EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, Vol. 35, pp. 1011–1023, 2012 doi:10.1111/j.1460-9568.2011.07920.x

Dissociating hippocampal and striatal contributions to sequential prediction learning

Aaron M. Bornstein[1] and Nathaniel D. Daw[2]
[1]Department of Psychology and
[2]Department of Psychology and Center for Neural Science, New York University, 4 Washington Pl. Suite 888, New York, NY 10003 USA

Keywords: associative learning, hippocampus, human, model-based fMRI, striatum

Abstract

Behavior may be generated on the basis of many different kinds of learned contingencies. For instance, responses could be guided by the direct association between a stimulus and response, or by sequential stimulus–stimulus relationships (as in model-based reinforcement learning or goal-directed actions). However, the neural architecture underlying sequential predictive learning is not well understood, in part because it is difficult to isolate its effect on choice behavior. To track such learning more directly, we examined reaction times (RTs) in a probabilistic sequential picture identification task in healthy individuals. We used computational learning models to isolate trial-by-trial effects of two distinct learning processes in behavior, and used these as signatures to analyse the separate neural substrates of each process. RTs were best explained via the combination of two delta rule learning processes with different learning rates. To examine neural manifestations of these learning processes, we used functional magnetic resonance imaging to seek correlates of time-series related to expectancy or surprise. We observed such correlates in two regions, hippocampus and striatum. By estimating the learning rates best explaining each signal, we verified that they were uniquely associated with one of the two distinct processes identified behaviorally. These differential correlates suggest that complementary anticipatory functions drive each region's effect on behavior. Our results provide novel insights as to the quantitative computational distinctions between medial temporal and basal ganglia learning networks and enable experiments that exploit trial-by-trial measurement of the unique contributions of both hippocampus and striatum to response behavior.

Introduction

Although a behavior may appear to stem from unitary processes, a primary thrust of cognitive neuroscience has been to fractionate its neural causes. In the case of memory, it is clear that different forms are subserved by separate systems (Packard *et al.*, 1989 Knowlton *et al.*, 1994; Knowlton *et al.*, 1996a,b); similarly, different networks relying on distinct representations appear to support distinct strategies for learned decision-making (Dickinson & Balleine, 2002; Niv *et al.*, 2006; Bornstein & Daw, 2011).

In particular, whereas much work has examined the brain's mechanisms for 'model-free' reinforcement learning (RL) of action policies (Houk *et al.*, 1995; Montague *et al.*, 1996), decisions may also be evaluated by anticipating their consequences using learned predictive associations among non-rewarding states or events, as for locations in a cognitive map (Doya, 1999; Daw *et al.*, 2005; Rangel *et al.*, 2008; Redish *et al.*, 2008; Fermin *et al.*, 2010). The psychological, computational and neural processes supporting such 'model-based' RL are comparatively poorly understood and under intensive current study (Balleine *et al.*, 2008; Daw *et al.*, 2011; McDannald *et al.*, 2011; Simon & Daw, 2011). However, in RL tasks, it has been difficult to disentangle the contributions of either strategy to choices (both, after all, ultimately seek reward). Here, we extended methods previously used for the study of RL to examine more directly the learning of sequential, non-rewarded predictive representations, an important subcomponent of model-based RL.

We used reaction times (RTs) as a trial-by-trial index of predictive learning in a serial RT task requiring human subjects to identify images presented in a probabilistic sequence. Predictive learning was demonstrated by facilitated RTs for more probable images (Bahrick, 1954; Harrison *et al.*, 2006; Bestmann *et al.* 2008; den Ouden *et al.*, 2010).

The trial-by-trial pattern by which RTs depended on experience was consistent with learning by a delta rule: a gradually decaying influence, on average, of images observed on previous trials (cf. Corrado *et al.*, 2005; Lau & Glimcher, 2005). A key feature of such a process is the learning rate: how much weight the system places on new information, relative to previous experience. Suggestively, RTs were best explained as resulting from a combination of *two* such processes, each with different learning rates. Such multiplicity might reflect multiple predictive representations supporting the behavior, e.g. response–response sequencing (procedural learning) and stimulus–stimulus associations (relational learning). Of these two sorts of

Correspondence: A. M. Bornstein, as above.
E-mail: aaronb@nyu.edu

Received 27 May 2011, revised 14 September 2011, accepted 21 September 2011

hypothetical representations, only the second might support model-based RL. Although our experimental design cannot explicitly distinguish between the two, we can use our behavioral signatures to uniquely identify neural correlates in regions suggestive of one type of mapping or the other.

We reasoned that if two distinct neural systems underlay this apparently dual-process estimation, then a computational analysis of functional magnetic resonance imaging (fMRI) data could dissociate their activity via this parameter (Gläscher & Büchel, 2005). We used a computational model of learning to analyse the fMRI data (O'Doherty *et al.*, 2007), seeking correlates of predictive learning and estimating the implied learning rates. Studies using related tasks have identified a distributed network of regions, including hippocampus and striatum, involved in contingency estimation (Strange *et al.*, 2005; Harrison *et al.*, 2006). Here, we decompose this apparently unitary network into separable subnetworks, each displaying a learning rate that corresponds uniquely to one of those estimated behaviorally. This approach allows us to dissociate the individual contributions of hippocampus and striatum to trial-by-trial response behavior, and, consequently, measure neural activity reflecting sequential predictive learning specific to each of these structures.

Methods

Participants

Twenty right-handed individuals (nine female; ages 18–32 years, mean 25) participated in the study. All had normal or corrected-to-normal vision. Participants received a fixed fee, unrelated to performance, for their participation. Participants were recruited from the New York University community as well as the surrounding area and gave informed consent in accordance with procedures approved by the New York University Committee on Activities Involving Human Subjects.

Exclusion criteria

Data from two participants were excluded due to failure to demonstrate learning of the sequential contingencies embedded in the task. Failure to learn was identified when a regression model with only nuisance regressors (the 'constant' model) proved a statistically superior explanation of participant RTs than any of the other models considered here, which each include regressors of interest specifying the estimated conditional probability of images (see Analysis, below). Statistical superiority was measured by the Bayesian information criterion (BIC; Schwarz, 1978), used to correct likelihood scores when comparing models with different numbers of parameters.

Task design

Participants performed a serial reaction time (SRT) task in which they observed a sequence of image presentations and were instructed solely to respond to each image using a pre-trained key-press assigned to that image. The stimulus set consisted of four grayscale photographs of natural landscapes that were matched for size, contrast and luminance (Wittmann *et al.*, 2008). Each participant viewed the same four images. During behavioral training, the keys corresponded to the innermost fingers on the home keys of a standard USA-layout keyboard (D, F, J, K). For the MRI sessions, the same finger keys were used on two MR-compatible button boxes (Fig. 1A). Participants were instructed to learn the responses as linking a finger and an image, rather than a key and an image (e.g. left index finger, rather than 'F'). The mappings between the four images and four keys were one-to-one, pseudorandomly generated for each participant prior to their training session, trained to the criterion prior to the fMRI session, and maintained constant during the experiment. The experiment was controlled by a script written in Matlab (Mathworks, Natick, MA, USA), using the Psychophysics Toolbox (Brainard, 1997).

At each trial, one of the pictures was presented in the center of the screen, where it remained for 2 s, plus or minus a small, pseudorandom amount of jitter time, up to 220 ms in increments of 55 ms. Participants were instructed to continue pressing keys until they responded correctly or ran out of time. Correct responses triggered a gray bounding box which appeared around the image for the lesser of 300 ms or the remaining trial time (Fig. 1B). Thus, each image presentation occurred for the programmed amount of time, regardless of participant response. The inter-trial interval consisted of 220 ms of blank screen.

The training phase of the experiment was conducted outside of the scanner, seated upright, with responses provided on a standard PC keyboard. During this phase, participants were trained to a criterion level of accuracy, defined as 75 correct first responses out of at most the previous 100 presentations.

A second practice session of 150 presentations was conducted inside the scanner, to ensure that participants attained reasonable comfort and proficiency with the magnet-safe button boxes used to collect responses. Neuroimaging data were not collected during this practice session. The finger-to-image response mappings generated for the training session were preserved for the scanner session.

The test phase of the scanning session proceeded with four blocks of 249 presentations. The first three blocks were followed by a rest period of participant-controlled length. During the rest period the participants were presented with a color image of a natural scene not among the study set, and a text reminder to pause and rest before beginning the next run. Scan blocks after the first were initiated manually by the operator only after the participant pressed any of the relevant keys twice to alert the operator that they were prepared to continue the task. Total experiment time – inclusive of training, practice and test – was approximately 1.5 h.

Stimulus sequence

For training, the sequence of images was selected pseudorandomly according to a uniform distribution. Participants were instructed to emphasize learning the mappings between image and finger, disregarding speed of response in favor of correctly identifying the image on the screen.

During the test phase of the experiment, the sequence of images was generated pseudorandomly according to a first-order Markov process, meaning that the probability of viewing a particular image was solely dependent on the identity of the previous image (Fig. 1C). Thus, the statistical structure of the sequence is fully specified by a 4×4 array containing the conditional probabilities that each picture would be followed by the others. Self-transitions were allowed. Participants were not informed of the existence of sequential structure in the task design.

