Psychological Review
2010, Vol. 117, No. 1, 197–209

0033-295X/10/$12.00 DOI: 10.1037/a0017808

Context, Learning, and Extinction

Samuel J. Gershman, David M. Blei, and Yael Niv
Princeton University

A. Redish et al. (2007) proposed a reinforcement learning model of context-dependent learning and extinction in conditioning experiments, using the idea of "state classification" to categorize new observations into states. In the current article, the authors propose an interpretation of this idea in terms of normative statistical inference. They focus on renewal and latent inhibition, 2 conditioning paradigms in which contextual manipulations have been studied extensively, and show that online Bayesian inference within a model that assumes an unbounded number of latent causes can characterize a diverse set of behavioral results from such manipulations, some of which pose problems for the model of Redish et al. Moreover, in both paradigms, context dependence is absent in younger animals, or if hippocampal lesions are made prior to training. The authors suggest an explanation in terms of a restricted capacity to infer new causes.

Keywords: classical conditioning, renewal, latent inhibition, Bayesian, hippocampus

An enduring problem in the study of classical conditioning is how animals learn about the causal structure of their environment (Blaisdell, Sawa, Leising, & Waldmann, 2006). Most theories frame conditioning as the learning of associations between stimuli and reinforcement (Pearce & Bouton, 2001; Rescorla & Wagner, 1972). Under a statistical interpretation, these associations are parameters of a generative model in which stimuli cause reinforcement (Kakade & Dayan, 2002). However, evidence suggests that animals may employ more flexible models, learning, for example, that some stimuli are causally unrelated to reinforcement (Dayan, Kakade, & Montague, 2000; Dayan & Long, 1998). A more radical departure are *latent cause* models (Courville, 2006; Courville, Daw, Gordon, & Touretzky, 2004; Courville, Daw, & Touretzky, 2002), in which both stimuli and reinforcement are attributed to causes that are hidden from observation. One motivation for such models is the finding that learned relationships between cues and reinforcement are not necessarily erased following extinction: Returning the animal to the original training context after extinction in a different context can lead to renewal of the conditioned response (Bouton, 2004; Bouton & Bolles, 1979). These and related data can be characterized by a latent cause model in which different latent causes are associated with the training and extinction contexts.

One problem with latent cause models is that the number of different latent causes is in general unknown. The challenge, then, is to formulate a learning algorithm that can infer new causes as it gathers observations, as well as learn the statistical relationships between causes and observations. Recently, Redish, Jensen, Johnson, and Kurth-Nelson (2007) formulated such a theory of extinction learning, combining the well-studied framework of temporal difference reinforcement learning (Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998) with a state classification mechanism that allows the state space to expand adaptively. In their model, states can be loosely interpreted as latent causes, serving to explain both stimuli and reinforcement in terms of an underlying discrete variable.

We suggest a new model of latent cause inference in animal learning that is based in a normative statistical framework. With this model, we address certain limitations of the theory of Redish et al. (2007), while still capturing its essential insights. We agree with their assertion that the computational problem the animal must solve is one of structure learning. We posit that the computational principles at work in structure learning are based on a generative model of the environment that specifies the animal's a priori beliefs about how its observations were generated by latent causes. Given a set of observations, the problem facing the animal is to combine its prior beliefs with the evidence provided by the observations to infer which causes were in action. At the algorithmic level, we identify several features of Redish et al.'s model that make it difficult to account for relevant data and show how these are obviated in our model. Finally, drawing on a suggestion by Redish et al., we make explicit the computational role played by the hippocampus in our model and use this to explain developmental changes in learning.

Redish et al.'s (2007) Model

The data motivating the model of Redish et al. (2007) came from a conditioning procedure studied by Bouton and Bolles (1979): In the conditioning phase, the animal is placed in Context A and exposed in each trial to both a stimulus cue and a reinforcer; eventually the cue comes to evoke a conditioned response. In the extinction phase, the animal is then placed in a new context (B) and exposed in each trial to the cue in the absence of reinforcement, until the cue ceases to evoke the conditioned response. It

Samuel J. Gershman and Yael Niv, Psychology Department and Neuroscience Institute, Princeton University; David M. Blei, Computer Science Department, Princeton University.

David M. Blei is supported by grants from Google and Microsoft Research. We thank Stephen Maren, Marie Monfils, and David Redish for comments on the article. We thank Nathaniel Daw, Michael Todd, Anatole Gershman, and Ken Norman for helpful discussions.

Correspondence concerning this article should be addressed to Samuel J. Gershman, Psychology Department, Princeton University, Princeton, NJ 08540. E-mail: sjgershm@princeton.edu

would appear, at first glance, that the animal has "unlearned" its response to the cue. However, if the animal is returned to the original context (A) in a test phase and presented with the cue, the response is restored, strongly suggesting otherwise. Rather, it seems that the animal has learned a new relationship between the cue and the reinforcer during extinction that was somewhat limited to Context B.

This phenomenon, known as *ABA renewal*, is explained by Redish et al.'s (2007) model in terms of two simultaneous processes: a value-learning process and a state-classification process. The first updates values associated with states (and potentially actions), using a form of the temporal difference learning algorithm (Sutton & Barto, 1998), which is closely related to the Rescorla-Wagner update rule (Rescorla & Wagner, 1972). A state's value represents a prediction about future reinforcement in that state. The temporal difference learning rule incrementally updates values in proportion to the discrepancy between predicted and received reinforcement (the *prediction error*). In the Pavlovian version of the renewal paradigm described here, the animal's conditioned response is presumed to be proportional to the current state's value (see Dayan, Niv, Seymour, & Daw, 2006). In the operant version modeled by Redish et al., the probability of the animal taking a particular action (e.g., lever press) is proportional to the state-action value.

The innovation of Redish et al. (2007) is to introduce a state-classification process that determines what state the animal is currently in and creates new states when the observation statistics change. The observations, in this case, are defined to be tuples consisting of {context, stimulus, immediate reinforcement, time since last reinforcement}. The actual mechanics of the state classification are quite sophisticated, and we refer the reader to the original article. In brief, a competitive learning model using radial basis functions and classifier expansion (Grossberg, 1976; Hertz, Krogh, & Palmer, 1991) partitions the input space into multivariate Gaussian state prototypes; temporal difference learning then operates on these states. For our purposes, the important aspect of this process is that each state is associated with an observation prototype, and a new observation is classified as a particular state on the basis of its match to the state's prototype. When the observation fails to match any prototype (as determined by a threshold), a new state/prototype is inferred. A local estimate of the average negative prediction error modulates this process: When prediction errors are tonically negative, a new state is more likely to be inferred.

According to this model, acquisition in the ABA renewal paradigm proceeds according to the value-learning process, with all training observations being assigned to the same state (because their statistics are homogeneous). During extinction, the absence of the predicted reinforcement results in tonic negative prediction errors. Combined with a context change, this results in the state-classification process creating a new state. Thus, one state is associated with reinforcement, and another state is associated with no reinforcement. When the animal is returned to the training context, it identifies its observations as belonging to the state associated with reinforcement (on the basis of the contextual cue) and therefore produces the conditioned response.

One implication of this model is that new states are unlikely to be inferred when prediction errors are tonically *positive*. Evidence in contradiction of this hypothesis comes from the context dependence of latent inhibition (Hall & Honey, 1989). The latent inhibition procedure is, in some sense, a concomitant manipulation to extinction: An animal is first exposed to a stimulus in the absence of reinforcement and is later conditioned with pairing of the stimulus and reinforcement. In this case, animals are slower to acquire a conditioned response compared with animals that have not been pre-exposed. However, if the pre-exposure and conditioning phases are conducted in different contexts, the latent inhibition effect is diminished. Here the conditioning phase is accompanied by positive prediction errors (as the reinforcement is unexpected following the pre-exposure), which, according to Redish et al.'s (2007) model, should not result in the inference of a new state. Hence, their model mispredicts that a shift in context will not affect latent inhibition.[1] This problem was also noted by Redish et al.

Another problem is that because the values associated with new states are initialized to zero, Redish et al.'s (2007) model does not predict ABC and AAB renewal (Bouton & Bolles, 1979; Bouton & King, 1983), in which the test trials occur in a completely new context. In both these cases, conditioned responding returns during the test phase. This can be fairly easily accomodated by initializing the values of new states to some prior belief about state values, as we discuss here in connection with our model.

Apart from these problems, the idea of state classification on the basis of observation statistics is a fundamental contribution. In formulating a quantitative theory of how animals solve this problem, we would like to understand the statistical principles underlying this insight. To this end, we propose a new model that is conceptually aligned with that of Redish et al. (2007) but more directly descended from the latent cause theory of Courville (2006).

