest success rate of the five values tested. A value of $\mu = 0.0025$ produces no successes, and $\mu = 0.01$ produces a success rate of nearly zero. **b** Larger values μ tend to have more “too-early” failures. **c** Only $\mu = 0.0025$ has at least one predict-too-late failure. The smaller μ values tend to have more no-prediction failures; $\mu = 0.0075$ and 0.01 results in zero no-prediction failures. **d** Error bars are from a matched sign test at 95 % confidence. Number of simulations: $n = 50$ for μ of 0.0025, 0.004, 0.0075, and of 0.01; $n = 100$ for $\mu = 0.005$

Fig. 8 Trace length affects first stable, correct prediction.

a Average number of trials until a stable, correct prediction. A trace interval of 500 ms requires the fewest trials to create a stable, correct prediction. The 800 ms interval requires the most trials to create a stable, correct prediction for these 8,000 neuron simulations.

b Variability of trials to learn. Note the distinctively large standard deviation from the 600 ms trace interval simulations, which is analyzed further in the next figure

two exclusive learning modes emerged, we quantify the number of US neurons firing on each trial of every simulation.

US neuron dynamics

The three dimensional US neuron activity plot (e.g. Figs. 3, 4, 5,6a) is designed to visualize learned performance, which is defined around the arbitrary 30 % threshold. However, these plots need to be supplemented when trying to understand the dynamic that occurs just before the formation of such strong US codes. In what follows we total the number of US neurons that fire at least once on each test trial. With this visualization, a higher order dynamic and phase transition is immediately revealed.

For the early trials, this total US-firing measurement quickly decreases from about 65 to nearly zero. The trial number when this value starts increasing implies that the network finds a path in neuron-firing-space from the CS neurons to some US neurons. The US neuron activity that produces this increased firing is called “learned US firing” (note that a high value of this calculation does not constitute a correct prediction, as these US neurons could be firing too early, too late, or not enough on the same time-step to attain the 30 % threshold).

A second unexpected observation is the small variance in the trial number when US neurons begin learned firing. For example (Fig. 10) with a 600 ms trace and 8,000

Fig. 9 First trial with a stable, correct prediction for the 600 ms trace interval. A bimodal histogram underlies the large variability at the 600 ms trace interval noted in Fig. 8. One group of simulations learns between trial 94 and trial 109 (median = 99, $n = 25$). Another group learns between trial 169 and trial 192 (median = 186, $n = 31$). No simulations begin a stable, correct prediction between trials 110 and 168. These results are for the 8,000 neuron simulations

neurons, 98 % of the simulations activate at least three US neurons on at least one trial between trials 83 and 95. This low variance is a function of the number of neurons in the simulation. With 1,000 neurons, the phenomenon of Fig. 10 is virtually unobservable. In the simulations with

2,000 neurons, the US neurons do not decrease their firings down to zero before they start to increase. In these simulations it is possible to see that the learned US neuron firings are rising around the same time, but it is impossible to quantify the trial at which the rise occurs because the initial decrease of US activity is not large enough. Thus, to study the higher order dynamics of US firings, the remainder of the paper is restricted to the simulations with 8,000 neurons.

Returning to Fig. 10, there are other characterizations that also display a surprising amount of synchrony across simulations. One of these is the plateau region that spans about 30 trials that the first cluster of learned US firings display. Another characteristic is the existence of two later clusters of learned US firings. The simulations in these two clusters participate in the initial increase in learned US firing, but the US neurons return to lower level (often near zero) of activity after trial 110. The first of these two clusters has learned US firings centered around trial 138, and the second cluster has learned US firings centered around trial 160.