To encourage continual learning, and to sample responses across a wide range of conditional probabilities, the transition matrices were changed twice, at evenly spaced intervals during the experimental session. Despite this, the program offered no explicit indication of the shift to a different transition matrix, nor were changes of matrix aligned with the onset of rest periods. Time to first keypress was

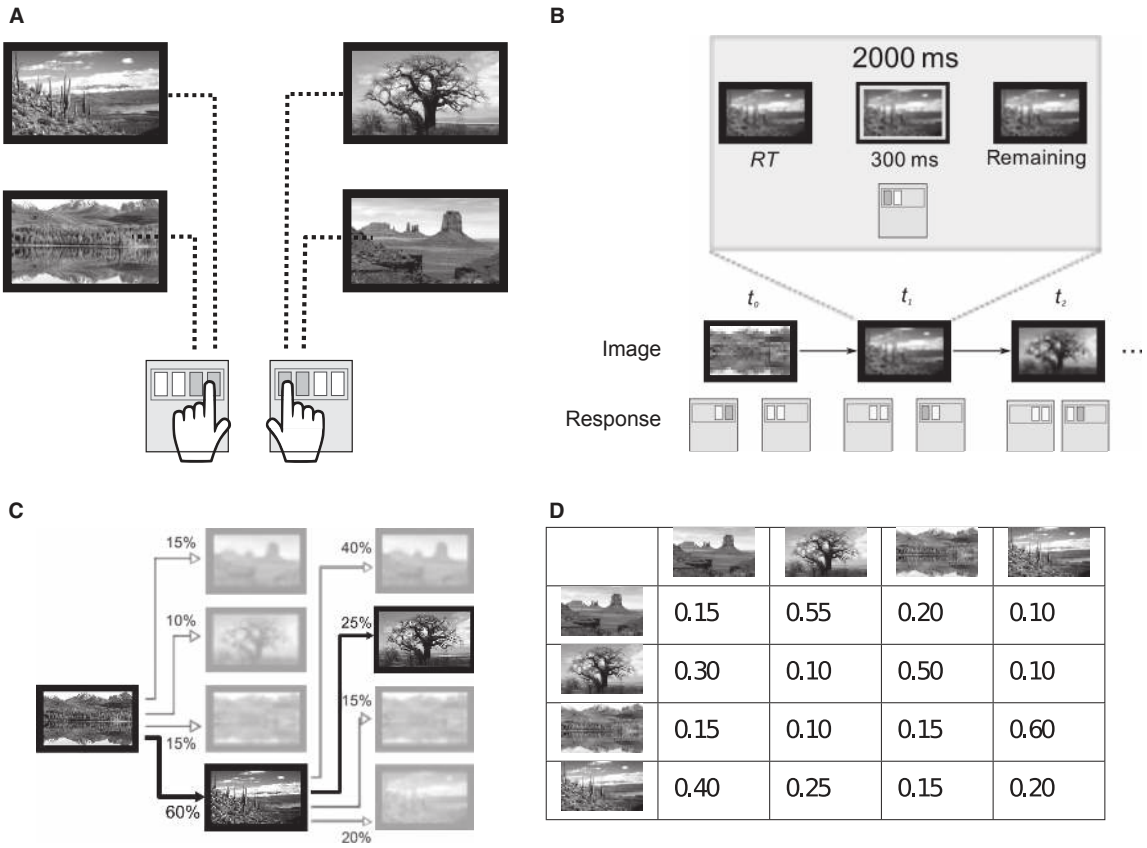

FIG. 1. Task design. (A) Training. Participants were first trained to deterministically associate each of four buttons with one of the stimulus images. Training proceeded until participants reached a fixed accuracy criterion. The associations between stimuli and responses were not varied during the course of the task. (B) Test. Images were presented one at a time for a fixed 2000 ms, regardless of the keypress response. At the first correct keypress, a gray bounding box appears around the image and remains displayed for 300 ms, or until the end of the fixed trial time, whichever is less. Reaction time was recorded to the first keypress. (C,D) Transition structure. Successive images were chosen according to a first-order transition structure, the existence of which was not instructed to the participants. This structure changed abruptly at two points during the task, unaligned to rest periods and with no visual or other notification.

recorded and used as our dependent behavioral variable. Participants were not informed that their RTs were being recorded, and no information was provided as to overall accuracy or speed either during or after the experiment. Trials on which the first keypress was incorrect were discarded from behavioral analysis.

Three transition matrices were generated pseudorandomly for each subject, in a manner designed to balance two priorities: (i) to equalize the *overall* presentation frequencies for each image over the long and medium term, while (ii) examining response properties across a wide sample of *conditional* image transition probabilities. Detailed information on the procedure used to generate matrices satisfying these constraints is available in the Supporting Information (*Transition Matrix Generation*).

Analysis

We employed a series of multiple linear regressions to investigate whether RTs reflected learning of the stimulus–stimulus conditional probabilities, and, subsequently, to examine the form of this learning. In particular, each participant's trial-by-trial RTs for correct identifications were regressed on explanatory variables including the estimated conditional probability of the picture currently being viewed given its predecessor, and the entropy of the distribution of conditional probabilities leading to this picture – defined, in separate models (described below), in a number of different ways representing different accounts of learning – together with several effects of no interest. Trials on which the first keypress was not correct were excluded from behavioral analysis. Effects of no interest included stimulus-self transitions, image-identity effects and a linear effect of trial number. These effects were identical across models – thus, each model-based analysis was uniquely identified by a proposed form for conditional probability estimates.

In our initial analysis, the conditional probabilities (and pursuant entropies) were specified as the ground-truth contingencies – the contingencies actually encoded in the transition matrix. Having established that RT reflected such learning by demonstrating a significant correlation with these asymptotic values (Fig. 2A) of probability (though not entropy, see Results), subsequent analyses used computational models to generate a timeseries of probability estimates such as would be produced by different learning rules with the same experience history as the participant (see Computational Models). These rules involved additional free parameters controlling the learning process (e.g. learning rates), which were jointly estimated together with the regression weights by maximum likelihood. For behavioral analysis, models were fit and parameters were estimated separately for each participant. At the group level, regression weights were tested for significance using a *t*-test on the individual estimates across participants (Holmes & Friston, 1998).

To generate regressors for fMRI analysis (below) we refitted the behavioral model to estimate a single set of the parameters that optimized the RT likelihoods aggregated over all participants (i.e. treating the behavioral parameters as fixed effects). This approach

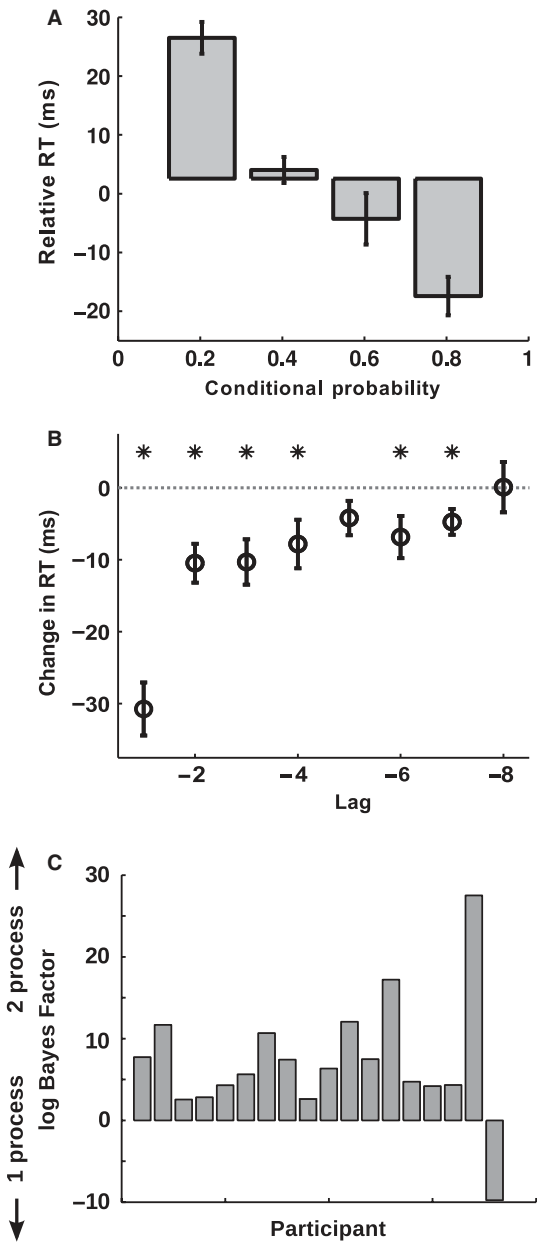

FIG. 2. Sequential learning. (A) Despite the fact that they were unaware of task structure, participant reaction times reflected the probabilities as designed – response time was commensurately lower as conditional likelihood of the image increased. (B) An analysis of the influence of prior responses on reaction time on the current trial shows a decaying effect of previous experience, with significant contributions from the seven most recent presentations of the current image. Reaction time for a given image–image transition was lowered by more recent experience with that transition; this effect showed an exponential relationship between recency of experience and reaction time. This pattern excludes models that do not incorporate forgetting of past experience. $*P < 0.05$. Error bars are SEM. (C) Model comparison. Individual log Bayes factors in favor of a model using two learning processes, vs. a single process. The two-process model is decisively favored for 14 of 18 subjects, and was a significantly better fit across the population (summed log Bayes factor 145, $P < 5 \times 10^{-5}$ by likelihood ratio test).

allowed us separately to characterize baseline learning-related activity and individual variation in neurally implied learning rates relative to this common baseline. For the former, in our experience (Daw *et al.*, 2006, 2011; Schoenberg *et al.*, 2007, 2010; Gershman *et al.*, 2009; Daw, 2010), enforcing common model parameters provides a simple regularization that improves the reliability of population-level neural results. To capture individual between-subjects variation in the learning rate parameter, over this baseline, we add, as an additional random effect across participants, the partial derivatives of the regressors of interest with respect to the learning rate (see Learning rate analysis, below).

Computational models

First, to investigate the contribution of past experience to expectation about the current stimulus, while making relatively few assumptions about the form of this dependence, we entered the past events themselves as explanatory variables in the analysis (Corrado *et al.*, 2005; Lau & Glimcher, 2005). In particular, we used a similar multiple linear regression to the above, but replaced the conditional probability regressor with ten regressors, each a timeseries of binary indicator variables. If $I(t)$ is the image displayed on trial t, then the indicator variables at trial t represented, for each of the ten most recent presentations of the *preceding* image $I(t-1)$, whether on that presentation it had or had not been followed by image $I(t)$. The logic for this assignment is that the expectation of image $I(t)$, conditional on having viewed image $I(t-1)$, should depend on previous experience with how often image $I(t)$ has in the past followed the predecessor image $I(t-1)$. Because error-driven learning algorithms such as the Rescorla & Wagner (1972) model predict that the coefficients for these indicators should decline exponentially with the number of image presentations into the past (Bayer & Glimcher, 2005), we fitted exponential functions to the regression weights.