The rest of this article is organized as follows: We first describe an infinite-capacity mixture model and a particle filter algorithm for performing inference in this model. We then present the results of simulations of latent inhibition and renewal paradigms, including developmental and hippocampal manipulations. In the discussion, we compare our model with that of Redish et al. (2007), as well as several other models. Finally, we discuss some limitations of our model and propose directions for its future development.

A New Model: Statistical Inference in an Infinite-Capacity Mixture Model

The central claim of our account is that, to adaptively predict events in their environment, animals attempt to partition observations into separate groups on the basis of their properties. This task is known as *clustering* in computer science and statistics, and hence, we call these groups *clusters*. We assume that the animal's goal is to assign observations to clusters such that the clusters correspond to different latent causes. Renewal can then be understood as the result of this clustering process. Specifically, we suggest that the animal learns to partition its observations on the basis of their features into two distinct clusters, corresponding to the acquisition and extinction phases (which are implicitly the causes of the animal's observations).

[1] Redish et al.'s (2007) model will demonstrate latent inhibition if the modulation of state classification by tonic prediction error is weak. In this case, state classification is driven primarily by the match between the current observation and the prototypes. However, this scenario is at odds with the central role played by tonic prediction error in Redish et al.'s model.

Our basic approach is to first formulate a set of assumptions that we impute to the animal and then describe how, on the basis of these assumptions, the animal can make rational inferences about latent causes given a set of observations. The set of assumptions constitutes the *generative model*, which represents the animal's prior beliefs about the statistical structure and probabilistic dependencies between variables (both hidden and observed) in the environment. The generative model expresses the state of the animal's beliefs *prior* to making any observations. Given a set of observations, we expect the animal's beliefs (or inference) about the actual causes of these observations to change. In particular, this new state of knowledge is expressed by a posterior distribution over unobserved (hidden) variables given the observed variables. We refer to this as the animal's *inference model*.[2] In the context of the classical conditioning data that we model, the inference model represents the animal's beliefs about the latent causes of its observations.

Generative Model

We assume that the animal's observation on trial t takes the form of a discrete-valued multidimensional vector $\mathbf{f}_t$, with the following dimensions: reinforcement ($f_{t,1}$), cue ($f_{t,2}$) and context ($f_{t,3}$). The reinforcement dimension represents a binary unconditioned stimulus delivered to the animal (e.g., $f_{t,1} \in$ {reinforcement, no reinforcement}). The cue dimension represents a typical Pavlovian cue (or its absence; e.g., $f_{t,2} \in$ {tone, no tone}).[3] The context dimension is an abstraction of the context manipulations typical in renewal paradigms (e.g., box color, shape, odor), which we simplify into discrete values: $f_{t,3} \in$ {Context A, Context B, Context C}.

The generative model we impute to the animal is one in which, on each trial, a single latent cause is responsible for generating all the observation features (reinforcement, cue, context). In such a *mixture model*, each trial is assumed to be generated by first sampling a cause c_t (from a known set of causes) according to a mixing distribution $P(c)$ and then sampling observation features conditioned on the cause from an observation distribution $P(\mathbf{f}|c_t)$. This type of generative model is a reasonable prior belief for many environments. In fact, it correctly expresses, to a first approximation, the process by which many conditioning procedures are generated: First a phase (e.g., conditioning, extinction, test) is selected, and then a set of stimuli are selected conditioned on the phase. If the animal assumes that each observation is probabilistically generated by a single latent cause, then clustering is the process of recovering these causes on the basis of its observations.[4]

The mixture model described so far implicitly assumes that the animal knows how many possible causes there are in the environment. This seems an unreasonable assumption about the structure of the animal's environment, as well as the animal's a priori knowledge about its observations. Furthermore, as we discuss later, there is evidence that animals can flexibly infer the existence of new causes as more observations are made. We thus use a generative model that allows for an unbounded (expanding) number of latent causes (an *infinite capacity* mixture model, as described below). In this model, the animal prefers a small number of causes but can, at any time, infer the occurrence of a new latent cause when the data support its existence and thus decide to assign its current observation to a completely new cluster.

Formally, let us denote a partition of observations (trials) $1, \ldots, t$ by the vector $\mathbf{c}_{1:t}$. A partition specifies which observations were generated by which causes, such that $c_t = k$ indicates that the observation t was assigned to cluster k. In our model, the animal's prior over partitions takes the form of a sequential stochastic generative process (Aldous, 1985; Pitman, 2002) that generates cause k with probability

$$P(\mathrm{c}_{t+1} = k|\mathrm{c}_{1:t}) = \begin{cases} \dfrac{N_k}{t+\alpha} & \text{if } k \leq K_t \text{ (i.e., } k \text{ is an old cause)} \\ \dfrac{\alpha}{t+\alpha} & \text{if } k = K_t + 1 \text{ (i.e., } k \text{ is a new cause),} \end{cases} \quad (1)$$

where N_k is the number of observations already generated by cause k (by default it is assumed that $c_1 = 1$) and K_t is the number of unique causes generated for observations 1 to t. The number of causes generating observations $1, \ldots, t$ is now a random variable and can be any number from 1 to t. The concentration parameter α specifies the animal's prior belief about the number of causes in the environment. When $\alpha = 0$, the animal assumes that all observations are generated by a single cause; when α approaches ∞, the animal assumes that each observation is generated by a unique cause. In general, for $\alpha < \infty$, the animal assumes that observations tend to be generated by a small number of causes.[5]

The animal further assumes that once a cause has been sampled for a trial, an observation is sampled from an observation distribution conditional on the cause. Each cause is associated with a multinomial observation distribution over features, parameterized by ϕ, where $\phi_{i,j,k}$ is the probability of observing value j (e.g.,

[2] This is also sometimes referred to as the *recognition model* (Dayan & Abbott, 2001).

[3] The choice of discrete-valued observations is not crucial to our formalism; we have used real-valued features and obtained essentially the same results.

[4] We use the term *cause* in connection with the generative model and the term *cluster* in connection with the inference procedure. The clusters inferred by the animal may not be identical to the true causes of its observations.

[5] This "infinite-capacity" distribution over causes is known in statistics and machine learning as a *Chinese restaurant process* (Aldous, 1985; Pitman, 2002). We use the Chinese restaurant process as an intra-agent prior on the structure of the environment, which is the basis (and provides the constraints) for the inference process once observations are seen. A Chinese restaurant process is a probability distribution over partitions of observations, where a partition is a vector indicating the latent cause of each observation. Its name comes from the following metaphor: Imagine a Chinese restaurant with an unbounded number of tables (causes). The first customer (observation) enters and sits at the first table. Subsequent customers sit at an occupied table with a probability proportional to how many people are already sitting there, and at a new table with probability proportional to alpha, a concentration parameter. Once all the customers are seated, one has a partition of observations into causes. In a Chinese restaurant process mixture model (also known as a Dirichlet process mixture model), each cause is associated with a parameterized distribution over features so that an observation's feature properties are determined by its latent cause (and its associated parameters). Observations generated by the same cause will tend to have similar features by virtue of sharing these parameters.

reinforcement) for feature i, given latent cause k. A common assumption in mixture models (which we adopt here) is that, in the generative model, each feature is conditionally independent of all the other features given a latent cause and the multinomial parameters. For instance, a latent cause that can be labeled as "training trial" might generate a cue with probability $\phi_{2,\text{tone},k = \text{"training"}} = 1$ and, independently, generate reinforcement with some probability $\phi_{1,\text{reinforcement},k = \text{"training"}}$ (possibly less than 1), whereas a latent cause labeled as "extinction trial" might generate a cue with probability 1 and reinforcement with probability 0. The conditional independence assumption expresses the idea that, given the identity of the latent cause, cues and reinforcement are separately generated, each according to its associated probability $\phi_{i,j,k}$.

We assume that the multinomial parameters themselves are drawn from a Dirichlet distribution (the conjugate prior for the multinomial distribution). This prior expresses the animal's predictions about the experiment before it has made any observations. Given that the animal is unlikely to have strong a priori predictions about the experiment before it has begun, we chose the parameters of the Dirichlet distribution so that all possible multinomial parameters have equal probability under the prior. Note that each cause is endowed with its own multinomial distribution over features; this allows different causes to be associated with different observation statistics. Every time a new cause is created by Equation 1, the parameters of its corresponding multinomial distribution are drawn from the Dirichlet prior.

