

Hippocampal Remapping as Learned Clustering of Experience

Honi Sanders (honi@mit.edu)

Center for Brains Minds and Machines, Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, 43 Vassar Street, Room 46-5233
Cambridge, MA 02139 USA

Matthew Wilson (mwilson@mit.edu)

Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, 43 Vassar Street, Room 46-5233
Cambridge, MA 02139 USA

Samuel J. Gershman (gershman@fas.harvard.edu)

Department of Psychology and Center for Brain Science
Harvard University, 52 Oxford Street, Room 295.05
Cambridge, MA 02138 USA

Abstract

The place cells of the hippocampus collectively form distinct maps of each context, a process known as hippocampal remapping. Past work has asked which features of an experience determine which map is used, but no consistent answer has been reached. However, this approach has ignored the place of context identification as part of a learning process. We suggest that context identification corresponds to an unsupervised clustering problem, where the animal receives a stream of observations and must cluster them in a data-driven manner. Each cluster corresponds to a particular context, and therefore a particular hippocampal map. Formalizing context learning as a clustering problem allows us to capture a range of experimental results that have not yet been explained by a single theoretical framework. In particular, our results highlight the role that learning plays in hippocampal remapping. This model also provides novel predictions, such as the effect of variability in training.

Keywords: hippocampus; Bayesian nonparametric clustering; learning; place cell remapping

Background

Hippocampal remapping is thought to be a neural correlate of context representation (Colgin, Moser, & Moser, 2008). Context-dependent behavior is ethologically important, and is at least partially hippocampal dependent (Holland & Bouton, 1999; Gershman, Blei, & Niv, 2010). Given this window into high-level representation of context, the field has sought to answer the question, “What aspects of an experience determine context boundaries?”

Here, we take a “normative” approach to capturing the phenomenon of hippocampal remapping: we ask what computational problem the hippocampus is solving and then see what insight the constraints on the solution to that problem provide. The question of how to assign observations to contexts, and therefore how to appropriately generalize context-specific knowledge, is a difficult question. One challenge is that there

are no objective context labels that the animal has access to. Rather, the animal receives a stream of unlabeled observations, which it must partition appropriately into a series of contexts. Another challenge is that there is no objective indication as to how many contexts the animal will be exposed to. Finally, there is no *a priori* information about what the nature (statistics) of a given context is: which features are variable and which are consistent, what range of feature values are typical, or in formal terms what distribution observations are drawn from in a given context.

Fundamentally, this corresponds to an unsupervised clustering problem, where observations are data points for which one must find a partition of those data points into clusters (i.e., contexts) that minimizes the number of clusters needed while maximizing variance captured. Each cluster can be seen as a “latent state”, which corresponds to a unique generative distribution that generates observations when the world is in that state.

The clustering algorithm that we will be focusing on is Bayesian nonparametric clustering (BNC) (Gershman & Blei, 2012). Other clustering methods will have qualitatively similar behaviors. The main features of BNC are that it allows for the number of clusters to be determined flexibly by the data, that clusters can have independent shapes, and that the algorithm provides probabilistic cluster assignments rather than hard assignments. These capabilities address the issues described above: that labels must be inferred and are not given, that the animal does not know *a priori* how many latent states exist, and that the animal does not know *a priori* the statistics of the generative distribution in each state.

Model

We model the animal as performing inference over a hidden state c given an observation y . The observation may be multi-dimensional, with each dimension corresponding to a feature of the observation, whether sensory or non-sensory, such as the color of the walls or the task that the animal is performing. The goal is to estimate the probability that each



observation y_t was generated by a particular latent state c_t , or equivalently what probability to assign a given list of latent state assignments \mathbf{c} for a given list of observations \mathbf{y} : $P(\mathbf{c}|\mathbf{y}) \propto P(\mathbf{c})P(\mathbf{y}|\mathbf{c})$.

There are two factors that go into that probability: one is how likely the proposed partition of observations into latent states is under the Chinese restaurant process (CRP) prior (Gershman & Blei, 2012) ($\mathbf{c} \sim \text{CRP}(\alpha)$, where α is the concentration parameter that controls the expected number of states; we use $\alpha = 10^{-3}$). The process defines a probability of a given assignment being assigned to a novel draw c_t given an assignment of labels to previous draws $\mathbf{c}_{1:t-1}$, which has K unique labels, the k th of which has m_k observations associated with it:

$$P(c_t = k | \mathbf{c}_{1:t-1}) = \begin{cases} \frac{m_k}{t-1+\alpha} & \text{if } k \leq K \\ \frac{\alpha}{t-1+\alpha} & \text{if } k = K+1 \end{cases} \quad (1)$$

Fewer latent states are more likely, and the extent to which fewer latent states are favored is a function of α .

