Available online at www.sciencedirect.com
SCIENCE DIRECT
Neural Networks 18 (2005) 1242–1264
Neural Networks
www.elsevier.com/locate/neunet

2005 Special issue

Interpreting hippocampal function as recoding and forecasting

William B Levy[a,*], Ashlie B. Hocking[a,b], Xiangbao Wu[a]

[a]Department of Neurosurgery, University of Virginia Health System, P.O. Box 800420, Neurosurgery, Charlottesville, VA 22908-0420, USA
[b]Department of Computer Science, University of Virginia, Charlottesville, VA, USA

Abstract

A model of hippocampal function, centered on region CA3, reproduces many of the cognitive and behavioral functions ascribed to the hippocampus. Where there is precise stimulus control and detailed quantitative data, this model reproduces the quantitative behavioral results. Underlying the model is a recoding conjecture of hippocampal computational function. The expanded conjecture includes a special role for randomization and, as recoding progresses with experience, the occurrence of sequence learning and sequence compression. These functions support the putative higher-order hippocampal function, i.e. production of representations readable by a linear decoder and suitable for both neocortical storage and forecasting. Simulations confirm the critical importance of randomly driven recoding and the neurocognitive relevance of sequence learning and compression. Two forms of sequence compression exist, on-line and off-line compression: both are conjectured to support neocortical encoding of context and declarative memory as described by Cohen and Eichenbaum (1993).

Keywords: Prediction; Context; Recoding; CA3; Synaptic modification; Randomization; Chaotic; Cognition; Chunking; Transverse patterning; Transitive inference; Trace conditioning; Cognitive mapping

1. The theory

The fundamental observations of Milner (1972) concerning hippocampal function and its subsequent refinement by Cohen and Squire (1980) and by Cohen and Eichenbaum (1993) direct our ideas about hippocampal function concerning learning and declarative memory. Based on their ideas and observations as well as the anatomical relationship of the hippocampus with neocortex and the intrinsic anatomy of the hippocampus itself, we have concentrated on a recoding theory of hippocampal function (Levy, 1985, 1989, 1990a, 1994). At the same time, the animal literature, with its emphasis on spatial, contextual, and configural learning, led us (Levy, 1989) to include a sequence prediction aspect to this theory as we incorporated the insights of Hirsh (1974), Kesner and Hardy (1983), O'Keefe and Nadel (1978), and eventually Rudy and Sutherland (1995).

Thus, our theory (Levy, 1989) arises from the confluence of several ideas: the basic function of the hippocampus as a cognitive map; the particular anatomy and detailed connectivity of the hippocampus (sparse recurrence in CA3 with divergence of entorhinal cortex (EC) inputs and recurrent signals vs. convergence coming out of CA3 via the sequential projections to CA1, subiculum, and EC); the general need for a device to find associations that the neocortex would have trouble creating due to a lack of connectivity (Levy, 1994); and last but not least, the need to encode correlations across behaviorally relevant time-spans for the purpose of forecasting. Thus, the combination of these perspectives leads to a hippocampal theory that conceptualizes a sequence learning device, as well as conjecturing a random recoder.

Others who model hippocampal function as sequence learning include Abbott and Blum (1996), Hasselmo's laboratory, e.g. Hasselmo et al. (2002), Mehta et al. (1997), Molyneaux and Hasselmo (2002), Schmajuk (2002), Treves (2004), and Tsodyks et al. (1996). On the other hand, McClelland et al. (1995), and Rolls et al. (1997) advocate more conventional pattern recognition models. In terms of pattern recall, speed of convergence is quite rapid for the integrate-and-fire model (Panzeri et al. 2001; Rolls and Treves, 1998; Treves, 1993; Treves et al., 1997). Moreover, recently the Rolls' laboratory has begun to investigate sequence learning in their models (Stringer et al., 2004). From our viewpoint, almost all neurons are

* Corresponding author. Tel.: +1 434 924 5014; fax: +1 434 982 3829.
E-mail address: wbl@virginia.edu (W.B Levy).

0893-6080/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2005.08.005

pattern recognition devices. Therefore, there is no argument about the existence of this ability in the hippocampus. Rather, sequence-learning models have pattern recognition as one of several capabilities (see Levy, 1996 for a summary).