It may at first appear odd that causes in Equation 1 are generated purely on the basis of how many times a particular cause was generated in the past and that features are generated independently from one another given a cause and multinomial parameters. Intuitively, one would expect that similar observations would be generated by the same cause. Indeed, this intuition is faithfully embodied in the model. An important point to keep in mind is that in the generative model, observations will be similar *because* they were generated by the same cause. Similarly, features will exhibit correlations because they are coupled by a common cause (e.g., the latent cause associated with the training phase of a conditioning experiment will tend to generate both the cue and the reinforcement). When faced with uncertainty about the latent causes of its observations, both of these properties will influence the animal's beliefs in the inference model (described in the next section). First, the animal will use the similarity between trials to infer what latent cause they came from. As a result, the belief about the causes of one trial will no longer be independent of the other trials. Second, the animal's beliefs about the future value of one feature (e.g., reinforcement) will depend in the inference model on its knowledge about other features (e.g., context and cue). In other words, observation features will be conditionally dependent when the latent cause is unknown (i.e., in all realistic scenarios).

Inference Model

There are two components to the inference problem facing the animal: identifying the latent causes of its observations and predicting reinforcement given a partial observation (context and cues). Because in our model prediction depends on inferences about latent causes, we address each of these in turn.

Denote the observations in trials $1, \ldots, t$ by the vector $\mathbf{F}_{1:t}$. Given a set of observations up to trial t, what are the animal's beliefs about the latent causes of these observations? According to Bayesian statistical inference, these beliefs are represented by the posterior distribution over partitions given the observations:

$$P(\mathbf{c}_{1:t}|\mathbf{F}_{1:t}) = \frac{P(\mathbf{F}_{1:t}|\mathbf{c}_{1:t})P(\mathbf{c}_{1:t})}{P(\mathbf{F}_{1:t})}. \quad (2)$$

Exact computation of the posterior in this model is computationally demanding. Moreover, for such a model to be plausibly realized by animals, learning and inference must be incremental and online. One approximate inference algorithm which is both tractable and incremental is the *particle filter* (Fearnhead, 2004). This algorithm approximates the posterior distribution over partitions with a set of weighted samples and has been used successfully to model a number of learning phenomena (Brown & Steyvers, 2009; Daw & Courville, 2008; Sanborn, Griffiths, & Navarro, 2006). The essential idea in particle filtering is to create a set of m hypothetical particles, each of which is a specific partition of all the trials into causes, and then weight these particles by how likely they are to have generated the particular set of observations that has been seen. The weights will depend on factors such as whether similar observations are clustered together in a particular particle and the number of latent causes in the partition. They will also depend on multiplicative interactions between features, such that a particle will receive larger weight to the extent that it predicts consistent *configurations* of feature values. A detailed description of the particle-filter algorithm can be found in the Appendix.

We assume that the animal's general goal in a classical conditioning experiment is to predict the probability of reinforcement when observing a "test" observation vector that lacks the first feature (i.e., where it is not yet specified whether reinforcement will or will not occur). This prediction can rely on the presence or absence of the other features (context and cue) as well as all of the animal's previous experience. In our model, this prediction is accomplished by augmenting each particle with a cluster assignment of the test observation and then averaging the probability of reinforcement over all the particles, weighted by the posterior probability of the test cluster assignment (see the Appendix for the corresponding equations). We assume that the animal's conditioned (Pavlovian) response is proportional to the predicted probability of reinforcement (Dayan et al., 2006) and so report the reinforcement prediction in the results.

Results

Except where otherwise mentioned, for the simulations reported here, we used uniform Dirichlet priors over all features and $\alpha = .1$ as the concentration parameter.[6] All the simulations used 3,000

[6] Although alpha can be learned straightforwardly with the particle filter, our simulations suggest that this added flexibility does not change the results substantially, so we have fixed it to a constant value.

particles.[7] For each phase (pre-exposure, conditioning, extinction), trials were identical replicas of each other (i.e., there was no noise injected into the observations). Although the output of the particle filter is stochastic (because of the sample-generating process), it returns effectively the same results on multiple runs by averaging over a large number of particles.

Acquisition and Extinction

As a preliminary illustration of the model's basic behavior, Figure 1 shows simulated conditioned responding before and after a context change. The solid line represents conditioned responding when the animal continues to be reinforced after the context change. The dashed line represents conditioned responding when the animal is no longer reinforced after the context change (i.e., extinction). As expected (and in agreement with most models of conditioning), responding increased during the conditioning phase and decreased during the extinction phase. Note that the conditioned response began at 50%, which follows from the uniform prior over the reinforcement feature. In other settings, a different prior may be more suitable (e.g., to express the animal's prior expectation that it will not get reinforced).

Renewal

Figure 2a shows experimental data from a renewal paradigm (Bouton & Bolles, 1979), in which rats were given training in Context A and extinction in Context B and were then tested in either the training context (A), extinction context (B) or a novel context (C). The conditioned response measured at test was, in this case, conditioned suppression (but similar results have been obtained with many other preparations; see Bouton, 2004). Conditioned responding was renewed both in the training context (ABA renewal) and in the novel context (ABC renewal) but not in the extinction context.

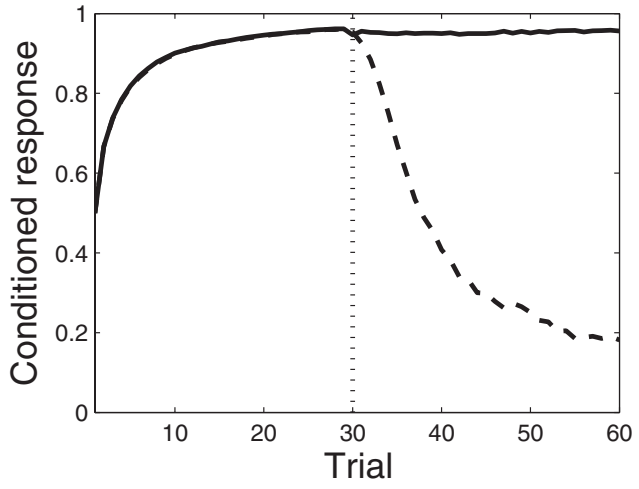

Figure 1. Acquisition and extinction. Simulated conditioned responding during conditioning and extinction in different contexts. The dotted line demarcates the context change. The solid line represents conditioned responding when the animal continues to be reinforced after the context change. The dashed line represents conditioned responding when the animal is no longer reinforced after the context change (extinction). Simulated conditioned responding is directly determined by degree of expectation of reinforcement.

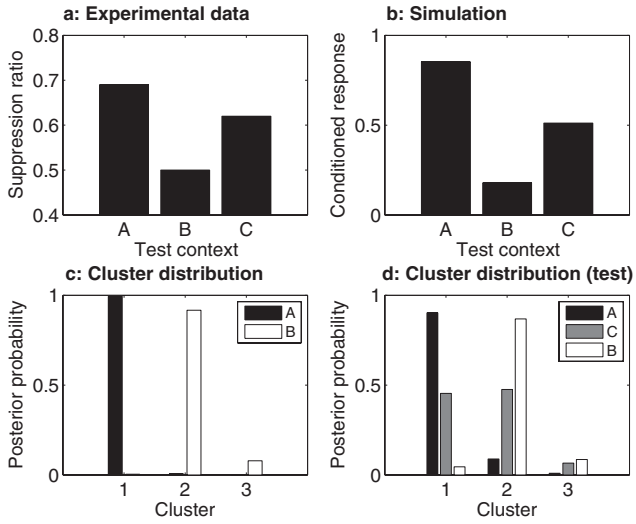

Figure 2. Renewal. Experimental (a) and simulated (b) conditioned responding to a stimulus during a test phase after conditioning and extinction. In all plots, experimentally observed conditioned responses are plotted using their original measurement units. a. Both returning the subject to the conditioning context and placing it in a novel context result in renewal of conditioned responding. Data replotted from Bouton and Bolles (1979). b. Simulated conditioned responding during test in the conditioning Context A, extinction Context B and a novel Context C. c. Posterior distribution of cluster assignments after conditioning in Context A and extinction in Context B. Conditioning and extinction trials tended to be assigned to different clusters, as evidenced by different modes of the posterior in the two phases. d. Posterior distribution of cluster assignments on the first test trial in Contexts A, B and C.

Figures 2c and 2d show the results of simulating our model with conditioning in Context A ($\mathbf{f}$ = [reinforcement, tone, A] for 20 trials); extinction in Context B ($\mathbf{f}$ = [no reinforcement, tone, B] for 50 trials); and testing in either A ($\mathbf{f}$ = [?, tone, B]), B ($\mathbf{f}$ = [?, tone, B]), or C ($\mathbf{f}$ = [?, tone, C]), demonstrating that our model replicates the ABA and ABC renewal effects. Similar in spirit to Redish et al.'s (2007) model, our model predicts ABA renewal as a consequence of the animal's inference that different latent causes are active during conditioning and extinction. When the animal is returned to the conditioning context in the test phase, it infers (because of the presence of contextual cues) that the first latent cause is once again active. Because trials with the same latent cause have similar properties, the animal predicts that reinforcement is likely to occur on the test trials and therefore emits the conditioned behavior. Thus, the importance of context in our theory derives from its usefulness in disambiguating the latent causes of observations (see also Bouton, 1993).