Following up on the results of this analysis, we considered a more constrained learning rule – again, of the form proposed by Rescorla & Wagner (1972; see also Gläscher *et al.*, 2010) – which updates entries in a 4×4 stimulus–stimulus conditional probability matrix in light of each trial's experience. The appropriate estimate from this matrix at each step was then used as an explanatory variable for the RTs in place of the ground-truth probabilities or binary indicator regressors. Formally, at each trial the transition matrix was updated according to the following rule, for each image i:

$$P(i|I(t-1)) = P(i|I(t-1)) + \alpha(1 - P(i|I(t-1))) \quad i = I(t) \quad (1)$$

$$P(i|I(t-1)) = P(i|I(t-1)) + \alpha(0 - P(i|I(t-1))) \quad i \neq I(t)$$

where $I(t)$ is the identity of the image observed at trial t and α is a free learning-rate parameter. This rule preserves the normalization of the estimated conditional distribution.

In addition, we examined the possibility that behavior may reflect the contributions of two parallel learning processes, by examining the fit of a transition matrix resulting from a weighted combination of two Rescorla–Wagner processes, each with different values of the learning rate parameter α. Each process updated its matrix as above, but the behaviorally expressed estimate of conditional probability was computed by combining the output of each process according to a weighted average with weight (a free parameter) π:

$$\pi \times P_1(I(t)|I(t-1)) + (1-\pi) \times P_2(I(t)|I(t-1)). \quad (2)$$

As the models considered here differ in the number of free parameters, we compared their fit to the reaction time data using Bayes factors (Kass & Raftery, 1995; the ratio of posterior probabilities of the model given the data) to correct for the number of free parameters fit. We approximated the log Bayes factor using the difference

between scores assigned to each model via the Laplace approximation to the model evidence, assuming a uniform prior probability for values of the free parameters between zero and one. In participants for whom the Laplace approximation was not estimable for any model (due to a non-positive definite value of the Hessian of the likelihood function with respect to parameters) the BIC was used for all models. Model comparisons were computed both per individual, and on the log Bayes factors aggregated across the population.

Finally, to evaluate the relative contribution of each process to explaining RTs, we report standardized regression coefficients for the slow and fast conditional probability estimates, e.g.

$$\beta_{p1\text{std}} = \beta_{p1} \times \text{SD}(P_1)/\text{SD}(\text{RT}). \quad (3)$$

fMRI methods

Acquisition

Imaging was performed on the 3T Siemens Allegra head-only scanner at the NYU Center for Brain Imaging, using a Nova Medical (Wakefield, MA, USA) NM011 head coil. For functional imaging, 33 T2*-weighted axial slices of 3 mm thickness and 3 mm in-plane resolution were acquired using a gradient-echo EPI sequence (TR = 2.0 s). Acquisition was tilted in an oblique orientation at 30° to the AC–PC line, consistent with previous efforts to minimize signal loss in orbitofrontal cortex and medial temporal lobe (e.g. Hampton *et al.* 2006). This prescription obtained coverage from the base of the orbitofrontal cortex and medial temporal lobes to the superior border of the dorsal anterior cingulate cortex. Four scans of 300 acquisitions each were collected, with the first four volumes (8 s) discarded to allow for T1 equilibration effects. We also obtained a T1-weighted, high-resolution anatomical image (MPRAGE, $1 \times 1 \times 1$ mm) for normalization and localizing functional activations.

Imaging analysis

Preprocessing and data analysis were performed using Statistical Parametric Mapping software version 5 (SPM5; Wellcome Department of Imaging Neuroscience, London, UK), and version 8 for final multiple comparison correction. EPI images were realigned to the first volume to compensate for participant motion, co-registered to a higher resolution field map with the anatomical image, and, to facilitate group analysis, spatially normalized to atlas space using a transformation estimated by warping the subject's anatomical image to match a template (SPM5 segment and normalize). To ensure that the original sampling resolution was preserved in the normalized space, images were resampled to $2 \times 2 \times 2$-mm voxels in the normalized space. Finally, they were smoothed using a 6-mm full-width at half-maximum Gaussian filter. For statistical analysis, data were scaled to their global mean intensity and high-pass filtered with a cutoff period of 128 s. Volumes on which instantaneous motion was > 0.25 mm in any direction were excluded from analysis. Data from two participants were excluded due to excessive motion on a large percentage of volumes.

Statistical analysis

Statistical analyses of functional time-series were conducted using general linear models, and coefficient estimates from each individual were used to compute random-effects group statistics. Delta-function onsets were specified at the beginning of each stimulus presentation, and – to control for lateralization effects – duplicate nuisance onsets were specified for presentations on which right-handed responses were required. This had the effect of mean-correcting these trials separately. All further regressors were defined as parametric modulators over the initial, two-handed stimulus presentation onsets. All regressors were convolved with SPM's canonical hemodynamic response function.

The remaining regressors were constructed as follows. First, to control for non-specific effects of RT (which, as demonstrated by our behavioral results, was correlated with our primary regressor of interest, the conditional probability), each trial's RT was entered into the design matrix as a parametric nuisance effect. We took advantage of the serial orthogonalization implicit in SPM's parametric regressor construction by placing this regressor first in the set of parametric modulators. As a result, all subsequent regressors, including all regressors of interest, were orthogonalized against this variable, ensuring that it accounted for any shared variance. We next included regressors of interest specifying the conditional probability and conditional entropy associated with the current image, and their partial derivatives with respect to the learning rate (see Learning rate analysis below). Finally, variance due to the effects of missed trials (those in which the participant did not press any keys in the allotted time) and error trials was modeled with additional nuisance regressors, entered last in orthogonalization priority.

Our regressors of interest were derived from the time-series of transition matrices estimated by the best-fitting behavioral model, the two-learning rate model of Eqns 1 and 2. In particular, we include the probability of the image $I(t)$ displayed at each trial t, conditional on its predecessor – $P(I(t) \mid I(t-1))$, and in addition the *entropy* of the distribution over the *subsequent* stimulus, given the image $I(t)$ currently viewed:

$$H(I(t+1)) = -\Sigma_{I(t+1)}[\log(P(I(t+1)|I(t))) \cdot P(I(t+1)|I(t))] \quad (4)$$

where $I(t)$ denotes the image actually displayed on trial t, but the sum is over all four possible identities of the as-yet-unrevealed subsequent image, $I(t+1)$. Whereas the conditional probability measures how 'surprising' is the current stimulus, this quantity, which we refer to as the 'forward entropy,' measures the 'expected surprise' for the next stimulus conditional on the current one, i.e. the uniformity of the conditional probability distribution. Because of the temporal dissociation between probability of the *current* stimulus and entropy of the distribution of *ensuing* stimuli, there was no inherent confound between these two regressors. However, as the entropy regressor was orthogonalized against probability, any shared variance was attributed to conditional probability.

In all analyses, unless otherwise stated, activations are reported for areas where we had a prior anatomical hypothesis at a threshold of $P < 0.05$, corrected for family-wise error (FWE) in a small volume defined by constructing an anatomical mask, comprising the regions of a priori interest, over the population average of normalized structural images. These are, bilaterally, hippocampus as defined by the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002), and anterior ventral striatum (caudate and putamen), defined (after Drevets *et al.*, 2001) by taking the portion of the relevant AAL masks below the ventral-most extent of the lateral ventricles [in Montreal Neurological Institute (MNI) coordinates, $Z < 1$], and anterior to the anterior commissure ($Y > 5$). This area corresponds, functionally, to the regions most often observed to reflect learning-related update signals in fMRI studies of reward learning tasks (Delgado *et al.*, 2000; Knutson *et al.*, 2001; O'Doherty *et al.*, 2003; McClure *et al.*, 2004). Activations outside regions of prior interest are reported if

they exceed a threshold of $P < 0.05$, whole-brain corrected for FWE. All voxel locations are reported in MNI coordinates, and results are displayed overlaid on the average over participants' normalized anatomical scans.

Learning rate analysis

In the best-fitting behavioral model, the learned transition matrix arises from two modeled learning processes each with a free parameter for its learning rate. We used three separate fMRI analyses (and, reported in the Supporting Information, two more specifications) to investigate this multiplicity of potential effects. First, seeking correlates for each of these two subprocesses hypothesized on the basis of the behavior separately, we conducted two generalized linear model (GLM) analyses, each using as regressors of interest the probability and entropy regressors constructed from one of the single learning rates identified from the behavior. These two analyses were conducted in separate GLMs due to correlation between regressors generated using different values of the learning rate parameter. Our third GLM addressed the problem of correlation between signals more directly with a single design that formally investigated the possibility that learning rates expressed across regions of the brain differed from one another. To this end, we employed a GLM that included additional regressors quantifying the effect of changes in the modeled learning rate on the regressors of interest. A detailed description of this analysis is available in the Supporting Information (*Learning rate analysis*).

Results

Behavioral results

Participants performed an SRT task, in which they were instructed to label each of a continuous sequence of image stimuli (Fig. 1B) according to a predetermined one-to-one mapping of each of four keys to each of four natural scene images (Wittmann *et al.*, 2008). Participants were first trained to map fingerpress responses to images (Fig. 1A) at a criterion level of performance (75 correct out of at most the 100 preceding trials). During this training phase, the mean number of trials to criterion was 102.4 [standard deviation (SD) 45.6]. The mean time to correct response was 889.9 ms (SD 443.2 ms).

The main experiment consisted of a sequence of image labeling trials. Images on each trial were selected according to a first-order Markov process, i.e. with a conditional probability determined by the identity of the previous image (see Methods). Participants were not instructed about the existence of sequential structure in the task. During the testing phase, errors – defined as trials in which the first keypress did not correspond to the presented image – were few (mean 2.6%, SD 1.5%), while the mean time to correct response fell relative to training, to 692.7 ms (SD 268 ms).

Ground-truth probability

Figure 2A shows the relationship between RT on correct trials and the ground-truth conditional probability of the image being identified, across the population. Here, for each participant, RTs were first corrected for the mean RT and a number of nuisance effects – estimated using a linear regression containing only these effects as explanatory variables and computing the residual RTs. Of the nuisance regressors, only the self-transition effect was significant across the population ($P < 1 \times 10^{-7}$; all others $P > 0.15$).

