Hippocampal Pattern Separation Supports Reinforcement Learning

Ian Ballard[1], Anthony D. Wagner[2], Samuel M. McClure[3]

1. Stanford Neurosciences Graduate Training Program, Stanford University. Stanford, CA 94305, USA
2. Department of Psychology, Stanford University, Stanford, CA 94305, USA.
3. Department of Psychology, Arizona State University, Tempe, AZ 85287, USA.

Corresponding Author
Ian Ballard
Department of Psychology
450 Serra Mall, Stanford CA 94305
iancballard@gmail.com

Conflicts of Interest:
The authors declare no competing financial interests.

Keywords: nonlinear discrimination, fMRI, striatum, prediction error, ambiguous feature, complementary learning systems, representational similarity analysis

1 *ABSTRACT*

Animals rely on learned associations to make decisions. Associations can be based on relationships between object features (e.g., the three-leaflets of poison ivy leaves) and outcomes (e.g., rash). More often, outcomes are linked to multidimensional states (e.g., poison ivy is green in summer but red in spring). Feature-based reinforcement learning fails when the values of individual features depend on the other features present. One solution is to assign value to multifeatural conjunctive representations. We tested if the hippocampus formed separable conjunctive representations that enabled learning of response contingencies for stimuli of the form: AB+, B-, AC-, C+. Pattern analyses on functional MRI data showed the hippocampus formed conjunctive representations that were dissociable from feature components and that these representations influenced striatal PEs. Our results establish a novel role for hippocampal pattern separation and conjunctive representation in reinforcement learning.

2 ACKNOWLEDGEMENTS

We thank the NSF GRFP (ICB), NSF IGERT (ICB) and Stanford Innovation Grants (ICB). We also thank Kim D'Ardenne for significant editing and Stephanie Gagnon, Karen LaRocque and Yuan Chang Leong for their feedback.

3 *INTRODUCTION*

Most North American hikers develop a reflexive aversion to poison ivy, which causes a painful rash, and learn to recognize its compound leaf with three leaflets that is green in summer and red in spring and autumn. The relationship between color and season distinguishes poison ivy from other plants like boxelder, which looks similar but is green in spring. Such learning problems are challenging because similar conjunctions of features can require different responses or elicit different predictions about future events. Responses and predictions also depend on the status of other features or context. In such problems, simple feature-response learning is insufficient and representations that include multiple features (leaf shape, color, season) must be learned.

Learning systems in the brain encode qualitatively distinct representations, and theories posit that reinforcement learning operates over multiple types of representations (1). Theoretical and empirical work suggest the hippocampus rapidly forms conjunctive representations of arbitrary sets of co-occurring features (2), making the hippocampus critical for episodic memory (3). During encoding of conjunctive representations, hippocampal computations likely establish minimal representational overlap between traces of events with partially shared features, so-called pattern separation (4, 5), which reduces interference between experiences with overlapping features. One solution to multifeatural learning problems that require stimuli with overlapping features to be associated with different outcomes is to encode neurally separable conjunctive representations, putatively through hippocampal-dependent computations, and to assign value to each "separated" representation, putatively through hippocampal-striatal interactions. The same circuit and computational properties that make the hippocampus vital for episodic memory can also benefit striatal-dependent reinforcement learning by providing separated conjunctive representations over which value learning can occur.

Stimulus-response learning occurs by the incremental adjustment of synapses on striatal neurons (6). Thalamic and sensory cortical inputs encode single stimuli, such as a light, and are strengthened in response to phasic dopamine reward prediction errors (PEs; 7-10). This system allows for incremental learning about individual feature values. Although the hippocampus is not critical for associating value with individual features or items (11), it provides dense input to the striatum (12). These synapses are strengthened by phasic dopamine release via D1 receptors (13)

and might represent conjunctions of features distributed in space or time (6). We used a non-spatial, probabilistic stimulus-response learning task including stimuli with overlapping features to test the role of the hippocampus and its interaction with the striatum in value learning over conjunctive codes. We hypothesized hippocampal pattern separation computations and hippocampal-to-striatal projections would form a conjunctive-value learning system that worked in tandem with a feature-value learning system implemented in sensory cortical-to-striatal projections.