ABA renewal is observed in our model to the extent that a test trial matches (in its observation features) trials from the conditioning phase more than trials from the extinction phase. ABC renewal

[7] We used a large number of particles to accurately approximate the posterior. For this reason, other inference algorithms, such as Gibbs sampling, will produce effectively the same predictions.

may be observed in at least three different scenarios. If C is substantially different from A or B, such that a new cluster is created, ABC renewal will be observed to the extent that the prior expectation of reinforcement in a new cluster is greater than zero. If C is not different enough to warrant a new cluster, ABC renewal may still be observed if C is equally similar to A and B, so that it gets assigned in equal proportions to their associated clusters. Yet another possibility[8] is that when there are many more A trials than B trials, the C observation will be assigned to the cluster associated with A because of Equation 1 (which in the inference model will tend to assign observations to more popular clusters). Although the simulations presented here manifest the second scenario (in which trials in C are associated equally with the training cluster and the extinction cluster), we note that different parameterizations (particularly the value of alpha) or feature representations may result in the first scenario, in which a new cluster is inferred, which would also lead to renewal.

When the animal is tested in the same context as the extinction phase, no renewal is observed (see Figure 2). Similarly, no renewal is observed when all three phases take place in the same context (results not shown). These results follow from the model's prediction that the same latent cause is active during extinction and test and hence predicts the absence of reward. Further insight into the mechanisms underlying renewal in our model can be gained by examining the posterior distribution of clusters, shown in Figure 2c for the conditioning phase and in Figure 2d for the extinction phase. As predicted, our model tends to assign the conditioning and extinction trials to different clusters. When the test trial occurs in Context A, the observation is assigned to the conditioning cluster, whereas when it occurs in Context B, it is assigned to the extinction cluster. When the test trial occurs in a new Context C, inference regarding its latent cause is divided between the conditioning and extinction clusters (and to a lesser extent a new cluster). This is due to the fact that as clusters accrue more observations, they come to dominate the posterior.

Latent Inhibition

Our model similarly explains the context dependence of latent inhibition in terms of the partition structure of the animal's experience. We simulated latent inhibition with 15 pre-exposure trials and 15 conditioning trials (Figures 3a and 3b). When the animal receives pre-exposure ($\mathbf{f}$ = [no reinforcement, tone, A]) and conditioning ($\mathbf{f}$ = [reinforcement, tone, A]) in the same context, it is more likely to attribute a common latent cause to both phases, and thus the properties of both pre-exposure and conditioning observations are averaged together in making predictions about reinforcement in the conditioning phase, leading to a lower prediction and slower acquisition. In contrast, when the animal receives pre-exposure ($\mathbf{f}$ = [no reinforcement, tone, A]) and conditioning ($\mathbf{f}$ = [reinforcement, tone, B]) in different contexts, it is more likely to assign observations from each phase to different clusters—that is, to infer that different latent causes were active during pre-exposure and conditioning. In this case, the reinforcement statistics learned from the conditioning trials are segregated from the reinforcement statistics of the pre-exposure trials, eliminating the retarding effect of pre-exposure on learning, as can be seen in Figures 3a and 3b.

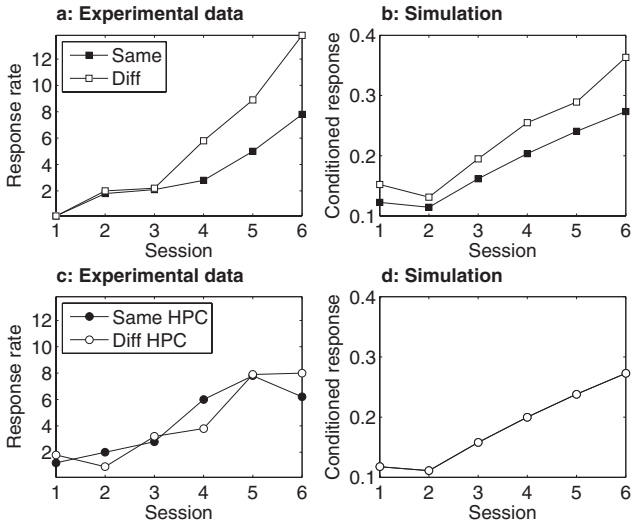

Figure 3. Latent inhibition. Experimental (a, c) and simulated (b, d) acquisition curves of conditioned responding to a stimulus paired with reinforcement as a function of whether unpaired stimulus pre-exposure occurred in the same or in a different (Diff) context. a. Pre-exposure in the same context as conditioning retards the acquisition of conditioned responding. This retarding effect is attenuated by pre-exposing the stimulus in a different context. b. Simulated responding using the mixture model. c. Subjects given hippocampal lesions before conditioning (HPC) show retarded acquisition regardless of whether pre-exposure is performed in the same or in a different context. Data replotted from Honey and Good (1993). d. Simulated responding using the mixture model after hippocampal lesions, which were simulated by restricting the model's ability to infer new clusters. Note that Same HPC is indistinguishable from Diff HPC.

Pathologies of the Model

Numerous studies have shown that damage to the hippocampus disrupts the context dependence of learning and extinction (for a review, see Ji & Maren, 2007). Animals with pretraining electrolytic lesions of the dorsal hippocampus fail to show renewal of conditioned responding (Ji & Maren, 2005). Likewise, animals with hippocampal lesions exhibit intact latent inhibition even when pre-exposure and conditioning occur in different contexts (Honey & Good, 1993). These findings are paralleled by a similar lack of context dependence in the behavior of the developing rat: Before the age of ~22 days, rats do not show renewal or the attenuation of latent inhibition by conditioning in a new context (Yap & Richardson, 2005, 2007). We propose a unified explanation for these phenomena in terms of a pathology in our model's capacity to infer new latent causes. Our theory also suggests an explanation for why the context dependence of renewal and latent inhibition is only impaired when both the conditioning and extinction phases (for renewal) or pre-exposure and conditioning phases (for latent inhibition) occur before maturation (Yap & Richardson, 2005, 2007).

Hippocampal lesions. We propose that hippocampal lesions disrupt the ability of the animal to infer new clusters, restricting its inference to already-established clusters. We implemented this by

[8] We thank an anonymous reviewer for pointing out this possibility.

setting alpha to zero at the time of the lesion. In latent inhibition, when this restriction was applied during pre-exposure, the pre-exposure and conditioning observations were assigned to the same cluster, regardless of the contexts that were in place for the two phases (in other words, our model degenerated into a single distribution over observation features). The prediction of reinforcement during conditioning was then based on an average of both pre-exposure and prior conditioning trials, leading to slower acquisition (see Figures 3c and 3d).

Although early studies reported intact ABA renewal with pretraining electrolytic lesions of the fimbria/fornix (Wilson, Brooks, & Bouton, 1995) or neurotoxic lesions of the entire hippocampus (Frohardt, Guarraci, & Bouton, 2000), Ji and Maren (2005) found that rats with pretraining electrolytic lesions of the dorsal hippocampus showed impaired renewal in the ABA paradigm.[9] Figure 4 shows these experimental data and simulated data from our model, demonstrating impaired renewal in our model after restricting the capacity to infer new clusters prior to training.

Developmental trajectories. Yap and Richardson (2005) have reported that in young rats, latent inhibition is context independent, with behavior being strikingly similar to that exhibited by rats with pretraining hippocampal lesions. As shown in Figure 5a, when rats were pre-exposed, conditioned and tested at 18 days postnatal (PN18), they showed slow acquisition regardless of whether pre-exposure and conditioning occurred in the same or different contexts. In a second experiment, Yap and Richardson (2005) found that if testing was conducted at PN25, the context independence of latent inhibition was still observed. In a third experiment, pre-exposure at PN18 with conditioning at PN24 and testing at PN25 resulted in intact context-dependent latent inhibition. We simulated these different conditions by once again restricting our model's capacity to infer new clusters (setting $\alpha = 0$) during the phases when the animal is younger than PN22, and instating this capacity (setting $\alpha = .1$) when the animal reaches PN22. Figure 5b shows that with this manipulation the mixture model demonstrates a pattern of context dependence similar to that observed experimentally. The explanation of these simulated results is the same as for the effects of pretraining hippocampal lesions described above.