The impression that RTs are faster for conditionally more probable images is confirmed by repeating the regressions with the ground-truth conditional probability included as an additional explanatory variable. Across participants (i.e. treating the regression weight as a random effect that might vary between individuals), the regression weight for this quantity was indeed significantly negative (one-sample *t*-test, $P < 2 \times 10^{-7}$; mean effect size 0.76 ms RT per percentage conditional probability) and, at an individual level, reached significance (at $P < 0.05$) for 15 of 20 participants. This analysis contained a second regressor of interest – the entropy of the conditional distribution leading to the current image. Although entropy has previously been shown to impact RTs (Strange *et al.*, 2005), the conditional entropy effect did not reach significance here either across the population ($P > 0.2$) or for any participant individually, and thus was discarded from further behavioral analyses.

This analysis indicates that participant responses were prepared in a manner reflecting some approximation of the programmed transition probabilities. As the probabilities were not instructed, we inferred that these quantities were estimated incrementally by learning from experience. The remainder of our analysis of behavior attempted to characterize the nature of this learning (den Ouden *et al.*, 2010).

Learning analysis

To test the general structure of a learning model, we first examined the contribution of past experience to current expectations, while making few assumptions about the form of this dependence. To this end, we employed a regression model in which previous events were explicitly included as explanatory variables for RT (Lau & Glimcher, 2005). In particular, for each trial, in which some image Y was presented having been preceded by some image X, we included explanatory variables corresponding to each of the last ten previous presentations of X, defined as 1 if that presentation was also followed by Y, and 0 otherwise. The resulting regression weights measure to what extent the RT to Y following X is affected by recent experience with the image pair $X \rightarrow Y$, relative to other pairs $X \rightarrow Z$. In Fig. 2B, the fitted regression weights, up to one more than the most remote to reach significance, are averaged across participants and plotted as a function of the lag into the past, counted as the number of presentations of X.

Consistent with experience-driven learning of conditional probabilities, recent experiences with the image pair $X \rightarrow Y$ predict faster responding. This dependence appears to decay rapidly, although it continues to contribute to current expectations for several presentations – regression weights are significantly non-zero across subjects through roughly the seventh previous observation of X (one sample *t*-tests – at lag 5 $P = 0.11$, all others from 1 to 7 $P < 0.04$; at lag 8 $P > 0.8$). As any image occurs, as expected, every four trials, this suggests that the experience of a transition has a detectable effect over an average window of some 16–28 trials.

This regression characterizes the form of RT dependence on past experiences as a weighted running average, here appearing reasonably exponential. Exponentially decaying weights are characteristic (Bayer & Glimcher, 2005; Corrado *et al.*, 2005; Lau & Glimcher, 2005) of an error-driven learning procedure for estimating conditional probabilities (Rescorla and Wagner 1972; here, Eqn 1), with the free learning rate parameter, α, determining the time constant ($1–\alpha$) of the decay. Thus, the learning rate is equivalent to a 'forgetting', or 'decay' rate (Rubin & Wenzel, 1996; Rubin *et al.*, 1999). The same equations also characterize the average decay behavior for models that update at varying rates or only sporadically (Behrens *et al.*, 2007). However, other sorts of learning rules predict qualitatively different weightings. For instance, because of exchangeability, ideal Bayesian estimation of a static transition matrix (Harrison *et al.*, 2006) or indeed a simple all-trials running average predict equal coefficients at all time lags. On the

basis of these results, we do not consider such models further (although the superiority of the Rescorla–Wagner model was also verified by directly comparing these models' fits to the RTs; analyses not reported).

However, in the domain of decision-making, it has previously been noted that the effects of lagged experiences on choices are better described by the weighted sum of two exponentials with different time constants, a pattern that was suggested to result from the superposition of two underlying processes learning at different rates (Corrado *et al.*, 2005). This is also true of the averaged regression weights in Fig. 2B (likelihood ratio test comparing one- and two-exponential fits, $P = 0.0024$).

Altogether, then, the form of the regression weights suggests that RTs superimpose conditional probability estimates learned using two error-driven learning processes with different learning rates. To verify that this appearance did not arise from the averaging over subjects in Fig. 2B, and to quantify the hypothesis directly in terms of its fit to RTs (rather than parameter estimates from an intermediate analysis), we considered the fit of one- and two-learning rate Rescorla–Wagner models to each participant's RT data, essentially equivalent to refitting the regression model while constraining each participant's weights to follow a one- or two-exponential form.

Figure 2C shows the difference in log Bayes factor between these two models. The two-learning rate model was favored over the one-learning rate model for 17 of 18 participants. Aggregated over subjects, the two-learning rate model was favored by a log Bayes factor of 145. The conditional probabilities learned by the two-learning rate Rescorla–Wagner model also explained the RTs considerably better than the 'ground truth' programmed probabilities (log Bayes factor 308.9 aggregated, and favoring the two-process model for 17 of 18 participants individually). For this two-learning rate model, the mean effect size implied by the regression weights was 1.06 ms RT per percentage point of combined conditional probability. Finally, we measured the relative contribution of each probability estimate to RT, by computing each effect's regression coefficients normalized for variance in each probability time-series. The resulting standardized coefficients were –0.075 for the slow regressor and −0.071 for the fast regressor (both means across the population, with standard error of the mean of 0.01). Thus, the probabilities learned by slower and faster learning rates appear to contribute roughly equally to explaining RTs.

For the purpose of conducting fMRI analyses measuring individual variations in neurally implied learning rate estimates relative to a common baseline (see Methods for an in-depth justification of this approach), we re-estimated the two-process model parameters as single, fixed effects across participants. The best-fit learning-rate parameters were 0.5499 and 0.0138, weighted at 0.4018 to the slow parameter. Additionally, we computed the population median of the random-effects learning rates (0.6054 and 0.008) and weight (0.35 to the slow parameter), and observed that they did not significantly differ from the best-fit fixed-effects values (all $P > 0.1$).

Together, these data suggest that participants learn predictions of conditional probability from experience by an error-driven learning process, and that the data are best explained by the superposition of two such processes learning in parallel.

fMRI results

We next sought signatures in neural activity of the two learning processes suggested by our behavioral analyses. On the basis of previous work on multiple systems involved in sequential learning (e.g. Poldrack *et al.*, 2001; Nomura *et al.*, 2007), we focused on the hippocampus and the anterior ventral striatum as areas of prior anatomical interest. We hypothesized that activity in these two areas might reflect learning at different rates, matching the two processes we inferred from behavior (Gläscher & Buchel, 2005).

We used a strategy of model-based fMRI analysis (Gläscher & O'Doherty, 2010), analogous to an approach often used with reward-related learning. In short, we exploited the fact that the models we fitted to behavior define *internal variables* – here, the learned transition matrices – hypothesized to underlie behavior – here, the RTs. The time-series of these variables from the fitted behavioral models may serve as signatures for ongoing neural processes related to the computations and thus provide quantitative tests for the hypothesized dynamics of these processes. This approach appears well suited to our study, in which the behavioral results suggest two parallel learning processes specifying distinct time-series of values for these internal variables. Analogous work in RL tasks often seeks both anticipatory (future value) and reactive (prediction error) measures of reward prediction; here, we define analogous regressors for stimulus predictions using the forward entropy (cf. Strange *et al.*, 2005; Harrison *et al.*, 2006) of the predicted stimulus distribution, and the conditional probability of an observed stimulus. We adopt the latter regressor rather than its log (a traditional measure of surprise in information-theoretic work), because of its relationship to the prediction error (Eqn 1; Gläscher *et al.*, 2010).

However, simply seeking the correlates of these time-series in a GLM (Tanaka *et al.*, 2004; Gläscher & Buchel, 2005) is statistically inefficient because the versions of our variables of interest calculated at different values of α are highly correlated (average correlation across participants between slow and fast learning rate processes, $R = 0.52$ for probability, 0.31 for entropy). We instead separately studied the regressors from the slow and fast learning rate processes, in distinct GLMs, to form an initial impression of how this activity may fractionate according to processes operating at different learning rates.

Next, to formally answer questions about whether the learning process that best explains blood oxygen-level dependent (BOLD) activity differs between areas, we measured the degree to which BOLD activity in a region reflects a value of the learning rate parameter α that was different from the one tested. In particular, we first identified areas with learning-related activity by assuming a value of the learning rate intermediate between those observed behaviorally, and then analysed residual activity explained by additional, orthogonal regressors representing how the modeled signal would change if the parameter that produced these regressors was increased or decreased.

Forward entropy

Our primary analysis sought regions where activity suggested anticipatory processing, operationalized by the uncertainty about the identity of the next stimulus conditional on the current one (forward entropy; cf. Strange *et al.*, 2005; Harrison *et al.*, 2006). BOLD signals correlating with this time-series may reflect spreading activation among (anticipatory retrieval of) multiple representations in an autoassociative memory network, or, similarly, simulation of future events using a forward model (Niv *et al.*, 2006).

When the regressors were computed according to the slow LR process, correlates emerged in the hippocampus. Activity correlated with the slow process is illustrated in Fig. 3B. In particular, a cluster of significantly correlating activity was identified in left anterior hippocampus (−26, −10, −18; peak FWE $P < 0.02$ small-volume

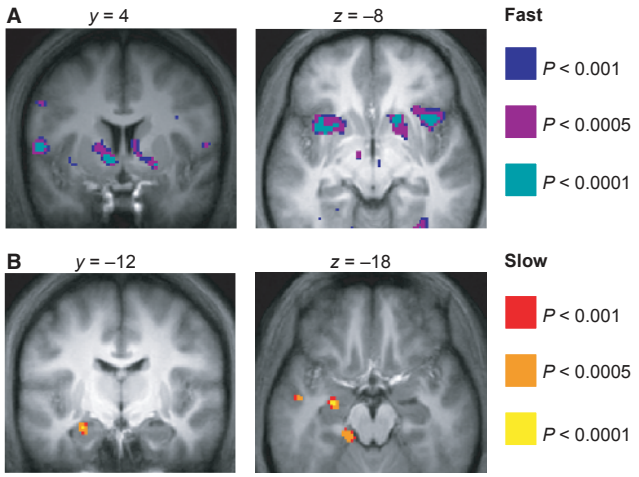

FIG. 3. Areas where the BOLD signal correlated with the entropy over the distribution of upcoming stimuli, generated at each of our analysed learning rates. Images are thresholded at $P < 0.001$, uncorrected, for display purposes. (A) Activation observed in the fast process GLM, with clusters of negative correlation in ventral striatum and anterior insula. (B) Activation observed in the slow process GLM, a positively correlated cluster in anterior hippocampus. The activation visible in posterior parahippocampal cortex did not survive correction for multiple comparisons.

corrected for FWE due to multiple comparisons over our mask of a priori regions of interest, cluster FWE $P < 0.04$).