Learned US firings coincide with activity oscillations

Correlated with US code development is an emergent behavior: oscillations in the total average activity per trial for the longer trace intervals (here we focus our comments on the 600 ms interval). These brief, across trial oscillations begin in concert with the first rise in learned US firings. In the case of the 28 simulations that have increased learned US firing that never declines (the early fully rising group of Fig. 10), the oscillation has an amplitude of 0.0035 (desired activity is 0.05) and lasts for one cycle, which is about 15 trials (Fig. 11a). The latter groups of simulations ($n = 72$) that do not form early, stable US firings have more dramatic oscillations, while as before the onset of oscillations occurs with the initial rise in learned US firing. Figure 11b displays a single simulation that is a late learner. Note the large amplitude oscillations of two cycles typical of this group. Here, the amplitude is 0.075 (desired activity is 0.05) and a cycle lasts about 19 trials. This large amplitude, oscillatory phenomenon is replicated in every long-trace-interval simulation that does not

Fig. 10 Clustering of trial number in which US neurons begin learned firing. Number of US neurons fired on each trial for the first 50 simulations: 8,000 neurons, 600 ms trace interval (the second 50 simulations are omitted to make visualization easier). Virtually all simulations participate in the first, initial up-swing around trial 90. About half of these simulations reach 55 US neurons by trial 100. The other half of the simulations decrease the number of US neurons

firing. For this latter group, US neurons begin firing again. A y-axis value of 80 corresponds to all US neurons firing at some point during the particular test trial of the particular simulation. This plot ignores multiple firings by the same neuron within a trial. A value of 0 corresponds to no US neurons ever firing during a test trial of a simulation. All plotted values are non-zero after trial 235

Fig. 11 Activity fluctuations distinguish early and late learners. **a** A small, brief activity fluctuation accompanies the formation of US firings by a typical early learner. The per trial count of US neuron firing (as in Fig. 10) is superimposed on the average trial activity of this simulation, which succeeds in learning a 600 ms trace on trial 97. Beginning near trial 90, the small rise, fall, and rise of activity is typical of such simulations. This minor activity oscillation coincides with the earliest increase in learned US firings. Note the stability of activity once three-fourths of US neurons are participating in the code. **b** Larger and more repetitious activity oscillations accompany simulations with delayed learning. Again, the per trial count of US neuron firing is superimposed on the average trial activity of this

simulation, which succeeds in learning a 600 ms trace interval beginning on trial 172. The activity oscillation, beginning near trial 90 and lasting for two cycles, is typical of such simulations. Note the onset of the activity oscillations correlates with the first increase in learned US firings and ceases long before US neurons begin new learned firings on trial 155. Five percent noise is enough to shuffle predict-too-early and correct prediction simulations without significantly changing average performance. Adding noise, however, does change the performance of individual seeds. For example, at the 600 ms trace interval, there are 12 seeds succeeding without noise but failing when noise is added. Conversely, there are 13 seeds succeeding with noise but failing without noise

generate an early stable prediction; in all cases the onset of the oscillations, which last for two or more cycles, are always coincident to the initial rise in US firings.

We conjecture that these oscillations in activity are due to the network creating profound and novel changes in firing patterns in a single trial, changes which the feedback activity-control mechanisms (K_{FB} and K_0) are unable to match.

The observed trial-by-trial activity oscillations can also be visualized without averaging within trials. Rather, plots of activity at a particular timestep across trials reveal the same oscillatory phenomenon. In fact, such fixed-timestep across-trial oscillations are of even greater amplitude than the average trial activity oscillations illustrated in Fig. 11b.

Introducing noise

In an attempt to destroy this clustering of learned US firings while still producing successful learning, a small amount of external noise is introduced throughout the trace

interval for all trials (see “Methods”). A variety of noise levels are evaluated, but 5 % noise is used in what follows because it maintains performance levels. That is, 5 % noise affects the prediction mode of half the 50 simulations examined, being beneficial for half of these (13 out of 25) and disruptive for the other 12 (Fig. 12). Noise up to 10 % is compatible with successful predictions at similar rates to no noise for the 600 ms trace interval.

Effects on learned US firings

The addition of 5 % external noise (an average 20 neurons out of a nominal 400 per timestep) fails to disrupt the clustering of learned US firings (Fig. 13 compared to Fig. 10). This observation is also true for 12.5 % noise although the middle cluster is sparse while the first and third clusters are heavily populated. (The apparent tightening of the clusters at 5 % noise is not observed at the higher noise value.)