Renewal has also been systematically investigated by Yap and Richardson (2007) in the developing rat. Figure 6a shows conditioned responding in Contexts A and B after conditioning in A and extinction in B at different ages, replotted from Yap and Richardson (2007). The main result is that if both conditioning and extinction are performed before maturity, no ABA renewal is observed, but if extinction is performed after maturity is reached, ABA renewal is intact. Figure 6b shows simulations of these experiments, demonstrating the same pattern of results. Only when our model's capacity for inferring new clusters is restricted during both conditioning and extinction will they be assigned to the same cluster. If extinction occurs after maturation, the animal can assign extinction observations to a new cluster, preventing interference between conditioning and extinction trials and thus enabling the conditioned response in the conditioning context to be renewed after extinction.

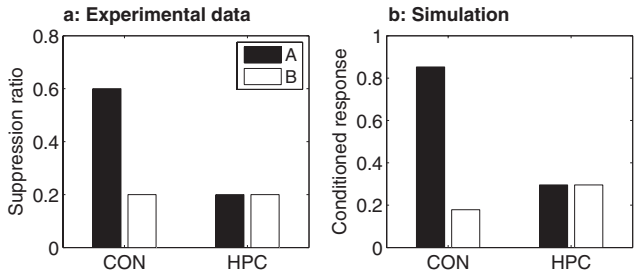

Figure 4. Effect of hippocampal lesions on ABA renewal. a. Experimental conditioned responding to a cue during the test phase in control rats (CON) and those that received pretraining electrolytic lesions of the dorsal hippocampus (HPC). Data replotted from Ji and Maren (2005). b. Simulated conditioned responding following restriction of the model's capacity to infer new clusters prior to training.

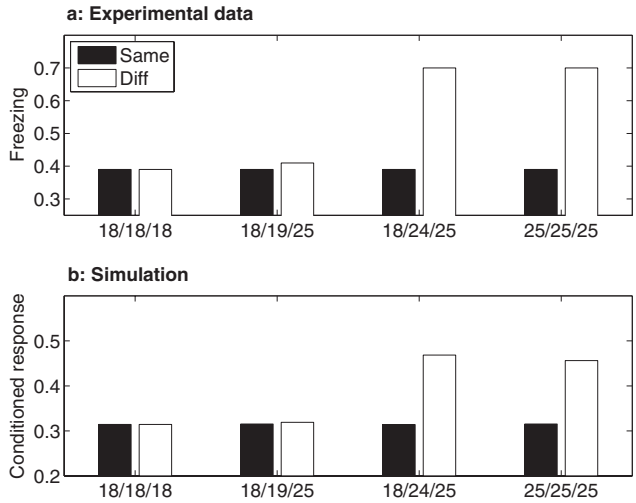

Figure 5. Development of latent inhibition. Experimental (a) and simulated (b) conditioned responding during the test phase following pre-exposure and conditioning in the same or in a different (Diff) context. Labels on the *x*-axis refer to the age (in days) at which each phase (pre-exposure/conditioning/test) was conducted. a. Freezing to the stimulus in the test context. Data replotted from Yap and Richardson (2005). b. Simulated conditioned responding.

Discussion

Starting from a normative statistical framework, we formalized a mixture model of animal learning in which context-dependent behavior is the result of inference over the latent causal structure of the environment. We showed that this model can explain several behavioral phenomena in latent inhibition and renewal paradigms. We also showed that restricting the model's capacity to infer new clusters can reproduce effects of hippocampal lesions and developmental changes in these paradigms.

[9] As discussed by Ji and Maren (2005), electrolytic lesions both damage neurons in the dorsal hippocampus and disrupt fibers of passage to subcortical structures, whereas fornix lesions only disrupt fibers of passage and neurotoxic lesions damage neurons, leaving fibers of passage intact. Thus, it is conceivable that these procedures failed to find impairment in renewal because it is necessary to damage both fibers of passage and hippocampal neurons.

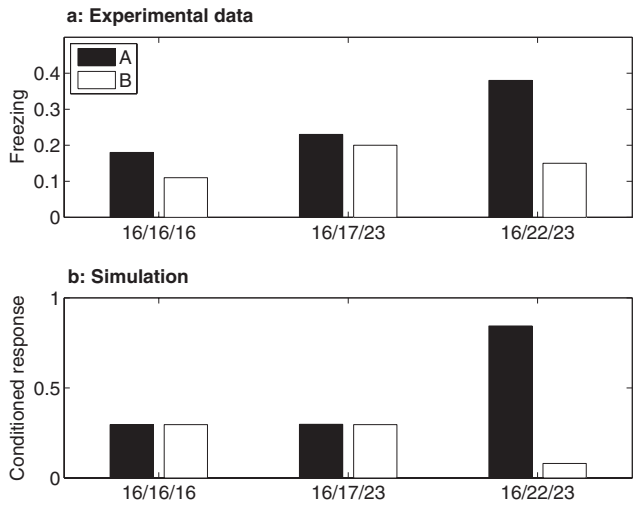

Figure 6. Development of renewal. Experimental (a) and simulated (b) conditioned responding during the test phase in Context A or B following conditioning in A and extinction in B. Labels on the *x*-axis refer to the rat's age (in days) at each phase (conditioning/extinction/test). a. Freezing to the cue in the test context. Data replotted from Yap and Richardson (2007). b. Simulated conditioned responding.

Our model extends and lends statistical clarification to the insights developed in the work of Redish et al. (2007). In addition, we have addressed some specific shortcomings of their model. It is important to note that the dependence of new state inference on negative prediction errors, which prevented Redish et al.'s model from explaining the context specificity of latent inhibition, is wholly absent from our account. We have also made a specific proposal regarding the role of the hippocampus in these tasks, which was alluded to by Redish et al. and is discussed in more detail to follow. But first, to understand the theoretical motivation for a mixture model and its relationship with other models, it is useful to consider a taxonomy of models organized along four dimensions: computational problem, causal structure, capacity, and inference algorithm. In the next four sections we detail each of these dimensions; we then discuss the role of hippocampus in learning and conclude with a discussion of limitations and possible extensions of our model.

Computational Problem: Generative Versus Discriminative

Marr (1980) argued that to understand an information-processing system, one must understand the computational problem it was designed to solve. Most models of animal learning are *discriminative*, implicitly or explicitly assuming that animals aim to predict the probability of reinforcement given the rest of their experience (Pearce & Bouton, 2001). *Generative* models, in contrast, assume that animals aim to learn the joint distribution over all variables in their internal model of the environment, including but not limited to reinforcement. Courville (2006) developed a generative model of animal learning using sigmoid belief nets that bears many similarities to our mixture model; we discuss similarities and differences in the following sections.

One might reasonably ask "Why favor a generative account over a discriminative one?" One problem with discriminative models is that they have no means of explaining behavioral phenomena in which the animal appears to learn information about the environment independent of reinforcement. A classic example of this is sensory preconditioning (Brogden, 1939): After initially pairing two neutral stimuli, A and B, in the absence of reinforcement, A is paired with reinforcement; subsequently, the animal exhibits significant responding to B, suggesting that an association between A and B was learned in the first phase, despite the absence of reinforcement. Courville (2006) reviewed numerous other phenomena that support a generative account.

Larrauri and Schmajuk (2008) proposed a discriminative connectionist model to account for renewal and several other context-dependent behaviors. They argued that a combination of attentional, associative and configural mechanisms can collectively account for these data. As pointed out by Courville and colleagues (Courville et al., 2002; Courville, 2006), many of the ideas behind configural mechanisms can be captured by latent variable models. Whereas in connectionist models, observation features are coupled via convergent projections to "configural" units, latent variable models capture this coupling generatively by having observation features share a common cause. Generalization between features is then accomplished by learning about their latent causes. In modeling the role of context in learning, we have adopted this same insight, showing how context can affect learning about reinforcers by means of a common latent cause.

A basic property of connectionist models, such as that of Larrauri and Schmajuk (2008), is that they effectively transpose the structure learning problem into a parameter learning problem by encoding all possible structures within the network, allowing the causal structure to be uncovered through experience-dependent adjustment of the connection weights (see also Gluck & Myers, 1993). One problem with this approach is that it ignores prior beliefs about the structure of the environment, which serve to constrain the kinds of structures that can be learned (Kemp & Tenenbaum, 2008; Courville et al., 2004). Our generative model is a middle ground between connectionist models that assume no prior structural beliefs and models that use hand-coded features for particular tasks (e.g., Brandon, Vogel, & Wagner, 2000). In our model, where exactly in this middle ground the animal's structural beliefs lie is determined both by its experience and its prior beliefs. The animal may initially expect only a small number of latent causes (specified by setting alpha close to zero), but its generative model is flexible enough to allow revision of this belief to accommodate more causes in light of new observations.