For regressors computed according to the fast LR process, we observed no activity positively correlated with forward entropy at our threshold. However, a significant cluster of *negatively* correlated activity – possibly reflecting a lower degree of response preparation for the upcoming trial – was identified in right putamen (18, 6, −6; peak marginal at $P < 0.06$ small-volume corrected for FWE, but cluster FWE $P < 0.04$).

Outside our areas of interest, correlates of fast LR forward entropy were also observed in left anterior insula (−40, 14, −4; peak $P < 0.05$ when corrected for FWE over a mask of the whole brain) and inferior frontal gyrus (36, 16, −4; peak $P < 0.05$ by whole-brain FWE correction; not shown). A complete list of clusters correlating with forward entropy can be found in Supporting Information Tables S1 and S2.

Together, these results suggest that anticipatory activity reflecting learned transition contingencies is visible across a number of brain regions. The distinct difference in activations observable in the SPMs from each process further suggests that distinct networks which include either hippocampus or striatum might be associated, respectively, with slow and fast LR estimates of sequential contingencies, and further correlated in opposing directions with the same measure, forward entropy, extracted from each process. In the Supporting Information, we also report results from analyses using probability and entropy regressors derived from the combined process rather than either separately. Correlates there do not reach similar levels of statistical significance, consistent with our interpretation that activity is related to either process separately.

However, it is important to stress that an apparent difference between thresholded statistical maps does not constitute formal demonstration that a difference exists ('the imager's fallacy'; Henson, 2005). Furthermore, the results presented thus far do not directly compare the two processes in a single GLM. Supporting Information Fig. S1 shows that qualitatively similar activations were observed only at a lower statistical threshold (as expected due to correlation between the regressors), when regressors from both processes were estimated within a single GLM. We report a different strategy to allow a more direct and statistically powerful investigation of this issue under *Divergent learning rates*, below.

Conditional probability

Next we sought regions with reactive rather than anticipatory activity, specifically those where BOLD signal correlated with the probability of the presently viewed image, conditional on the identity of the preceding image. Such activity might reflect the degree of expectation or response preparation, or (in the case of negative correlations) surprise or prediction error, which is decreasing in the predicted probability of the observed image (Gläscher *et al.*, 2010). Note that although the conditional probability was shown to correlate with RT in the behavioral analyses above, RT effects do not confound the fMRI analyses presented here, as all regressors of interest were first orthogonalized against RT.

When regressors were computed from the fast LR process, activation correlating with this time-series was observed in right putamen (18, 14, −4; $P < 0.04$ peak FWE corrected over small volume) (Fig. 4).

In the slow LR process, no activity correlating with conditional probability was observed in our regions of interest, and those clusters above our observation threshold did not survive whole-brain correction.

A complete list of clusters correlating with conditional probability can be found in Supporting Information Table S3.

Divergent learning rates

Our results to this point suggest distinct neural processes operating at different learning rates, but we have not yet provided statistical evidence explicitly supporting the claim that these rates are different. We now quantitatively evaluate the claims that activity in these regions is (i) best explained by different rates and (ii) these rates are uniquely consistent with one of each identified in our behavioral analysis.

Here, again, we consider time-series drawn from a single process learning at a single rate, and pose questions about neural activity, relative to that rate. The dependence of the modeled learning-related activity (conditional probability or entropy time-series) on the learning rate parameter is non-linear. We adopt a linear approximation to this dependence so as to pose statistical questions about the learning rate that would best explain the neural activity in terms of a standard random-effects GLM. Specifically, having generated regressors for our variables of interest according to a baseline learning rate (the midpoint of the behavioral rates), we estimated weights for additional regres-

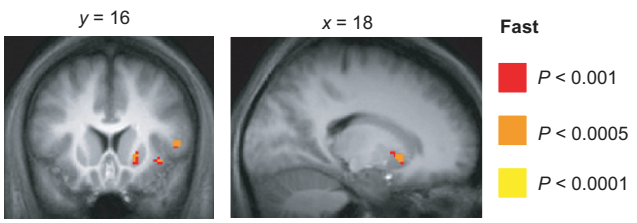

FIG. 4. Areas where the BOLD signal correlated with the conditional probability of the current stimulus, generated at each of our analysed learning rates. Images are thresholded at $P < 0.001$, uncorrected, for display purposes. Using the fast process GLM, significant activation was observed in ventral striatum. No clusters significantly correlated with conditional probability were observed in the slow process GLM.

sors, representing the change in each variable that would result from a change in learning rate – formally, this is the partial derivative, with respect to learning rate, of the variable (Friston *et al.*, 1998). A positive beta value estimated for this regressor in a given voxel would imply that the activity in that voxel is best explained by a time-series generated using a higher learning rate than the one used as a baseline, and a negative beta value would imply a lower value for this best-fit learning rate. For a detailed description of the analysis, see Supporting Information, *Learning rate derivatives*.

To allow comparison between regions, we first identified voxels displaying learning activity in either region. For this, we performed a whole-brain regression using regressors of interest generated from a baseline learning rate parameter set to the midpoint of the fast and slow rate values identified in behavior, so chosen to have a symmetric chance of detecting activity related to either putative process. We then selected the peak active voxels in our two a priori regions of interest for the regressors that elicited significant activity in our previous analyses – forward entropy in hippocampus and conditional probability and forward entropy in ventral striatum – and examined the beta weights estimated for the corresponding derivative regressors. Before testing, these were scaled by the main effect of the regressor of interest to produce an estimate with units of learning rate.

Figure 5 displays the pattern of results. Comparing weights between areas, the activity in hippocampus related to forward entropy was best explained by a smaller learning rate, as assessed by the partial derivative regressor, than that in striatum related either to entropy (paired two-sample, two-tailed *t*-test across subjects, $P = 0.023$) or probability ($P = 0.0251$). These tests support the conclusion that activity in each region is described by incremental learning processes with distinctly different dynamics.

To clarify the unique association of each region with our behaviorally obtained learning rates, we additionally compared the parameter implied by the responses measured in each neural structure to the fast and slow rates identified behaviorally, and also to the average used to select voxels, α_0. Compared with α_0, the implied α_{BOLD} in hippocampus was significantly lower (one-sample, two-tailed *t*-tests across subjects, $P = 0.049$), while activity in striatum implied a value that trended towards being significantly higher than α_0 for entropy ($P = 0.058$), but not for probability ($P = 0.37$). The rate implied by activity in striatum was significantly higher than the slow rate for entropy ($P < 5 \times 10^{-8}$) and probability ($P = 0.007$); neither was significantly different from the fast rate (entropy, $P = 0.53$; probability, $P = 0.86$). Symmetrically, the α_{BOLD} estimated for hippocampus was significantly lower than the fast rate ($P = 0.017$) but not significantly different from the slow rate ($P = 0.56$).

Taken together, these results suggest that learning-related activity in the hippocampus and striatum was, respectively, consistent with the slow and fast LR processes hypothesized on the basis of our behavioral model fits.

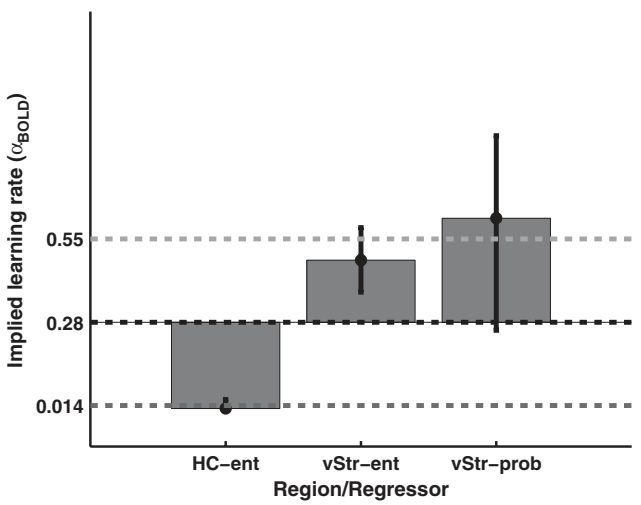

FIG. 5. Comparison of learning rates implied by activity in our primary regions of interest. These values were computed by first identifying voxels in our a priori regions of interest (hippocampus and ventral striatum) which were maximally responsive to our model regressors (probability and entropy) when generated at the midpoint of our behaviorally obtained learning rates (black dotted line), then estimating best-fitting learning rates by deviations from this baseline (see Supporting Information). Bars represent the average implied learning rate across subjects, at a single voxel for each combination of region and regressor: left hippocampus (−26, −14, 24) and right ventral striatum (entropy 18, 16, −6; probability 20, 6, −2). Error bars represent the positive and negative confidence intervals, across subjects.

Discussion

We provide evidence that learned expectations expressed in serial response behavior comprise dissociable contributions from anatomically distinct networks learning at different rates. Effects of serial expectation on RTs (Bahrick, 1954) and BOLD responses (Huettel *et al.*, 2002; Harrison *et al.*, 2006; Schendan *et al.* 2003) are well established; we exploit these effects to study how participants learned expectations trial-by-trial. Thus our approach parallels recent work in reward learning and decision-making (O'Doherty *et al.*, 2003; Barraclough *et al.*, 2004; Samejima *et al.*, 2005; Lohrenz *et al.*, 2007), but we apply these methods to investigate behavior that results from sequential contingency learning, while minimizing the influence of reinforcement. As feedback about correctness was downplayed and the response behavior was well practised and maintained near ceiling, the RT effects and associated neural modulations are likely to reflect fluctuating serial contingency predictions and are not explicable in terms of differential reward expectations. Such behavior appears well described by a weighted combination of two error-driven learning processes.