The other factor is how consistent different observations from the same latent state are with each other. More formally, it is the likelihood of observing a set of observations from a given cluster marginalizing over all possible parameter settings of the generative distribution:

$$P(\mathbf{y}|\mathbf{c}) = \int_{\theta} \left[\prod_t F(\mathbf{y}|\theta_{c_t}) \prod_k G_0(\theta_k) \right] d\theta \quad (2)$$

We take the generative distribution F to be a Gaussian whose parameters are drawn from G_0 , a Normal-Gamma distribution.

This prior and likelihood correspond to the generative model in Fig. 1A.

One thing to re-emphasize is that the outputs of BNC are probabilistic. This is important from a normative perspective because much of the time it is impossible to ascertain with certainty whether observations are drawn from the same distribution or not. Future information may change that assessment. The representation of uncertainty about latent state assignment corresponds in an important way with the result that hippocampal maps during two experiences are almost never entirely overlapping nor entirely independent. This ‘‘partial remapping’’ corresponds with the need to represent the inherent uncertainty about whether different observations are actually drawn from identical distributions. We can think of the subset of place fields that remap between two observations as being related to the hypothesis that the observations are drawn from independent distributions, and the place fields that do not remap as related to the hypothesis that the observations are drawn from the same distribution (Fuhs & Touretzky, 2007).

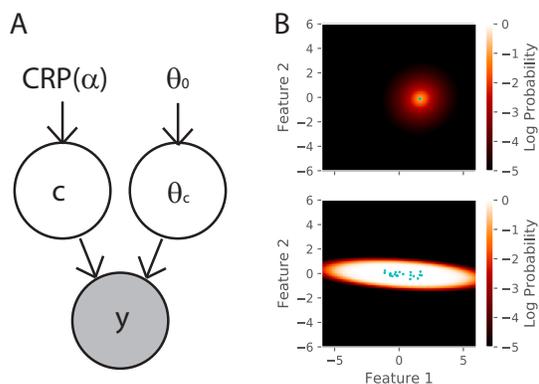


Figure 1: A) Generative model. Observations (y) are drawn from a distribution parametrized by θ_c , the parameters associated with latent state c . The latent state index c is drawn from a Chinese restaurant process parametrized by α . The θ parameters for each latent state are drawn from a prior distribution parametrized by θ_0 . B) Demonstration of inference. Observations are cyan dots, background color is log probability of assigning a new observation to same latent state as cyan dots. Top panel is after 1 observation, bottom panel is after 20 observations.

Results

Instructive Example of Model Function

We begin with a simple example of the latent state inference of this model. We generate observations from a 2D Gaussian with $\mu = [0, 0]$ and $\sigma = [1, 0.2]$. A single observation is observed with feature values $[1.6, -0.1]$ (cyan dot in Fig. 1B top). We can measure the posterior probability that any other observation would be drawn from the same latent state as the first observation $p(c_2 = c_1 | y_2, y_1)$ for every possible observation y_2 . This probability is shown using the heat of colors in the background of Fig. 1B top. New observations close to the original observation are more likely to be clustered together. The peak value is controlled by α and the spread is controlled by θ_0 .

The probability of clustering novel observations with past observations changes over the course of learning. We can apply the same inference process after the model has seen 20 draws from the same distribution (cyan dots in Fig. 1B bottom). The model now has sufficient data so that the posterior probability of a novel observation being clustered with the past observations captures the shape of the Gaussian distribution from which the observations were drawn: it has high spread in feature 1 and low spread in feature 2, corresponding to the statistics of the distribution.

This example captures the trade-off of learning: it can capture the shape of the data, but requires experience in order to do so.