Causal Structure: Products Versus Mixtures

Recall that in a mixture model, observations are assumed to be caused by a generative model in which a single discrete cause is sampled and then an observation is sampled conditional on this cause. An alternative generative model, known as a *product* model, assumes that observations are generated by a linear combination of several latent causes, any number of which can be present at the same time.

As far back as the influential model of Rescorla and Wagner (1972), the predominant mathematical representation in models of animal learning is the product. In discriminative models, for ex-

ample, the prediction of reinforcement is computed by taking a linear combination of feature variables. This is true not just for the Rescorla-Wagner model but for most models in the statistical and connectionist traditions as well (Dayan & Long, 1998; Schmajuk, Buhusi, & Gray, 1996). In Courville's (2006) generative model, the probability of each observed variable is a linear combination of latent variables passed through a logistic sigmoid function.

An exception is the competitive mixture of experts model of Dayan and colleagues (Dayan & Long, 1998; Dayan et al., 2000), in which reinforcement is assumed to be generated by a single *observable* cause. In that model, the probability of reinforcement is the sum of conditioned probabilities of reinforcement given each cause, weighted by the probability of observing that cause (its "mixing probability"). One motivation for using mixtures rather than products, articulated by Dayan et al. (2000), is that inference within a mixture provides an elegant model of competitive attentional allocation in animal learning (whereby stimulus features are attended in proportion to the posterior probability that they caused reinforcement) and may be necessary to explain effects like downward unblocking (Holland, 1988). Our model, although consistent with a competitive attentional account, puts mixtures to a different use by assuming that reinforcement is generated by a *latent* cause. There are many situations in which this assumption is reasonable. Indeed, if one contemplates the designs of most conditioning experiments (including those modeled by Dayan & Long, 1998), the stimulus patterns presented to the animals are generated by discrete, latent phases of the experiment (e.g., conditioning, extinction); the animal never directly observes these phases, but inferring them is key to predicting reinforcement.

Fuhs and Touretzky (2007) proposed a latent cause theory to explain hippocampal place cell remapping that is similar in spirit to our mixture model. They defined context as a statistically stationary distribution of observations and context learning as the task of clustering observations together into groups with local statistics that are stationary in time. In contrast with our static mixture model, they used a dynamic mixture model and formalized context learning in terms of Bayesian model selection, showing that this can predict when place cells will remap in response to environmental change. As with Courville's (2006) model, they selected the best finite-capacity mixture model (see the next section), whereas we employ an infinite-capacity mixture model that automatically selects the number of clusters on the basis of its observations. Because they applied this model to neural data and behavioral paradigms somewhat removed from our focus in this article, a direct comparison between the two models is difficult. Nonetheless, the idea that hippocampal place cells are important for inferring latent causes is consonant with the general view of the hippocampus set forth in this article (see **The Hippocampus and Context** section).

Capacity: Finite Versus Infinite

A special problem vexes models with latent variables in which the number of latent variables is unknown. One can almost always increase the likelihood of the data under a model by increasing the number of parameters in the model. The number of parameters, or more generally the complexity of the model, is sometimes referred to as its "capacity." Increasing capacity can lead to a phenomenon known as "overfitting," wherein extra parameters are just capturing noise, leading to poor predictive power. A principled statistical approach to this problem is to represent uncertainty over the model's structure explicitly and infer both the structure and the values of the latent variables. This was the approach adopted by Courville (2006), who used a Markov chain Monte Carlo algorithm to select the best finite capacity model (a model with a fixed number of parameters) given the data.

An alternative to selecting between different finite capacity models is to allow the number of parameters to grow with the data (i.e., infinite capacity). This is, in fact, the spirit of Redish et al.'s (2007) model, in which heuristic modifications to a reinforcement learning algorithm allow it to increase its capacity (by expanding the state space) during learning. To control overfitting, one can place a prior distribution over parameters that expresses a preference for simpler models. This approach, adopted in our model, satisfies certain intuitions about an animal's representation of its environment. It seems unreasonable to assume that the animal knows in advance how many hidden causes to which it might be exposed. A more reasonable assumption is that it infers that a new hidden cause is active when the statistics of its observations (e.g., lights, tones, odors) change, which is precisely the inference procedure imputed to the animal by the mixture model. Similar arguments have also been made by Sanborn et al. (2006) in their mixture model of human categorization.

Another aspect that our model shares with Redish et al.'s (2007) model is that cluster assignment (state classification) and cluster creation (expansion of the state space) are both determined by the similarity between the current observation and the existing states. A current observation is assigned to an existing cluster to the extent that it is similar to the other observations assigned to that cluster; if no cluster is sufficiently similar, a new cluster is created for that observation. In essence, the particle-filter algorithm attempts to create clusters with maximal within-cluster similarity and minimal between-cluster similarity. The state-classification mechanism in Redish et al.'s model also attempts to achieve this goal, but it lacks a direct statistical interpretation in terms of a well-defined inference procedure.

Redish et al.'s model does not represent uncertainty about the state classifications, whereas the particle filter maintains an approximation of the full posterior distribution over clusters.[10] This is potentially important in cases where previous clustering needs to be reevaluated in light of later information. For example, imagine coming home and seeing the house flooded. You could classify this as either resulting from the (latent) cause "it has rained" or the a priori much less probable (latent) cause "there was a fire and fire trucks sprayed my house." Later hearing on the news that it had been an exceptionally hot and dry day, you might reevaluate the fire hypothesis. Such *retrospective revaluation* phenomena (in which a previously disfavored interpretation becomes favored in light of new information) support the idea that humans and animals represent uncertainty about past interpretations, rather than making hard assignments (Daw & Courville, 2008).

[10] When the number of particles is small, particle filtering will behave similarly to hard assignment (see Daw & Courville, 2008; Sanborn et al., 2006).

Inference Algorithm: Batch Versus Incremental

One of the reasons for appealing to statistical models of learning is that they provide a formal description of the computational problem that the learning system is designed to solve. However, a complete analysis of an information processing system requires descriptions at two other levels (Marr, 1980). The *algorithmic* level specifies the operations and representations required to solve the computational problem. In a statistical model, the representations are probability distributions and the operations are usually some form of the product and sum rules from probability theory. The *implementational* level specifies how these computations are physically realized (e.g., in the brain). Statistical models of animal learning vary in their plausibility at these two levels of analysis (we discuss the implementational level in the next section).

At the algorithmic level, the main desideratum for plausibility is that the inference procedure be able to incorporate new data incrementally (Anderson, 1991; Sanborn et al., 2006). Reinforcement learning and connectionist updates satisfy this desideratum. The batch Markov chain Monte Carlo algorithm proposed by Courville et al. (2002, 2004), which must be rerun on all past observations after each trial, suffers in this regard, although later work attempted to remedy this drawback (Courville, 2006; Daw & Courville, 2008). We used the particle filter to perform inference in our model because it provides a cognitively plausible incremental algorithm for animal learning. However, it would be premature to commit to the particle filter as an algorithmic-level description of the conditioning data that we model, because given the large number of particles we use, this algorithm will make behavioral predictions essentially identical to those made by any other algorithm that adequately approximates the posterior (e.g., Markov chain Monte Carlo sampling). With fewer samples, the particle filter approximates the posterior only crudely. It has been argued that this might be the reason for certain kinds of resource limitations on behavior (Brown & Steyvers, 2009; Daw & Courville, 2008; Sanborn et al., 2006); it is an open question whether such resource limitations are evident in the renewal or latent inhibition data.

The Hippocampus and Context

The hippocampus has long been implicated in context learning, but theories have differed in their formal characterization of this role (Fuhs & Touretzky, 2007; Hasselmo & Eichenbaum, 2005; Hirsh, 1974; Howard, Fotedar, Datey, & Hasselmo, 2005; Jarrard, 1993; Nadel, 1995; Nadel & MacDonald, 1980; O'Keefe & Nadel, 1978; Redish, 1999; Rudy & O'Reilly, 1999). Here we have proposed one possible role for the hippocampus in inferring latent causes. We showed that restricting our model's ability to infer new clusters results in behavior qualitatively similar to that observed in rats with hippocampal lesions (see also Love & Gureckis, 2007, for a similar interpretation of human data). We believe that the hippocampus is suited for this role, with its ability to extract sparse codes from sensory inputs, which could support the learning of discrete latent causes (Doboli, Minai, & Best, 2000). In particular, sparse projections from the dentate gyrus to CA3 are thought to be crucial for *pattern separation* (Marr, 1971), an operation that could serve to separate different observations (inputs) into distinct activation patterns in CA3. When a partial pattern (e.g., a stimulus and context) is presented, the missing part of the pattern (e.g., reinforcer) is activated by means of recurrent connections in CA3, which may function as an attractor network (McNaughton & Morris, 1987). These attractors may thus correspond to inferred clusters, with new attractors being formed when the input statistics change dramatically.