It is possible that the two-process nature of our results may be rooted in the fact that a serial RT task confounds both response–response and stimulus–stimulus sequencing. For instance, the long-lasting effects of transitions on hippocampal activity might reflect learning there of stimulus–stimulus predictive relations – a key building block for model-based RL (Gläscher *et al.*, 2010) – while the faster decaying effects in striatum might reflect more transitory learning of response–response biases, with both affecting RTs. Directly testing this suggestion would require a different task that separately manipulated these two sorts of contingencies.

Similar distinctions in the timescale over which expectations are drawn have been observed between disparate brain structures processing common information, for example between subcortical and cortical association structures (Gläscher & Buchel, 2005), and between different parts of the sensory cortex (Hassan *et al.*, 2008), and even between different neurons within an area (Bernacchia *et al.*, 2011). However, this dissociation has not previously been drawn between hippocampus and striatum, and in none of these cases were the neural timescales linked to dissociable effects on behavior (see Kable & Glimcher, 2007; for a discussion of the importance of connecting putatively separate neural processes to distinct behavioral influences). In contrast, the most prominent issue addressed using SRT tasks like ours has been the status of learning as explicit or implicit

(e.g. Nissen *et al.*, 1989). In examining this distinction, researchers have explored the proposal that multiple processes underlie learning. A recent model (Keele *et al.*, 2003) proposes that sequential learning involves the parallel operation of two cognitive systems, each constructing distinct representations – a multidimensional representation in one, entrained by the unidimensional representations of the other. Although it is unclear whether this hierarchical arrangement results in different timescales of integration (which would be reflected by different learning rates), the proposal that these systems correspond to ventral and dorsal networks, containing hippocampus and basal ganglia, respectively, is broadly consistent with our results. Even more closely matching our conclusions, Davis & Staddon (1990) advance a dual-system, two-learning-rate architecture to explain pigeons' choices on a reversal learning task, which display both long- and short-timescale dependencies on experience.

Finally, neuropsychological and animal lesion studies have repeatedly observed that removal or damage to the hippocampus does not impair gross sequential RT effects in similar tasks (e.g. Curran, 1997). Our results are consistent with these observations, in that either system's predictions may encourage responses, although our model predicts a quantitative difference in trial-by-trial adjustments in RTs. Crucially, two studies on a serial RT task resembling the one employed in the present experiment observed that lesions to the dorsal hippocampus of rodents actually *improved* performance (Eckart *et al.* 2011) – i.e. producing a steeper learning curve – while striatal lesions severely diminished the rate of learning (Eckart *et al.* 2011); this pattern is consistent with slow and fast contingency learning in hippocampus and striatum, respectively.

Model-based simulation

One interpretation of the hippocampal activity in the present study is that it is driven by retrieval, which is ubiquitous, rather than encoding, which (as we discuss below) may be rare. For instance, if hippocampus retrieves likely subsequent pictures at each step, then more widespread activation would occur on trials with higher forward entropy, i.e. those in which activation spreads more evenly among more potential successors.

The suggestion that our hippocampal BOLD effects reflect preparatory 'prefetching' of the anticipated next elements in the sequence coincides with observations of 'preplay' activity in this structure in rodents (Ferbinteanu & Shapiro, 2003; Diba & Buszáki, 2007; Johnson & Redish, 2007). Such activity has been suggested to support decision-making by evaluating the anticipated consequences of candidate actions, a strategy formalized by model-based reinforcement learning (Doya, 1999; Daw *et al.*, 2005; Niv *et al.*, 2006; Johnson and Redish 2005; Rangel *et al.*, 2008; Balleine *et al.*, 2008; Doll *et al.*, 2009) and also known as constructive episodic simulation (Tulving & Thomson, 1971; Schacter & Addis, 2007). Our results indicate that signals reflecting this activity may be parametrically modulated by a measure of the associative complexity of the trace being constructed or simulated.

The model-based approach contrasts with the model-free algorithms for reinforcement learning prominently associated with striatum and its dopaminergic afferents (Houk *et al.*, 1995; Schultz *et al.*, 1997). Note that, due to the minimal involvement of rewards (or errors) in our task, our striatal results are unlikely to relate to the reward prediction errors posited by these models (Delgado *et al.*, 2000; Knutson *et al.*, 2001; O'Doherty *et al.*, 2003; McClure *et al.*, 2004; Hare *et al.*, 2008), but could relate to other non-reward correlates in striatum (Zink *et al.*, 2006; Wittmann *et al.*, 2008).

Arbitration between multiple systems in learning and control

Lesions in rodents support a dissociation between decision strategies along the lines of model-based vs model-free learning, supported by distinct networks neurally (Balleine & Dickinson, 1998; Gerlai, 1998; Corbit & Balleine, 2000; Balleine *et al.*, 2008). However, the neural basis of the dissociation is, as yet, far clear in humans, and some work even seems to suggest overlapping substrates (Valentin *et al.*, 2007; Frank *et al.*, 2009; Gläscher *et al.*, 2010; Tricomi *et al.* 2010; Simon & Daw, 2011; Daw *et al.*, 2011), perhaps because model-based and model-free evaluations of reward expectancy (and thus their anticipated BOLD correlates) are typically quite similar. To the extent that our two processes do indeed map differentially to stimulus–stimulus and response–response associations, our study suggests an additional difference between the systems, in their timescale of learning. This difference may allow their predictions in a reinforcement learning context to be more easily distinguished. Thus, a potentially fruitful avenue for further research is to use the tools provided herein to identify the use of either system's learned predictions in the service of reward-guided decision-making.

The nature of competition or collaboration between these systems in the control of behavior has been a topic of much empirical (Poldrack *et al.*, 2001) and theoretical (Daw *et al.*, 2005) inquiry. Our results suggest that activity in both hippocampus and striatum is mediated by the uncertainty (i.e. entropy) about anticipated ensuing stimuli, and that this activity may differently drive fluctuating signals in each area – positively in hippocampus, negatively in striatum. Such activity may reflect differential engagement of either system under different conditions of uncertainty (Daw *et al.*, 2005). In a traditional response or choice task with fixed contingencies, the overall trend would be towards sharper expectations (lower uncertainty) over time, giving rise to a decrease in hippocampal and commensurate increase in striatal activity. Indeed, such a pattern bears strong similarity to that repeatedly observed in probabilistic association learning tasks (Poldrack *et al.*, 2001; Poldrack & Packard, 2003).

Learning rates and associative representations

So far, we have stressed differences in the timescales over which neural activity reflects past events. However, in the error-driven learning model (Eqn 1), as indeed in estimation more generally, encoding is also forgetting. Thus, a long-timescale dependence of learned predictions on events (slow decay, high $1–\alpha$) goes hand in hand with slow, incremental encoding (a small learning rate α). Viewed from this perspective, perhaps the most surprising aspect of our data is that we measure faster learning rates in striatum than hippocampus, given the traditional association of hippocampus with fast, single-shot episodic encoding (McClelland *et al.*, 1995), and striatum with more incremental procedural learning (Knowlton *et al.*, 1996a).

There are at least two possible answers to this question. One is that a fundamental hippocampal function in learning relations (Cohen & Eichenbaum, 1993) comprises not only relating events occurring simultaneously in an episode – a fast-timescale encoding – but also discovering event relations obtained stochastically between temporally separated events (Shohamy *et al.*, 2009; Hales & Brewer, 2010), as with state transition contingencies in model-based RL. If so, different tasks or task variants might induce learning over different timescales, depending on the relations involved (Komorowski *et al.*, 2009). In particular, learning a probabilistic transition structure, such as imputing the equivalence relationships in Shohamy *et al.* (2009) acquired equivalence task, requires integrating events across time rather than within an episode.

A second interpretation of the hippocampal result, suggested by the episodic learning literature, is that the slow learning rate we measure for representations in hippocampus reflects not continual, incremental updating, but instead the average rate of learning over trials in which associations might be formed rapidly but only sporadically. In particular, the Rescorla–Wagner equation also describes the expected update for the predictions under sporadic encoding (e.g. on a given trial, the probability α of fully re-encoding a categorical stimulus–stimulus link, with no learning otherwise), or more generally some in-between process where α modulates over some fraction of trials. For simplicity and to enable comparing many sorts of learning in a single framework, we operationalized the distinction between the systems in terms of a nominally constant learning rate. If updates are sporadic or step sizes are time-varying, the fit parameter value will instead characterize their average rate (Behrens *et al.*, 2007). Our learning rule thus comprises (at least in expectation) a spectrum of possibilities between incremental learning of a probabilistic relation and sporadic encoding of a categorical relation, the latter similar to discrete state space models that have previously been applied to hippocampal function (Law *et al.*, 2005; Prerau *et al.*, 2008; Wirth *et al.*, 2009).

This hypothesis that our task produces sporadic hippocampal encoding is supported by work suggesting that hippocampus forms associations preferentially at the detection of sufficiently large deviations from expected input (Tulving *et al.*, 1996; Lisman & Grace, 2005; Bakker *et al.*, 2008), or when task demands enhance the motivational salience of expectancy violations (Duncan *et al.*, 2009).

This interpretation of the current result is difficult to test directly using the present data set and methods, particularly because it is technically challenging to test the fit of such a model without a specific hypothesis about which trials may have encouraged encoding or not. A more direct approach would be to manipulate factors expected to impact the tendency to form new episodes (Behrens *et al.*, 2007; Bakker *et al.*, 2008; Duncan *et al.*, 2009; Nassar *et al.*, 2010; Wilson *et al.*, 2010), and seek an effect on the measured hippocampal learning rate. For instance, in environments with largely stationary associative structure, as in our task, learning may appear 'slow' on average. However, tasks with more frequent changes may produce a correspondingly larger value of the hippocampal learning rate. This prediction parallels previous work demonstrating that humans (Behrens *et al.*, 2007; Speekenbrink & Shanks, 2010) and animals (Dayan *et al.*, 2000; Courville *et al.*, 2006; Preuschoff & Bossaerts, 2007) modulate their learning rates in response to the volatility of changes in associational structure. The exact response of each individual system to environmental volatility is a potentially fruitful avenue for future research.