Fig. 12 Five percent noise is enough to shuffle predict-too-early and correct prediction simulations without significantly changing average performance. Adding noise, however, does change the performance of individual seeds. For example, at the 600 ms trace interval, there are 12 seeds succeeding without noise but failing when noise is added. Conversely, there are 13 seeds succeeding with noise but failing without noise

Fig. 13 External noise does not disperse clustering of learned US firings, but does delay their onset. US neuron firing behavior over 250 trials with 5 % external noise during the trace interval: 8,000 neurons, 600 ms trace interval. Note that the initial increase in US firing is nearly 20 trials later than in the previous figure. Later clustered increases are also delayed and appear somewhat tighter than the previous figure. Data plotted as in previous figure. All plotted values are non-zero after trial 204

Another observation worth noting is that noise delays the onset of learned US firings. This effect is approximately linear over the range of noise investigated (Fig. 14). For example, at the 600 ms trace interval, the first learned US firings without noise begin on trial 89. As noise increases from 2.5 % to 15 % (in increments of 2.5 %), the trial with

Fig. 14 External noise interacts with trace length to affect number of trials before a learned US firing. Longer trace-lengths require fewer trials before US neurons begin learned firing. Adding progressively more external noise during the trace interval slows down the learning process and delays the trial at which the US neurons begin learned firing

first learned US firings are 97, 104, 111, 120, 129, and 137, respectively.

Discussion

At the heart of our hippocampal research is a presumed mapping from the firing of particular neurons to a behavioral output. In the case of successful trace conditioning, we assume that hippocampal cell firing must anticipate the occurrence of the US by firing the neurons that represent the US in the model. The particular neurons of the model that code for the US are defined by external inputs much like the entorhinal cortex (EC) projection to CA3 and the dentate gyrus. In some not fully specified manner, but using the reciprocal connections between hippocampal-EC regions and the neocortex, we assume that intrinsic CA3 activation of these US-representing neurons will produce a neocortical representation of the US even without the presence of an actual US. Presumably, this neocortical representation can drive a behavioral response (CR), and it possibly can drive training of the cerebellum. The time window for a useful US prediction via such a CA3 representation is not known, so there is some arbitrariness to the time window used here to define a correct response. Likewise, the fraction of US neurons needed to fire is somewhat arbitrary. However, we should point out that such arbitrariness in terms of timing and strength of response also exists in the behavioral studies themselves (McEchron and Disterhoft 1997; Kishimoto et al. 2001).

The metaphor of CS–US sequence learning as an abrupt phase transition seems viable for the early learners. The late learners are more like a phase transition that has difficulty aligning compatible domains as in condensation of certain alloys. However, more interesting for us than the phase transition metaphor, is the testable prediction produced by the model: There exists a trace interval, a little beyond 500 ms, that will produce a clear bimodal result concerning the subject population and trials to learn a conditioned response. The prediction seems quite strong in the sense of Fig. 9 where the early and late learners are each tightly clustered in trials to learn with a large gap between the clusters totally devoid of any simulations.

If the bimodal behavioral result is not produced, then there will be little motivation to find the neurophysiological predictions such as the larger, across-trial activity oscillations associated with the late learners and the initial, transient and small learned US firing. Furthermore if none of the predictions obtain, then we must re-think our minimal model and consider enhancing the model with additional features. The choices of features are many, including more inhibitory neurons, more realistic primary cell firing patterns, a spectrum of axonal delays, appropriate neuronal time-constants, CA3 a,b,c topology (Lorente de Nó 1934), and many others.

The across trial activity oscillations themselves are apparently due to the multiple timescales present in the model including synaptic modification rates (both excitatory and inhibitory) and the rate at which K_0 is modified. We suspect that multiple time scales exist in the hippocampus itself and would welcome some data that would lead to a rational introduction of rate constants beyond the number of trials to learn, which is used to adjust synaptic modification rates.

