

A familiarity-based learning procedure for the establishment of place fields in area CA3 of the rat hippocampus*

Szabolcs Káli^{1,2} and Peter Dayan¹

¹Gatsby Computational Neuroscience Unit, UCL,
17 Queen Square, London WC1N 3AR, U.K.

²Department of Brain and Cognitive Sciences, Massachusetts
Institute of Technology, Cambridge, MA 02139, U.S.A.

szabolcs@gatsby.ucl.ac.uk

Abstract

That familiarity should gate learning is a neurobiological and computational commonplace, and has been particularly investigated in the context of the hippocampus. We show its critical importance in continuous attractor models, using the formation of place fields in hippocampal area CA3 as an example. Without it, biased sampling of places in an environment, coming from purely random exploration, leads to degenerate attractors and prevents the formation of place fields. Conversely, appropriate use of familiarity information during learning can counteract the sampling bias, resulting in a normal place cell representation.

Introduction

Recurrent neural models with 1, 2, or multi-dimensional continuous attractors have recently been used to explain a host of phenomena such as the head direction system,¹² the neural integrator for eye position,¹⁰ orientation tuned simple cells in primary visual cortex,^{1,3,11} and place cells in area CA3 of the hippocampus.^{5,6,9} As has occasionally been noted,^{8,12} the trickiest problem in realising such networks is preventing attractor drift, *ie*, making states along the continuous attractor equally stable. This problem is rarely to the fore, since the recurrent weights, which define the attractor, are typically set by fiat, and so can be made arbitrarily regular.

In previous work modeling the formation of hippocampal place fields, we have shown that standard Hebbian plasticity applied to location and head-direction sensitive inputs from the entorhinal cortex (EC) can be used to learn feedforward weights in the perforant path from EC to area CA3 and recurrent weights within CA3 that generate place cells as states in a multidimensional continuous attractor.⁶ This model accounts for many experimentally observed properties

*We are grateful to an anonymous reviewer of our previous model for encouraging us to study the effects of sampling. Support was from the Gatsby Charitable Foundation.

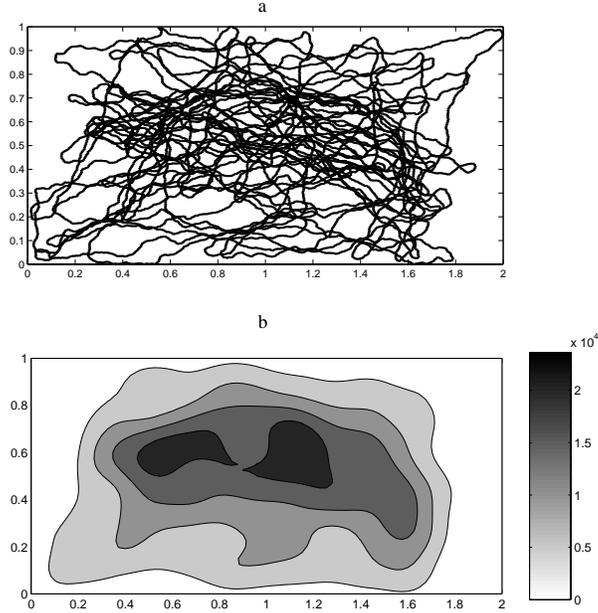


Figure 1: **a**: An example trajectory; **b**: Exposure to locations in the apparatus ($g(\mathbf{x}^*, t)$) after 5 minutes of exploration ($\sigma = 0.075$).

of place fields, including the complex effects of environmental manipulations.⁷ Plasticity is critical in the model, as there is otherwise no obvious way of establishing appropriate attractors.

However, in the first implementation of this model, we assumed that all locations (and head directions) in an environment are visited equally often, an assumption that can clearly be violated in realistic training procedures. In this paper, we show that inevitable uneven sampling of the environment leads to a catastrophic attractor degeneracy. The previous model used a binary familiarity measure to switch between a plastic mode (for novel environments) and a standard mode (for familiar environments). We also explore whether we can use a more graded measure of familiarity within an environment to ameliorate this degeneracy.

Results

We simulated sample paths in an experimental paradigm in which the rat, for 5 minutes, chases food pellets thrown into random locations in a rectangular apparatus.² An example trajectory is shown in Figure 1a. We quantified variations in exposure to different parts of the apparatus by convolving the sample path with a Gaussian, yielding

$$g(\mathbf{x}^*, t) = \int_0^t e^{-\frac{(\mathbf{x}(t') - \mathbf{x}^*)^2}{2\sigma^2}} dt'$$

where t is time since the beginning of exploration, $\mathbf{x}(t)$ is the rat's position at time t , and σ is the width of the Gaussian. This measures sampling density as