Representations in the model-based system

The question of whether sequential predictions are categorical or graded bears directly on how they would support decisions. In computational RL, state–state world models are probabilistic (to support exact computations of expected future value in stochastic Markov decision tasks) but more psychological accounts of deliberative processing, to which model-based RL might correspond, often take their representations to be rule-based or categorical, in contrast to more graded representations learned in an incremental fashion by (model-free) procedural systems (Packard & Knowlton, 2002; Maddox & Ashby, 2004). A similar binary view is taken in 'state space' models (Lau & Glimcher, 2005) that have been applied to hippocampal associative representations.

However, the above considerations notwithstanding, the extreme case of a process in which hippocampus learns only categorical predictive associations seems unlikely to explain our observations, because neural activity here is seen to relate to forward entropy (see also Strange *et al.*, 2005; Harrison *et al.*, 2006). Although the Rescorla–Wagner rule describes the *expected* time-course for the stimulus predictions even under sporadic encoding, the time-course of their entropy in this case is not the same as the entropy of the expected predictions, as the entropy is a non-linear function of the predictions. More concretely, if the neural representation over the next stimulus were always fully categorical (i.e. as a probability distribution, deterministic, although undergoing sporadic stepwise changes) then the implied entropy would be always minimal and never modulate, unlike the hippocampal signal we detect. It is possible that our observations could be produced by a process of learning graded predictions via sporadic yet still incremental changes, at a learning rate that is a multiple of that measured here.

Either way, our data invite an interpretation in which hippocampal representations reflect the statistics of the environment in graded form, a view more conducive to model-based RL and also consistent with research on its involvement in learning of statistical task structure (Gluck and Myers 1993; Courville *et al.* 2004; Gershman and Niv 2009).

Supporting Information

Additional supporting information can be found in the online version of this article:

Data S1. Methods.

Fig. S1. Non-orthogonalized statistical parametric mapping.

Fig. S2. Combined process statistical parametric mapping.

Table S1. Areas of *negative* correlation with the forward entropy regressor in the fast process GLM

Table S2. Areas of correlation with the forward entropy regressor in the slow process GLM

Table S3. Areas of correlation with the conditional probability regressor in the fast process GLM

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset by Wiley-Blackwell. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Acknowledgements

We are grateful to Todd Gureckis, David Heeger, Rich Ivry, Michael Landy, Brian McElree, Wendy Suzuki, Daphna Shohamy and Klaas Stephan for helpful conversations, and Samuel Gershman, Jian Li and Dylan Simon for valuable technical assistance. This work was funded by a McKnight Scholar Award, NIMH grant 1R01MH087882-01, part of the CRCNS program, Human Frontiers Science Program Grant RGP0036/2009-C, and a NARSAD Young Investigator Award.

Abbreviations

BIC, Bayesian information criterion; BOLD, blood oxygen-level dependent; fMRI, functional magnetic resonance imaging; FWE, family-wise error; GLM, generalized linear model; RL, reinforcement learning; RT, reaction time; SD, standard deviation; SRT, serial reaction time.

References

Bahrick, H.P. (1954) Incidental learning under two incentive conditions. *J. Exp. Psychol.*, **47**, 170–172.

Bakker, A., Kirwan, C.B., Miller, M. & Stark, C.E.L. (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, **319**, 1640–1642.

Balleine, B.W. & Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, **37**, 407–419.
Balleine, B.W., Daw, N.D. & O'Doherty, J.P. (2008). Multiple forms of value learning and the function of dopamine. In Glimcher, P.W., Camerer, C., Poldrack, R.A. & Fehr, E. (Eds), *Neuroeconomics: Decision Making and the Brain*. Academic Press, New York, NY, pp. 367–387.
Barraclough, D.J., Conroy, M.L. & Lee, D. (2004) Prefrontal cortex and decision-making in a mixed-strategy game. *Nat. Neurosci.*, **7**, 404–410.
Bayer, H.M. & Glimcher, P.W. (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, **47**, 129–141.
Behrens, T.E., Woolrich, M.W., Walton, M.E. & Rushworth, M.F. (2007) Learning the value of information in an uncertain world. *Nat. Neurosci.*, **10**, 1214–1221.
Bernacchia, A., Seo, H., Lee, D. & Wang, X.-J. (2011) A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci.*, **14**, 366–372.
Bestmann, S., Harrison, L.M., Blankenburg, F., Mars, R.B., Haggard, P., Friston, K.J. & Rothwell, J.C. (2008) Influence of contextual uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr. Biol.*, **18**, 775–780.
Bornstein, A.M. & Daw, N.D. (2011) Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Curr. Opin. Neurobiol.*, **21**, 374–380.
Brainard, D.H. (1997) The Psychophysics Toolbox. *Spat. Vis.*, **10**, 433–436.
Cohen, N.J. & Eichenbaum, H. (1993) *Memory, Amnesia, and the Hippocampal System*. MIT Press, Cambridge.
Corbit, L.H. & Balleine, B.W. (2000) The role of the hippocampus in instrumental conditioning. *J. Neurosci.*, **20**, 4233.
Corrado, G.S., Sugrue, L.P., Sebastian Seung, H. & Newsome, W.T. (2005) Linear-nonlinear-poisson models of primate choice dynamics. *J. Exp. Anal. Behav.*, **84**, 581–617.
Courville, A.C., Daw, N., Gordon, G.J. & Touretzky, D.S. (2004) Model uncertainty in classical conditioning. In Thrun, S., Saul, L. & Schölkopf, B. (Eds), *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, pp. 977–984.
Courville, A.C., Daw, N.D. & Touretzky, D.S. (2006) Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.*, **10**, 294–300.
Curran, T. (1997) Higher-order associative learning in amnesia: evidence from the serial reaction time task. *J. Cogn. Neurosci.*, **9**, 522–533.
Davis, D.G. & Staddon, J.E.R. (1990) Memory for reward in probabilistic choice. *Behaviour*, **114**, 1–4.
Daw, N.D. (2010) Trial-by-trial data analysis using computational models. In Phelps, E.A., Robbins, T.W. & Delgado, M. (Eds), *Affect, Learning and Decision Making, Attention and Performance XXIII*. Oxford University Press, Oxford.
Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, **8**, 1704–1711.
Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B. & Dolan, R.J. (2006) Cortical substrates for exploratory decisions in humans. *Nature*, **441**, 876–879.
Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. (2011) Model-based influences on human's choices and striatal prediction errors. *Neuron*, **69**, 1204–1215.
Dayan, P., Kakade, S. & Montague, P.R. (2000) Learning and selective attention. *Nat. Neurosci.*, **3**, 1218–1223.
Delgado, M.R., Nystrom, L., Fissel, C., Noll, D. & Fiez, J. (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.*, **84**, 3072–3077.
Diba, K. & Buszáki, G. (2007) Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.*, **10**, 1241–1242.
Dickinson, A. & Balleine, B.W. (2002) The role of learning in the operation of motivational systems. In Pashler, H. & Gallistel, R. (Eds), *Stevens Handbook of Experimental Psychology Vol. 3: Learning, Motivation and Emotion*. John Wiley & Sons, New York, pp. 497–533.
Doeller, C.F., King, J.A. & Burgess, N. (2008) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl Acad. Sci. USA*, **105**, 5915–5920.
Doll, B.B., Jacobs, W.J., Sanfey, A.G. & Frank, M.J. (2009) Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Res.*, **1299**, 74–94.
Doya, K. (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw.*, **12**, 961–974.
Drevets, W.C., Gautier, C., Price, J.C., Kupfer, D.J., Kinahan, P.E., Grace, A.A., Price, J.L. & Mathis, C.A. (2001) Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biol. Psychiatry*, **49**, 81–89.
Duncan, K., Curtis, C. & Davachi, L. (2009) Distinct memory signatures in the hippocampus: Intentional States distinguish match and mismatch enhancement signals. *J. Neurosci.*, **29**, 131.
Eckart, M., Huelse-Matia, M. & Schwarting, R. (2011) Dorsal hippocampal lesions boost performance in the rat sequential reaction time task. *Hippocampus*, doi: 10.1002/hipo.20965.
Ferbinteanu, J. & Shapiro, M.L. (2003) Prospective and retrospective memory coding in the hippocampus. *Neuron*, **40**, 1227–1239.
Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J. & Doya, K. (2010) Evidence for model-based action planning in a sequential finger movement task. *J. Mot. Behav.*, **42**, 371–379.
Frank, M.J., Doll, B.B., Oas-Terpstra, J. & Moreno, F. (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat. Neurosci.*, **12**, 1062–1068.
Friston, K.J., Josephs, O., Rees, G. & Turner, R. (1998) Nonlinear event-related responses in fMRI. *Magn. Reson. Med.*, **39**, 41–52.
Gerlai, R. (1998) A new continuous alternation task in T-maze detects hippocampal dysfunction in mice: a strain comparison and lesion study. *Behav. Brain Res.*, **95**, 91–101.
Gershman, S.J., Pesaran, B. & Daw, N.D. (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *J. Neurosci.*, **29**, 13524–13531.
Gershman, S.J. & Niv, Y. (2010) Learning latent structure: carving nature at its joints. *Neurobiology*, **20**, 1–6.
Gläscher, J. & Buchel, C. (2005) Formal learning theory dissociates brain regions with different temporal integration. *Neuron*, **47**, 295–306.
Gläscher, J. & O'Doherty, J.P. (2010) Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdiscip. Rev. Cogn. Sci.*, **1**, 501–510.
Gläscher, J., Daw, N.D., Dayan, P. & O'Doherty, J.P. (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, **66**, 585–595.
Gluck, M.A. & Myers, C.E. (1993) Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus*, **3**, 491–516.
Hales, J. & Brewer, J. (2010) Activity in the hippocampus and neocortical working memory regions predicts successful associative memory for temporally-discontiguous events. *Neuropsychologia*, **48**, 3351–3359.
Hampton, A.N., Bossaerts, P. & O'Doherty, J.P. (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, **26**, 8360–8367.
Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W. & Rangel, A. (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.*, **28**, 5623–5630.
Harrison, L.M., Duggins, A. & Friston, K.J. (2006) Encoding uncertainty in the hippocampus. *Neural Netw.*, **19**, 535–546.
Hassan, U., Yang, E., Vallines, I., Heeger, D.J. & Rubin, N. (2008) A heirarchy of temporal receptive windows in human cortex. *J. Neurosci.*, **28**, 2539–2550.
Henson, R. (2005) A mini-review of fMRI studies of human medial temporal lobe activity associated with recognition memory. *Q. J. Exp. Psychol.*, **58**, 340–360.
Holmes, A. & Friston, K. (1998) Generalisability, random effects & population inference. *Neuroimage*, **7**, S754.
Houk, J., Adams, J. & Barto, A. (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Houk, J.C., Davis, J.L. & Beiser, D.G. (Eds), *Models of Information Processing in the Basal Ganglia*. MIT Press, Cambridge, MA, pp. 249–270.
Huettel, S.A., Mack, P.B. & McCarthy, G. (2002) Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.*, **5**, 485–490.
Johnson, A. & Redish, A.D. (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.*, **27**, 12176.
Johnson, A. & Redish, A.D. (2005) Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Netw.*, **18**, 1163–1171.
Kable, J.W. & Glimcher, P.W. (2007) The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, **10**, 1625–1633.
Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *J. Am. Stat. Assoc.*, **90**, 773–795.
Keele, S.W., Ivry, R., Mayr, U., Hazeltine, E. & Heuer, H. (2003) The cognitive and neural architecture of sequence representation. *Psychol. Rev.*, **110**, 316–339.
Knowlton, B.J., Squire, L.R. & Gluck, M.A. (1994) Probabilistic classification learning in amnesia. *Learn. Mem.*, **1**, 106–120.