To establish and refine the viability of our theoretical perspective, we have produced a series of computer simulations. The first successful simulations of sequence learning were reported in Minai and Levy (1993b) and Minai et al. (1994). More recently we have been able to move beyond such qualitative observations. Our hippocampal model is now able to reproduce the quantitative data of hippocampal-dependent phenomena where the relevant stimuli are under precise control (see Appendix A). Here, we emphasize the stochastic dependence of the recoding dynamics.

The fundamental recoding by the hippocampal formation occurs in the CA3 subregion of the hippocampus (Levy, 1989). Therefore, we study hippocampal function with models that emphasize this subregion.

2. A family of models

Instead of a single model, we use a family of CA3 models. All the members of this family share certain basic biological properties that are summarized in Table 1. The primary neurons are spike passing, i.e. communication is binary {0,1}. Where we use McCulloch-Pitts neurons, the updating of internal excitation and the axonal communication lag are the same. This is not true for models using integrate-and-fire neurons. The input layer (Fig. 1a) is a merging of the entorhinal cortex (EC) and dentate gyrus (DG) inputs to CA3. Fundamentally, the CA3 model is a sparsely interconnected feedback network (Fig. 1b), typically with thousands of neurons in a simulation. All direct, recurrent connections between primary cells are excitatory. There is an interneuron mediating feedforward inhibition, and one mediating feedback inhibition. Inhibition is of the divisive form, but activity in the free-running models is only imperfectly controlled because of a delay in the feedback that activates these inhibitory neurons.

To date, region CA3 is modeled as a randomly connected network. Each excitatory neuron randomly connects to approximately $n \cdot c$ other neurons, where n is the number of neurons and c is the connectivity ratio (usually set to 0.1 but lower connectivities also work, e.g. Levy et al., 2005; Sullivan and Levy, 2004). Given the output of neuron i at time t, here $z_i(t)$, the net internal excitation of neuron j, $y_j(t)$, is

$$y_j(t) = \frac{\sum_{i=1}^{n} w_{ij} c_{ij} \phi(z_i(t-1))}{\sum_{i=1}^{n} w_{ij} c_{ij} \phi(z_i(t-1)) + K_{\mathrm{FB}} \sum_{i=1}^{n} w_{iI} z_i(t-1) + K_0 + K_{\mathrm{FF}} \sum_{i=1}^{n} x_i(t)} \tag{1}$$

where w_{ij} represents the weight value between neurons i and j at time $t-1$, and c_{ij} is a binary variable {0,1}, indicating whether or not there is a connection from neuron i to j. The term $\sum w_{ij} c_{ij} \phi(z_i(t-1))$ represents the excitatory synaptic conductance for the jth neuron. Parameters K_{FB} and K_{FF} are constants that scale the feedback and feedforward inhibitions, respectively. The constant K_0 controls the magnitude and stability of activity oscillations and is analogous to a shunting rest conductance (Smith et al., 2000). Weights w_{iI} are the positively valued synaptic strengths between each pyramidal cell i and the feedback inhibitory neuron at time $t-1$. The binary external input to neuron j at time t is indicated by $x_j(t)$. If either $x_j(t)=1$ or $y_j(t) \geq \theta$, neuron j fires (i.e. $z_j(t)=1$), where θ is a threshold fixed at 0.5.

Synaptic failures can be included via a synaptic failure channel represented by the function $\phi(z_j(t))$ for the connection from neuron i to neuron j (Sullivan & Levy, 2003a, 2004). Here, $\phi(z_j=0)=0$. A synaptic failure, $\phi(z_j=1)=0$, occurs with probability f, and successful synaptic activation, $\phi(z_j=1)=1$, with probability $(1-f)$; i.e. the failure process is a Bernoulli random variable that acts independently on each synapse at each time-step. The addition of failures allows successful simulations to run at lower activity levels (Sullivan and Levy, 2004).

The model uses a biologically-inspired postsynaptic associative modification rule with potentiation and

Table 1
A minimal hippocampal CA3 model

1	Neurons are threshold elements with inputs that are weighted and summed; the output is binary, a spike when threshold is exceeded and no spike otherwise
2	Most connections are excitatory
3	Synapses modify associatively based on a local Hebbian rule that is time-spanning between pre-and postsynaptic activations and includes LTP and LTD-like processes
4	Recurrent excitation is sparse and randomly connected
5	Recurrent excitation is stronger than external excitation
6	One or more randomization processes exist
7	Inhibitory neurons control activity, approximately
8	Activity is low but not too low