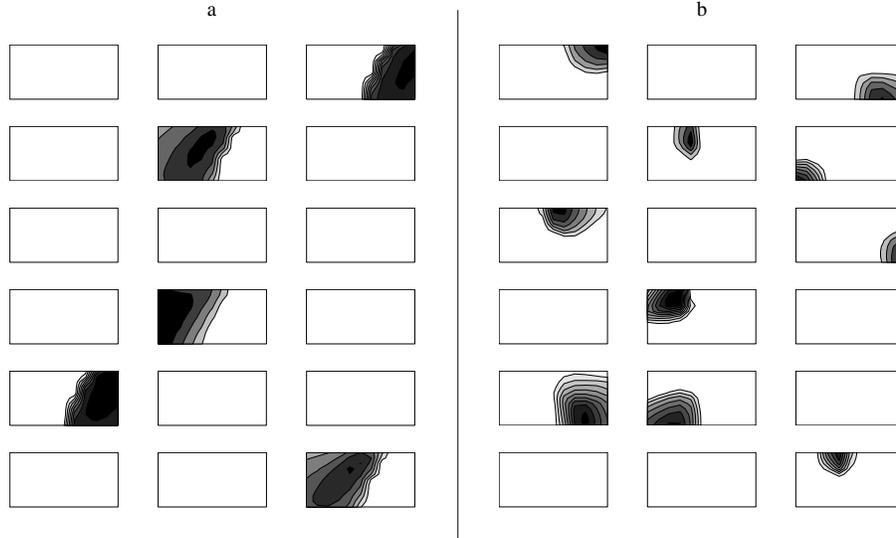


Figure 2: The spatial activity patterns of 12 randomly selected CA3 place cells after the exploration shown in Figure 1a, **a**: using simple Hebbian learning, and **b**: using the familiarity-based learning procedure. The place fields in **(b)** closely resemble experimental place fields, and provide good coverage of the whole environment. Conversely, the spatial firing patterns in **(a)** reflect essentially two different attractor states containing only a small proportion of the neurons, perturbed to some extent by the feedforward input.

a function of position, and is displayed in Figure 1b. It deviates both randomly and systematically from the uniform sampling density which, in conjunction with a Hebbian learning rule, is known to result in a good place cell representation. Further simulations indicated that the local random fluctuations in sampling density are essentially harmless since they do not destroy the overall attractor structure. However, the fact that, on average, the animal spends several times as much time at a location near the center of the apparatus than at a location near the edges, causes the naive Hebbian learning procedure to break down completely, resulting in a very poor place cell representation. An example of this is given in Figure 2a – the network possesses just two distinct attractors, and only neurons which are active in one of these attractors ever become active in this environment. Furthermore, the effect is persistent with respect to local manipulations of the learning rule involving only pre- and post-synaptic activities and the current efficacy of the weight.

An obvious solution to this problem is to vary the degree of plasticity in proportion to the novelty of a location in an on-line manner. We define the novelty of the current location $\mathbf{x}(t)$ at time t as

$$n(t) = e^{-\alpha g(\mathbf{x}(t), t)}$$

where α is the rate at which novelty decreases with increasing exposure, and introduce the learning rule

$$\dot{W}_{ij} = n(t) u_i(t) r_j(t)$$

where u_i is the depolarization of the i th postsynaptic unit, r_j the activation of the j th presynaptic unit, and W_{ij} is the weight connecting the two. The parameters σ and α were optimized numerically so that the amount of learning that occurs in different parts of the apparatus is as uniform as possible after 5, or more, minutes of exploration. The weight structure becomes increasingly uniform, given more exploration, and the place fields duly become increasingly regular. The application of this learning procedure results in an attractor structure not very much different from the one we got in the case of even sampling, and, as shown in Figure 2b, leads to a good place cell representation after just 5 minutes of exploration, in agreement with experimental data.

Discussion

In our previous model, and indeed most previous models of hippocampal plasticity, novelty has been seen as gating learning (possibly via GABA-ergic and cholinergic modulation from the septum) in a rather binary fashion, switching between different functional modes. Here we have shown how we can make use of a more graded representation of familiarity. Further, it has been suggested⁴ that such a familiarity signal might partly be computed by a feedback loop in which septal activity may in turn be modulated by activity in the CA1 region of the hippocampus. Similar processes could be used in our model to maintain and use a graded neural representation of familiarity. In fact, while the above simulations use a table lookup representation of familiarity, a variant of the model in which familiarity is maintained in the form of a more neurally plausible radial basis function representation works almost equally well.

Future work on the model includes more theoretical investigations of attractor degeneracy and the true effects of the random noise in the sampling density, and also a study on how the novelty signal $n(t)$ that gates learning should depend on local environmental exposure $g(\mathbf{x}^*, t)$.

References

- [1] Ben-Yishai, R., Bar-Or, R.L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 3844-3848.
- [2] Brunel, N. and Trullier, O. (1998). Plasticity of directional place fields in a model of rodent CA3. *Hippocampus*, 8, 651-65.
- [3] Carandini, M. and Ringach, D.L. (1997). Predictions of a recurrent model of orientation selectivity. *Vision Res.*, 37, 3061-71.
- [4] Hasselmo, M.E., Wyble, B.P., and Wallenstein, G.V. (1996). Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus*, 6, 693-708.
- [5] Káli, S. and Dayan, P. (1998). The formation of direction independent place fields in area CA3 of the rodent hippocampus using Hebbian plasticity in a recurrent network. *Soc. Neurosci. Abstr.*, 24, 931.

- [6] Káli, S. and Dayan, P. (1999). Spatial representations in related environments in a recurrent model of area CA3 of the rat. In Proceedings of ICANN99.
- [7] O'Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381, 425-428.
- [8] Pouget, A., Zhang, K., Deneve, S., and Latham P.E. (1998). Statistically efficient estimation using population coding. *Neural comp.*, 10, 373-401.
- [9] Samsonovich, A. and McNaughton, B.L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.*, 17, 5900-5920.
- [10] Seung, H.S. (1996). How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U.S.A.*, 93, 13339-44.
- [11] Somers, D.C., Nelson, S.B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, 15, 5448-65.
- [12] Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.*, 16, 2112-2126.