Learned Distinctions

One initial focus of the study of remapping was to ask which sensory aspects had “control” over the hippocampal map, but the answer was invariably that it depended on prior experience (O’Keefe & Speakman, 1987; Knierim, Kudrimoti, & McNaughton, 1998; Bostock, Muller, & Kubie, 1991). One prime example of this is the question of the role of “environmental geometry” or the shape of the recording arena. The first group to record throughout the course of learning found that there was not a consistent relationship between the shape of the recording arena and inferred latent state (Lever, Wills, Cacucci, Burgess, & O’Keefe, 2002). The experiment they performed was to record place cells of rats who were alternately placed in square and circle boxes in the same location in the recording room day after day. What they found was that early in learning, there was limited remapping, indicating that the animal considered the observations to be drawn from the same latent state. Only after extensive experience did the animals remap between the two enclosures (Fig. 2A). This indicates that latent state boundaries change with experience. These effects are hard to explain under a framework where latent state has *a priori* boundaries. However, it is consistent with the clustering perspective. As emphasized above, clustering allows for capturing arbitrary distinctions, but it requires sufficient data for the evidence in favor of creating a distinction to outweigh the bias against unnecessarily creating extra latent states.

We model these experiments in the following way. We take observations to be 1D for simplicity, where the single dimension is the feature along which the distinction is learned. For example, in the circle-square experiment (Lever et al., 2002), the dimension would be the shape of the enclosure. We generate observations from two Gaussians with $\mu_1 = -0.5, \mu_2 = 0.5, \sigma_1 = \sigma_2 = 0.15$ (Fig. 2B). These two Gaussians correspond to the two experimenter-defined conditions. We alternate drawing observations from each distribution. After each pair of draws, we ask the model what the relative probability of the hypothesis that all observations up to that point were drawn from a single latent state against the hypothesis that all observations up to that point had been drawn from two alternating latent states (Fig. 2C). Early in training, the single latent state hypothesis is more probable. Having two latent states is less probable under the CRP prior, and the observations are not yet sufficiently inconsistent with the single latent state hypothesis. In other words, the likelihood of the observations marginalized over all possible parameters of a single Gaussian is not yet sufficiently small to outweigh the CRP prior bias against adding a second latent state. Only after sufficient data has been observed are the observations sufficiently inconsistent with the single Gaussian hypothesis for its probability to decrease below the two-latent state hypothesis. In this way, the clustering perspective on remapping explains why it takes time to begin to remap between two distinct enclosures even if the distinction is apparent to the animal from the beginning: only through experience can the animal tell if observed varia-

tion is consistent enough to merit creating new latent states.

Other Phenomena and Predictions

Our model also captures a wide range of other phenomena observed in the hippocampal remapping literature. The observation that hippocampal maps take time to stabilize (Frank, Stanley, & Brown, 2004; S. Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Law, Bulkin, & Smith, 2016) can be taken as reflecting the fact that experience is required for accurate estimates of the parameters of the generative distribution for each latent state. The observation of partial remapping in response to changes in movement direction on a linear track but not in an open field (Markus et al., 1995) can be taken as reflecting the fact that inference of distinct latent states would only occur when the observations are clustered. The observation that different animals express different remapping behaviors in response to identical training (Bostock et al., 1991; Lever et al., 2002; Wills et al., 2005) can be understood if each animal has a unique setting of the α parameter controlling willingness to infer novel latent states.

The insight provided by this model leads us to several experimental predictions. We suggest that the range of behaviors observed in the different morph experiments (J. K. Leutgeb et al., 2005; Wills et al., 2005; Colgin et al., 2010) can be recreated in a single experiment by manipulating the amount of pretraining before the morph testing. We also make predictions about the role of variance in training experiences in controlling later remapping behavior.

Broader Significance

Hippocampal remapping provides a window into the animal’s subjective sense of context. Our work provides a motivated framework for making sense of a range of empirical observations associated with remapping. It emphasizes the role of learning and experience in context segmentation, and highlights the trade-off between flexibility and speed of learning.

Acknowledgments

This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

References

- Bostock, E., Muller, R. U., & Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, *1*(2), 193–205. doi: 10.1002/hipo.450010207
- Colgin, L. L., Leutgeb, S., Jezek, K., Leutgeb, J. K., Moser, E. I., McNaughton, B. L., & Moser, M.-B. (2010, 7). Attractor-Map Versus Autoassociation Based Attractor Dynamics in the Hippocampal Network. *Journal of Neurophysiology*, *104*(1), 35–50. doi: 10.1152/jn.00202.2010
- Colgin, L. L., Moser, E. I., & Moser, M.-B. (2008, 9). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, *31*(9), 469–77. doi: 10.1016/j.tins.2008.06.008