Finally, the model has given us an interesting example of ‘error-correction’ without ‘error-detection’. If a neurophysiologist had made the cell firing and behavioral observations predicted by the simulations producing the late learners, it would be easy to misinterpret the observations and to postulate an error correction process governing synaptic modification. However, the synaptic modification rule is not of the error-corrector type, and there is no error signal (Levy and Steward 1979). Thus, this result serves as a warning against such interpretations.

In sum, our minimal hippocampal model has predicted specific behavioral and cell firing dynamics that are, for us, well beyond what intuition can predict. These results are reasonably suited for immediate experimental tests. If only a portion of this model’s predictions are correct, we have support for a computational modeling approach that spans biophysics to behavior.

References

- August DA, Levy WB (1996) Temporal sequence compression by a hippocampal network model. In: INNS world congress on neural networks pp 1299–1304
- August DA, Levy WB (1999) Temporal sequence compression by an integrate-and-fire model of hippocampal area CA3. *J Comput Neurosci* 6(1):71–90
- Howe AG, Levy WB (2007) A hippocampal model predicts a fluctuating phase transition when learning certain trace conditioning paradigms. *Cogn Neurodyn* 1(2):143–155
- Kishimoto Y, Kawahara S, Mori H, Mishina M, Kirino Y (2001) Long-trace interval eyeblink conditioning is impaired in mutant mice lacking the NMDA receptor subunit $\epsilon 1$. *Eur J Neurosci* 13(6):1221–1227
- Lester RA, Clements JD, Westbrook GL, Jahr CE (1990) Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature* 346:565
- Levy WB (1989) A computational approach to hippocampal function. *Comput Model Learn Simple Neural Syst* 23:243–305
- 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, Sederberg PB (1997) A neural network model of hippocampally mediated trace conditioning. *Int Conf Neural Netw IEEE* 1:372–376
- Levy WB, Steward O (1979) Synapses as associative memory elements in the hippocampal formation. *Brain Res* 175(2):233–245
- Levy WB, Colbert CM, Desmond NL (1990) Elemental adaptive processes of neurons and synapses: a statistical/computational perspective. In: Gluck MA, Rumelhart DE (eds) *Neuroscience and connectionist theory*. Lawrence Erlbaum, Hillsdale, New Jersey, pp 187–235
- Levy WB, Sanyal A, Wu X, Rodriguez P, Sullivan DW (2005) The formation of neural codes in the hippocampus: trace conditioning as a prototypical paradigm for studying the random recoding hypothesis. *Biol Cybern* 92(6):409–426
- Lorente de Nó R (1934) Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J Psychol Neurol* 46:113–177
- 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(2):1030–1044
- Moyer JR, Deyo RA, Disterhoft JF (1990) Hippocampectomy disrupts trace eye-blink conditioning in rabbits. *Behav Neurosci* 104(2):243–252
- Rodriguez P, Levy WB (2001) A model of hippocampal activity in trace conditioning: where’s the trace? *Behav Neurosci* 115(6):1224–1238
- Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry* 20(1):11–21
- Smith MC, Coleman SR, Gormezano I (1969) Classical conditioning of the rabbit’s nictitating membrane response at backward, simultaneous, and forward CS–US intervals. *J Compar Physiol Psychol* 69(2):226–231
- Smith AC, Wu X, Levy WB (2000) Controlling activity fluctuations in large, sparsely connected random networks. *Netw Comput Neural Syst* 11(1):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(5):729–744

- Sullivan DW, Levy WB (2003) Synaptic modification of interneuron afferents in a hippocampal CA3 model prevents activity oscillations. *Proc Int Joint Conf Neural Netw IEEE* 3:1625–1630
- Sullivan DW, Levy WB (2004) Quantal synaptic failures enhance performance in a minimal hippocampal model. *Netw Comput Neural Syst* 15(1):45–67
- Traynelis SF et al (2010) Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol Rev* 62(3):405–496
- Wu X, Levy WB (2005) Increasing CS and US longevity increases the learnable trace interval. *Neurocomputing* 65:283–289