Our model may also shed new light on a long-standing question about the hippocampus and memory in general (Marr, 1971; McNaughton & Morris, 1987): When a new observation is made, under what circumstances is a new trace encoded or an old trace retrieved? Our model frames this as a choice between assigning an observation to an existing cluster or to a new cluster. O'Reilly and McClelland (1994) extensively analyzed a model of the hippocampus and argued that its anatomical and physiological properties might serve to minimize the trade-off between pattern separation (encoding) and pattern completion (retrieval). Our model offers a normative motivation for how this trade-off should be balanced on the basis of the animal's observation statistics and prior beliefs, and future work should be directed at connecting it to the underlying neurophysiological mechanisms identified by O'Reilly and McClelland (1994), as well as the roles of theta oscillations and cholinergic input discussed by Hasselmo, Bodelón, and Wyble (2002).

We would like to emphasize that the ostensibly nonstatistical functions of the hippocampus, such as rapid conjunctive encoding (McClelland, McNaughton, & O'Reilly, 1995), are not incompatible with a statistical account. Most distinctions of this sort have identified statistical learning with extraction of the covariation structure of sensory inputs by neocortex (cf. Gluck & Myers, 1993). In neural network models, this is implemented through gradual synaptic weight change. Our model attempts to broaden this view of statistical learning to include the learning of discrete partition structure, a function that we argued fits with existing computational models of the hippocampus.

The fact that infant rats show a lack of context dependence similar to rats with hippocampal damage (Yap & Richardson, 2005, 2007) suggests that the same causal inference mechanism may underlie both phenomena (Martin & Berthoz, 2002; Rudy, 1993), but more research on the behavioral consequences of hippocampal maturation is needed to test this idea. Other brain structures, notably the prefrontal cortex, also undergo maturation during this period, and it is unclear what specific contributions they may make to context-dependent learning and extinction (Quirk, Garcia, & Gonzàlez-Lima, 2005; Rhodes & Killcross, 2007).

We have also said little about one of the main motivations for Redish et al.'s (2007) model, namely the role of the dopamine system in learning. Evidence has begun to accumulate suggesting that the hippocampal and dopamine systems are intricately intertwined (Lisman & Grace, 2005); however, the behavioral significance of this relationship is poorly understood (but see Foster, Morris, & Dayan, 2000; Johnson, van der Meer, & Redish, 2007). Finally, it is important to note that we do not view the role of the hippocampus in causal inference suggested here to be an all-encompassing functional description of the hippocampus. The hippocampus may perform several functions or some more general function that includes causal inference as a subcomponent. Furthermore, inference may rely on the interaction between the hippocampus and other regions in the medial temporal lobe and elsewhere (see Corbit & Balleine, 2000).

Limitations and Extensions

In their article, Redish et al. (2007) also modeled the partial reinforcement extinction effect, the observation that extinction is slower when stimuli are only intermittently paired with reinforcement during training (Capaldi, 1957, 1958). Our model, without further assumptions, cannot demonstrate this effect, which depends crucially on using reinforcement *rate* as a contextual cue. Our model assumes that reinforcements across trials are conditionally independent give their latent causes, and thus it has no representation of reinforcement rate. The essential explanation given by Redish et al. and others (e.g., Courville, 2006) is that the training and extinction contexts are harder to discriminate in the partial reinforcement condition because of smaller differences in reinforcement rate, and thus extinction trials are less likely to be assigned to a new cluster. Redish et al. were able to show this effect primarily because they included the time since last reinforcement, which is inversely correlated with reinforcement rate, in their prototype representation. We have found in simulations (not shown here) that augmenting the observation vector with an additional contextual feature that differs between training and extinction (which could be interpreted as a reinforcement rate cue) is sufficient to produce the partial reinforcement extinction effect. However, an alternative approach to modeling this phenomenon is to incorporate an explicit model of dynamics and change over time. Other extinction phenomena also depend on a richer representation of time than we have employed here. For example, in spontaneous recovery, simply waiting 48 hr after extinction is enough to produce renewed responding to the cue. We leave development of a temporally sophisticated mixture model to future work (see, e.g., Ren, Dunson, & Carin, 2008).

Finally, we would like to note that although the formalism employed here appears to be a substantial departure from the type of reinforcement learning model used by Redish et al. (2007), the difference is not so great at it seems. Note that learning about reinforcement in our model essentially requires that the animal maintain and update a set of sufficient statistics about its beliefs—specifically, the average reinforcement in each cluster for each feature value. We suspect that such sufficient statistics might be learned by a mechanism similar to temporal difference learning and, hence, may similarly rely on the dopamine system (see Daw, Courville, & Touretzky, 2006, for related ideas). However, the potentially rich connections between these formalisms remain to be studied more thoroughly.

Conclusions

We have argued that a wealth of behavioral data are consistent with an account of animal learning in which the animal infers the latent causes of its observations. Drawing on insights from Redish et al. (2007), we formalized this idea as a mixture model and showed how a particle-filter algorithm can be used to perform inference. Simulations show that this framework can reproduce patterns of context-dependent behavior in latent inhibition and renewal paradigms. We also showed that restricting the model's ability to infer new clusters can reproduce patterns of hippocampal damage and developmental change. Our model places context-dependent learning phenomena in a normative statistical framework, which we see as providing a computational-level analysis of the same problems addressed by Redish et al. (2007).

References

Aldous, D. (1985). Exchangeability and related topics. In *École d'Été de probabilités de Saint-Flour xiii* (pp. 1–198). Berlin, Germany: Springer.

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006, February 17). Causal reasoning in rats. *Science, 311,* 1020–1022.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80–99.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory, 11,* 485–494.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10,* 445–466.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: Tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavioral Processes, 9,* 248–265.

Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioral Brain Research, 110,* 67–72.

Brogden, W. (1939). Sensory preconditioning. *Journal of Experimental Psychology, 25,* 323–332.

Brown, S., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology, 58,* 49–67.

Capaldi, E. (1957). The effect of different amounts of alternating partial reinforcement on resistance to extinction. *American Journal of Psychology, 70,* 451–452.

Capaldi, E. (1958). The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *Journal of Comparative and Physiological Psychology, 51,* 367–371.

Corbit, L. H., & Balleine, B. W. (2000). The role of the hippocampus in instrumental conditioning. *Journal of Neuroscience, 20,* 4233–4239.

Courville, A. C. (2006). *A latent cause theory of classical conditioning.* Unpublished doctoral dissertation, Pittsburgh, PA.

Courville, A. C., Daw, N., Gordon, G. J., & Touretzky, D. S. (2004). Model uncertainty in classical conditioning. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (pp. 977–984). Cambridge, MA: MIT Press.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2002). Similarity and discrimination in classical conditioning: A latent variable account. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (pp. 313–320). Cambridge, MA: MIT Press.

Daw, N., & Courville, A. (2008). The rat as particle filter. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (pp. 369–376). Cambridge, MA: MIT Press.

Daw, N., Courville, A., & Touretzky, D. (2006). Representation and timing in theories of the dopamine system. *Neural Computation, 18,* 1637–1677.

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems.* Cambridge, MA: MIT Press.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience, 3*(Suppl), 1218–1223.

Dayan, P., & Long, T. (1998). Statistical models of conditioning. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (pp. 117–123). Cambridge, MA: MIT Press.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. (2006). The misbehavior of value and the discipline of the will. *Neural Networks, 19,* 1153–1160.

Doboli, S., Minai, A. A., & Best, P. J. (2000). Latent attractors: A model for context- dependent place representations in the hippocampus. *Neural Computation, 12,* 1009–1043.

Fearnhead, P. (2004). Particle filters for mixture models with an unknown number of components. *Journal of Statistics and Computing, 14,* 11–21.

Foster, D. J., Morris, R. G., & Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus, 10,* 1–16.

Frohardt, R. J., Guarraci, F. A., & Bouton, M. E. (2000). The effects of neurotoxic hippocampal lesions on two effects of context after fear extinction. *Behavioral Neuroscience, 114,* 227–240.

Fuhs, M. C., & Touretzky, D. S. (2007). Context learning in the rodent hippocampus. *Neural Computation, 19,* 3173–3215.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis.* Boca Raton, FL: Chapman and Hall/CRC.

Gluck, M., & Myers, C. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus, 3,* 491–516.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121–134.

Hall, G., & Honey, R. C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes, 15,* 232–241.

Hasselmo, M. E., Bodelón, C., & Wyble, B. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation, 14,* 793–817.