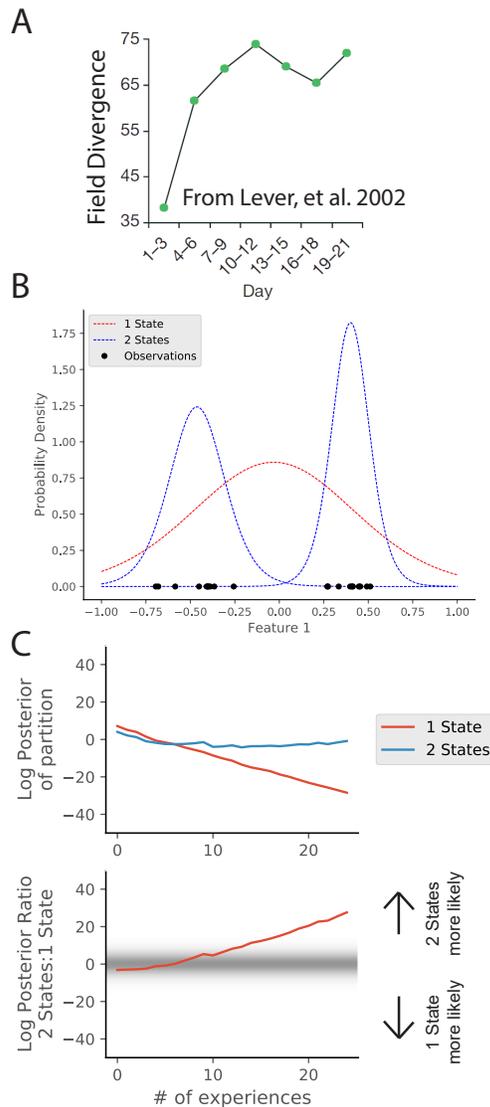


Figure 2: A) From Lever, et al., 2002. They compared place cell representations between alternating presentations of square and circle boxes. Field Divergence is expressed in percent and represents the fraction of place fields that remap between the two enclosures. The representations are initially similar, but diverge with learning. B) Observations (black dots) are generated from Gaussians centered at -0.5, 0.5. The model compares the posterior probability of the observations coming from 1 hypothesized latent state (red) or 2 hypothesized latent states (blue). C) The relative probability assigned to the observations coming from 1 latent state or 2 latent states is shown as a function of amount of experience. Early on, 1 latent state is more probable, whereas later 2 latent states is more probable, similar to the empirical observations.

- Frank, L. M., Stanley, G. B., & Brown, E. N. (2004, 9). Hippocampal Plasticity across Multiple Days of Exposure to Novel Environments. *Journal of Neuroscience*, 24(35), 7681–7689. doi:10.1523/JNEUROSCI.1958-04.2004
- Fuhs, M. C., & Touretzky, D. S. (2007). Context learning in the rodent hippocampus. *Neural computation*, 19(12), 3173–3215. doi:10.1162/neco.2007.19.12.3173
- Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1), 1–12. doi:10.1016/j.jmp.2011.08.004
- Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological review*, 117(1), 197–209. doi:10.1037/a0017808
- Holland, P. C., & Bouton, M. E. (1999). Hippocampus and context in classical conditioning. *Current Opinion in Neurobiology*, 9(2), 195–202. doi:10.1016/S0959-4388(99)80027-0
- Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1998). Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *Journal of neurophysiology*, 80(1), 425–446. doi:10.1037/0735-7044.110.4.643
- Law, L. M., Bulkin, D. A., & Smith, D. M. (2016, 12). Slow Stabilization of Concurrently Acquired Hippocampal Context Representations. *Hippocampus*, 00(12), 1560–1569. doi:10.1002/hipo.22656
- Leutgeb, J. K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C. A., McNaughton, B. L., ... Moser, E. I. (2005, 10). Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron*, 48(2), 345–348. doi:10.1016/j.neuron.2005.09.007
- Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., & Moser, E. I. (2004, 8). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, 305(5688), 1295–8. doi:10.1126/science.1100265
- Lever, C., Wills, T. J., Cacucci, F., Burgess, N., & O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature*, 416(6876), 90–94. doi:10.1038/416090a
- Markus, E. J., Qin, Y. L., Leonard, B., Skaggs, W. E., McNaughton, B. L., & Barnes, C. A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *The Journal of Neuroscience*, 15(11), 7079–7094.
- O'Keefe, J., & Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research*, 68(1), 1–27. doi:3691688
- Wills, T. J., Lever, C., Cacucci, F., Burgess, N., O'Keefe, J., & Keefe, J. O. (2005, 5). Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723), 873–6. doi:10.1126/science.1108905