Knowlton, B.J., Mangels, J.A. & Squire, L.R. (1996a) A neostriatal habit learning system in humans. *Science*, **273**, 1399–1402.
Knowlton, B.J., Squire, L.R., Paulsen, J.S., Swerdlow, N.R., Swenson, M. & Butters, N. (1996b) Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology*, **10**, 538–548.
Knutson, B., Adams, C.M., Fong, G.W. & Hommer, D. (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.*, **159**, 1–5.
Komorowski, R.W., Manns, J.R. & Eichenbaum, H. (2009) Robust conjunctive item-place coding by hippocampal neurons parallels learning what happens where. *J. Neurosci.*, **29**, 9918–9929.
Lau, B. & Glimcher, P.W. (2005) Dynamic response-by-response models of matching behavior in Rhesus monkeys. *J. Exp. Anal. Behav.*, **84**, 555–578.
Law, J.R., Flanery, M.A., Wirth, S., Yanike, M., Smith, A.C., Frank, L.M., Suzuki, W.A., Brown, E.N. & Stark, C.E.L. (2005) Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *J. Neurosci.*, **25**, 5720–5729.
Lisman, J.E. & Grace, A.A. (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, **46**, 703–713.
Lohrenz, T., Mccabe, K., Camerer, C.F. & Montague, P.R. (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl Acad. Sci. USA*, **104**, 9493–9498.
Maddox, W.T. & Ashby, F.G. (2004) Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behav. Processes*, **66**, 309–332.
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. *Psychol. Rev.*, **102**, 419–457.
McClure, S.M., Laibson, D.I., Loewenstein, G. & Cohen, J.D. (2004) Separate neural systems value immediate and delayed monetary rewards. *Science*, **306**, 503.
McDannald, M.A., Lucantonio, F., Burke, K.A., Niv, Y. & Schoenbaum, G. (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.*, **31**, 2700–2705.
Montague, P.R., Dayan, P. & Sejnowski, T.J. (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.*, **16**, 1936–1947.
Nassar, M.R., Wilson, R.C., Heasly, B. & Gold, J.I. (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.*, **30**, 12366–12378.
Nissen, M., Willingham, D. & Hartman, M. (1989) Explicit and implicit remembering: when is learning preserved in amnesia? *Neuropsychologia*, **27**, 341–352.
Niv, Y., Joel, D. & Dayan, P. (2006) A normative perspective on motivation. *Trends Cogn. Sci.*, **10**, 375–381.
Nomura, E., Maddox, W., Filoteo, J.A.D., Gitelman, D.R., Parrish, T.B., Mesulam, M.M. & Reber, P.J. (2007) Neural correlates of rule-based and information-integration visual category learning. *Cereb. Cortex*, **17**, 37–43.
O'Doherty, J.P., Dayan, P., Friston, K.J., Critchley, H. & Dolan, R.J. (2003) Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**, 329–337.
O'Doherty, J.P., Hampton, A. & Kim, H. (2007) Model-based fMRI and its application to reward learning and decision-making. *Ann. N Y Acad. Sci.*, **1104**, 35–53.
den Ouden, H.E.M., Daunizeau, J., Roiser, J., Friston, K.J. & Stephan, K.E. (2010) Striatal prediction error modulates cortical coupling. *J. Neurosci.*, **30**, 3210–3219.
Packard, M.G., Hirsh, R. & White, M. (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. *J. Neurosci.*, **9**, 1465.
Packard, M.G. & Knowlton, B.J. (2002) Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.*, **25**, 563–593.
Poldrack, R.A. & Packard, M.G. (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, **41**, 245–251.
Poldrack, R.A., Clark, J., Pare-Blagoev, E.J., Shohamy, D., Moyano, J.C., Myers, C. & Gluck, M.A. (2001) Interactive memory systems in the human brain. *Nature*, **414**, 546–550.
Prerau, M.J., Smith, A.C., Yanike, M., Suzuki, W.A. & Brown, E.N. (2008) A mixed filter algorithm for simultaneously recorded continuous-valued and binary observations. *Biol. Cybern.*, **99**, 1–14.
Preuschoff, K. & Bossaerts, P. (2007) Adding prediction risk to the theory of reward learning. *Ann. N Y Acad. Sci.*, **1104**, 135–146.
Rangel, A., Camerer, C. & Montague, P.R. (2008) A framework for studying the neurobiology of value-based decision-making. *Nat. Rev. Neurosci.*, **9**, 545–556.
Redish, A.D., Jensen, S. & Johnson, A. (2008) A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.*, **31**, 415–437.
Rescorla, R.A. & Wagner, A.R. (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A.H. & Prokasy, W.F. (Eds), *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, New York, NY, pp. 64–99.
Rubin, D.C. & Wenzel, A.E. (1996) One hundred years of forgetting: a quantitative description of retention. *Psychol. Rev.*, **103**, 734–760.
Rubin, D.C., Hinton, S. & Wenzel, A. (1999) The precise time course of retention. *J. Exp. Psychol. Learn. Mem. Cogn.*, **25**, 1161–1176.
Samejima, K., Ueda, Y., Doya, K. & Kimura, M. (2005) Representation of action-specific reward values in the striatum. *Science*, **310**, 1337–1340.
Schacter, D.L. & Addis, D.R. (2007) The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc.*, **362**, 773.
Schendan, H.E., Searl, M.M., Melrose, R.J. & Stern, C.E. (2003) An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, **37**, 1013–1025.
Schoenberg, T., Daw, N.D., Joel, D. & O'Doherty, J.P. (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision-making. *J. Neurosci.*, **27**, 12860.
Schoenberg, T., O'Doherty, J.P., Joel, D., Inzelberg, R., Segev, Y. & Daw, N.D. (2010) Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. *Neuroimage*, **49**, 772–781.
Schultz, W., Dayan, P. & Montague, P.R. (1997) A neural substrate of prediction and reward. *Science*, **275**, 1593.
Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.
Shohamy, D., Myers, C.E., Hopkins, R.O., Sage, J. & Gluck, M.A. (2009) Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *J. Cogn. Neurosci.*, **21**, 1820–1832.
Simon, D.A. & Daw, N.D. (2011) Neural correlates of forward planning in a spatial decision task in humans. *J. Neurosci.*, **31**, 5526–5539.
Speekenbrink, M. & Shanks, D.R. (2010) Learning in a changing environment. *J. Exp. Psychol. Gen.*, **139**, 266–298.
Strange, B.A., Duggins, A., Penny, W., Dolan, R.J. & Friston, K.J. (2005) Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw.*, **18**, 225–230.
Tanaka, S.C., Doya, K., Okada, G., Ueda, K., Okamoto, Y. & Yamawaki, S. (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.*, **7**, 887–893.
Tricomi, E., Balleine, B.W. & O'Doherty, J.P. (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.*, **29**, 2225–2232.
Tulving, E. & Thomson, D.M. (1971) Retrieval processes in recognition memory: effects of associative context. *J. Exp. Psychol.*, **87**, 116–124.
Tulving, E., Markowitsch, H.J., Craik, F.I.M., Habib, R., Houle, S., York, N. *et al.* (1996) Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cereb. Cortex*, **6**, 71–79.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. & Joliot, M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**, 273–289.
Valentin, V.V., Dickinson, A. & O'Doherty, J.P. (2007) Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.*, **27**, 4019–4026.
Wilson, R.C., Nassar, M.R. & Gold, J.I. (2010) Bayesian online learning of the hazard rate in change-point problems. *Neural Comput.*, **22**, 2452–2476.
Wirth, S., Avsar, E., Chiu, C.C., Sharma, V., Smith, A.C., Brown, E., Suzuki, W. (2009) Trial outcome and associative learning signals in the monkey hippocampus. *Neuron*, **61**, 930–940.
Wittmann, B.C., Daw, N.D., Seymour, B. & Dolan, R.J. (2008) Striatal activity underlies novelty-based choice in humans. *Neuron*, **58**, 967–973.
Zink, C.F., Pagnoni, G., Chappelow, J., Martin-Skurski, M. & Berns, G.S. (2006) Human striatal activation reflects degree of stimulus saliency. *NeuroImage*, **29**, 977–983.