Hasselmo, M. E., & Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks, 18,* 1172–1190.

Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation*. Boston, MA: Addison Wesley.

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioral Biology, 12,* 421–444.

Holland, P. C. (1988). Excitation and inhibition in unblocking. *Journal of Experimental Psychology: Animal Behavioral Processes, 14,* 261–279.

Honey, R. C., & Good, M. (1993). Selective hippocampal lesions abolish the contextual specificity of latent inhibition and conditioning. *Behavioral Neuroscience, 107,* 23–33.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological Review, 112,* 75–116.

Jarrard, L. (1993). On the role of the hippocampus in learning and memory in the rat. *Behavioral and Neural Biology, 60,* 9–26.

Ji, J., & Maren, S. (2005). Electrolytic lesions of the dorsal hippocampus disrupt renewal of conditional fear after extinction. *Learning and Memory, 12,* 270–276.

Ji, J., & Maren, S. (2007). Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus, 17,* 749–758.

Johnson, A., van der Meer, M. A., & Redish, A. D. (2007). Integrating hippocampus and striatum in decision-making. *Current Opinion in Neurobiology, 17,* 692–697.

Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review, 109,* 2002.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 10687–10692.

Larrauri, J. A., & Schmajuk, N. A. (2008). Attentional, associative, and configural mechanisms in extinction. *Psychological Review, 115,* 640–676.

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron, 46,* 703–713.

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience, 7,* 90–108.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society London B, Biological Sciences, 262,* 23–81.

Marr, D. (1980). *Vision.* San Francisco, CA: Freeman.

Martin, P. D., & Berthoz, A. (2002). Development of spatial firing in the hippocampus of young rats. *Hippocampus, 12,* 465–480.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences, 10,* 408–415.

Nadel, L. (1995). The role of the hippocampus in declarative memory: A comment on Zola-Morgan, Squire, and Ramus (1994). *Hippocampus, 5,* 232–239.

Nadel, L., & MacDonald, L. (1980). Hippocampus: Cognitive map or working memory? *Behavioral and Neural Biology, 29,* 405–409.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* New York, NY: Oxford University Press.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus, 4,* 661–682.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111–139.

Pitman, J. (2002). *Combinatorial stochastic processes* (Notes for Saint Flour Summer School; Technical Report no. 621). Department of Statistics, University of California, Berkeley.

Quirk, G. J., Garcia, R., & Gonzàlez-Lima, F. (2006). Prefrontal mechanisms in extinction of conditioned fear. *Biological Psychiatry, 60,* 337–343.

Redish, A. (1999). *Beyond the cognitive map: From place cells to episodic memory.* Cambridge, MA: MIT Press.

Redish, A., Jensen, S., Johnson, A., & Kurth-Nelson, A. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review, 114,* 784–805.

Ren, L., Dunson, D. B., & Carin, L. (2008). The dynamic hierarchical dirichlet process. In *Proceedings of the 25th international conference on machine learning* (pp. 824–831). Helsinki, Finland: ACM.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rhodes, S. E., & Killcross, A. S. (2007). Lesions of rat infra-limbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *European Journal of Neuroscience, 25,* 2498–2503.

Rudy, J. W. (1993). Contextual conditioning and auditory cue conditioning dissociate during development. *Behavioral Neuroscience, 107,* 887–891.

Rudy, J. W., & O'Reilly, R. C. (1999). Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. *Behavioral Neuroscience, 113,* 867–880.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 726–731). Mahwah, NJ: Erlbaum.

Schmajuk, N. A., Buhusi, C., & Gray, J. A. (1996). An attentional-configural model of classical conditioning. *Journal of Mathematical Psychology, 40,* 358.

Schultz, W., Dayan, P., & Montague, P. R. (1997, March 14). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Wilson, A., Brooks, D. C., & Bouton, M. E. (1995). The role of the rat hippocampal system in several effects of context in extinction. *Behavioral Neuroscience, 109,* 828–836.

Yap, C. S., & Richardson, R. (2005). Latent inhibition in the developing rat: An examination of context-specific effects. *Developmental Psychobiology, 47,* 55–65.

Yap, C. S., & Richardson, R. (2007). Extinction in the developing rat: An examination of renewal effects. *Developmental Psychobiology, 49,* 565–575.

Appendix

Particle Filter Algorithm

Recall that for trials 1 . . . t, the vector $\mathbf{c}_{1:t}$ denotes a partition of the trials into clusters and the vector $\mathbf{F}_{1:t}$ denotes the observations for these trials. In our implementation, the particles are generated by sampling from the generative model; this involves, for each particle 1 . . . m, sequentially drawing a hypothetical set of cluster assignments from Equation 1. The posterior is then approximated by a weighted sum of delta functions placed at the particles:

$$P(\mathrm{c}_{1:t} = \mathrm{c}|\mathbf{F}_{1:t}) \approx \sum_{l=1}^{m} w_t^{(l)} \delta[\mathrm{c}_{1:t}^{(l)}, \mathrm{c}], \tag{A1}$$

where $\mathbf{c}_t^{(l)}$ is the partition in particle l and $\delta[\,\cdot\,,\cdot\,]$ is 1 when its arguments are equal and 0 otherwise. The importance weight $w_t^{(l)}$ is proportional to the likelihood of observation f_t under the partition in particle l:

$$w_t^{(l)} \propto P(\mathrm{f}_t|\mathrm{c}_{1:t}^{(l)}, \mathbf{F}_{1:t-1}) = \prod_i P(f_{t,i}|\mathrm{c}_{1:t}^{(l)}, \mathbf{F}_{1:t-1}). \tag{A2}$$

Note that the weight depends only on the likelihood of the current observation because the particles are resampled according to their weights at the beginning of each trial (see below); after resampling, the (unweighted) particles are distributed according to the posterior. Using a standard calculation for the Dirichlet-Multinomial model (Gelman, Carlin, Stern, & Rubin, 2003), we can analytically integrate out the multinomial parameters ϕ_k associated with each cause to obtain the following expression for the likelihood:

$$P(f_{t,i} = j|\mathrm{c}_t^{(l)} = k, \mathrm{c}_{1:t-1}^{(l)}, \mathbf{F}_{1:t-1}) = \int_{\phi_k} P(f_{t,i} = j|\mathrm{c}_t^{(l)} = k, \mathrm{c}_{1:t-1}^{(l)}, \mathbf{F}_{1:t-1}, \phi_k) P(\phi_k) d\phi_k = \frac{N_{i,j,k}^{(l)} + 1}{\sum_j (N_{i,j,k}^{(l)} + 1)}, \tag{A3}$$

where $N_{i,j,k}^{(l)}$ is the number of observations with feature value j on dimension i that were generated by cause k in particle l. Note that $N_{i,j,k}^{(l)}$ depends on $\mathbf{F}_{1:t-1}$. Although not immediately evident in these equations, learning occurs through maintaining and updating the sufficient statistics of each cluster, namely the cluster-feature co-occurrence counts (encoded by $N_{i,j,k}^{(l)}$). As we show below, these sufficient statistics can be used to predict reinforcement given a subset of the observation features (i.e., a test observation).

The particle filter algorithm proceeds on each trial by

1. Sampling (with replacement) from the current set of particles according to the importance weights.
2. For each particle, sampling a hypothetical cluster assignment for the next observation using Equation 1.
3. Recomputing the weights given the next observation using Equation A2.

Two things should be noted about this algorithm. First, particles will receive higher weight to the extent that observations assigned to the same cluster are similar; this can be seen in Equation A3. Second, the features interact multiplicatively in Equation A2: A particle will receive a large weight only if all the observed features are likely under the particle's partition.

The probability of reinforcement for a test observation is calculated according to

$$P(f_{t,1} = \text{reinforcement}|\mathrm{f}_{t,2:3}, \mathbf{F}_{1:t-1}) = \sum_{\mathrm{c}_{1:t}} P(f_{t,1} = \text{reinforcement}|\mathrm{f}_{t,2:3}, \mathbf{F}_{1:t-1}, \mathrm{c}_{1:t}) \approx \sum_{l=1}^{m} r_t^{(l)} P(f_{t,1} = \text{reinforcement}|\mathbf{F}_{1:t-1}, \mathrm{c}_{1:t}^{(l)}), \tag{A4}$$

where the predictive weight $r_t^{(l)}$ is proportional to the likelihood of the observed features (context and cue):

$$r_t^{(l)} \propto \prod_{i \in \{2,3\}} P(f_{t,i}|\mathrm{c}_{1:t}^{(l)}, \mathbf{F}_{1:t-1}). \tag{A5}$$

Received March 12, 2009
Revision received August 31, 2009
Accepted August 31, 2009 ■