We compared hippocampal representational codes to those of three other cortical areas that could contribute to learning in our task: perirhinal (PRc) and parahippocampal (PHc) cortices and inferior frontal sulcus (IFS). The PRc and PHc gradually learn representations of individual items (14, 15). Cortical learning is generally too slow to form representations linking multiple items (2), and pattern separation likely depends on hippocampal computations (5). We therefore predicted PRc and PHc would not form pattern-separated representations of conjunctions with overlapping features. The IFS supports the representation of abstract rules (16, 17) that often describe conjunctive relationships (e.g., "respond to stimuli with both features A and B", (18), but because our task biased subjects away from rule-based learning we predicted the IFS would not form pattern-separated representations of conjunctions. We designed our task and analyses to test for a hippocampal role in encoding conjunctive representations that serve as inputs for striatal associative learning.

4 RESULTS

4.1 Behavioral Results

Subjects learned stimulus-outcome relationships that required the formation of conjunctive representations. Our task was based on the "simultaneous feature discrimination" rodent behavioral paradigm (19). Task stimuli consisted of four feature configurations (AB, AC, B, C). We used a speeded reaction time (RT) task in which a target "go" stimulus was differentially predicted by the four stimuli (Figure 1a). AB and C predicted the target 70% of the time and B and AC 30% of the time. To earn money, subjects pressed a button within a limited response window after target onset. Each feature was associated with the target 50% of the time, but stimuli were more (70%) or less (30%) predictive of the target. Optimal performance

required learning the value of stimuli as distinct conjunctions of features (i.e., *conjunctive representations*).

We tested whether subjects learned stimulus-outcome relationships with four computational models:

1) No Learning Model: subjects ignored predictive information and responded as fast as possible after target.
2) Feature RL: subjects learned values for individual features but not conjunctions. For multifeatural cues, value was updated for each feature.
3) Conjunctive RL: subjects learned values for each distinct stimulus. Value was updated for one representation on each trial (for "AB", value updated for AB but not A or B).
4) Hybrid RL: subjects learned values of stimuli but confused stimuli that shared common features (e.g., AB and B). This model spreads value updates between stimuli that shared features (for AB trial, some of value update was applied to B).

Stimuli that are highly predictive of targets should be associated with faster responses, permitting us to fit each model to the RT data. We first compared the Conjunctive model, which implements the experimenter-defined optimal task strategy, with the Feature and No Learning models. The Feature model uses a simpler and commonly used learning strategy (20). We assessed model fits using a cross-validated predictive likelihood method. The Conjunctive model outperformed the No Learning model ($T = 136$, $p = .028$, Wilcoxon test, Figure 1c) but was only marginally better than the Feature Model ($T = 158$, $p = .08$, Figure 1c). We next assessed the relative fits of these three models with a random-effects Bayesian procedure that gives probabilities that each model would generate the data of a random subject (21). We found the most likely model was the Conjunctive model (protected exceedance probabilities (pEP): Conjunctive 92.3%, Feature 3.9%, No Learning 3.7%), which suggests subjects learned predictive relationships. Overall, there was mixed evidence in support of learning about conjunctions.

We reasoned that these results could be explained by subjects forming conjunctive representations but also confusing stimuli with overlapping features. This behavior could arise if hippocampal pattern separation was partially effective in encoding distinct representations for each stimulus (5) and/or stimulus representations in the hippocampus and feature representations in cortex were simultaneously reinforced during learning (22). We fit a Hybrid model that

allowed for value updates to spread between stimuli with overlapping features. A parameter ω specifies the degree to which value updates spread to other stimuli with shared features. This model outperformed both the Conjunctive ($T = 124$, $p = .015$) and Feature models ($T = 115$, $p = .009$; Wilcoxon tests, Figure 1c) on the cross-validation analysis. Bayesian model comparison confirmed the Hybrid model was the most likely model (pEP: 89.9%, Figure 1b). Fits of spread parameter ω (mean: .44, SD: .25, Table S1) indicated that for any given value update to the current stimulus (e.g., AB), about half that update was also applied to overlapping stimuli (e.g., B). These results support subjects forming conjunctive representations of multifeatural stimuli but blending value learning across stimuli with shared features.

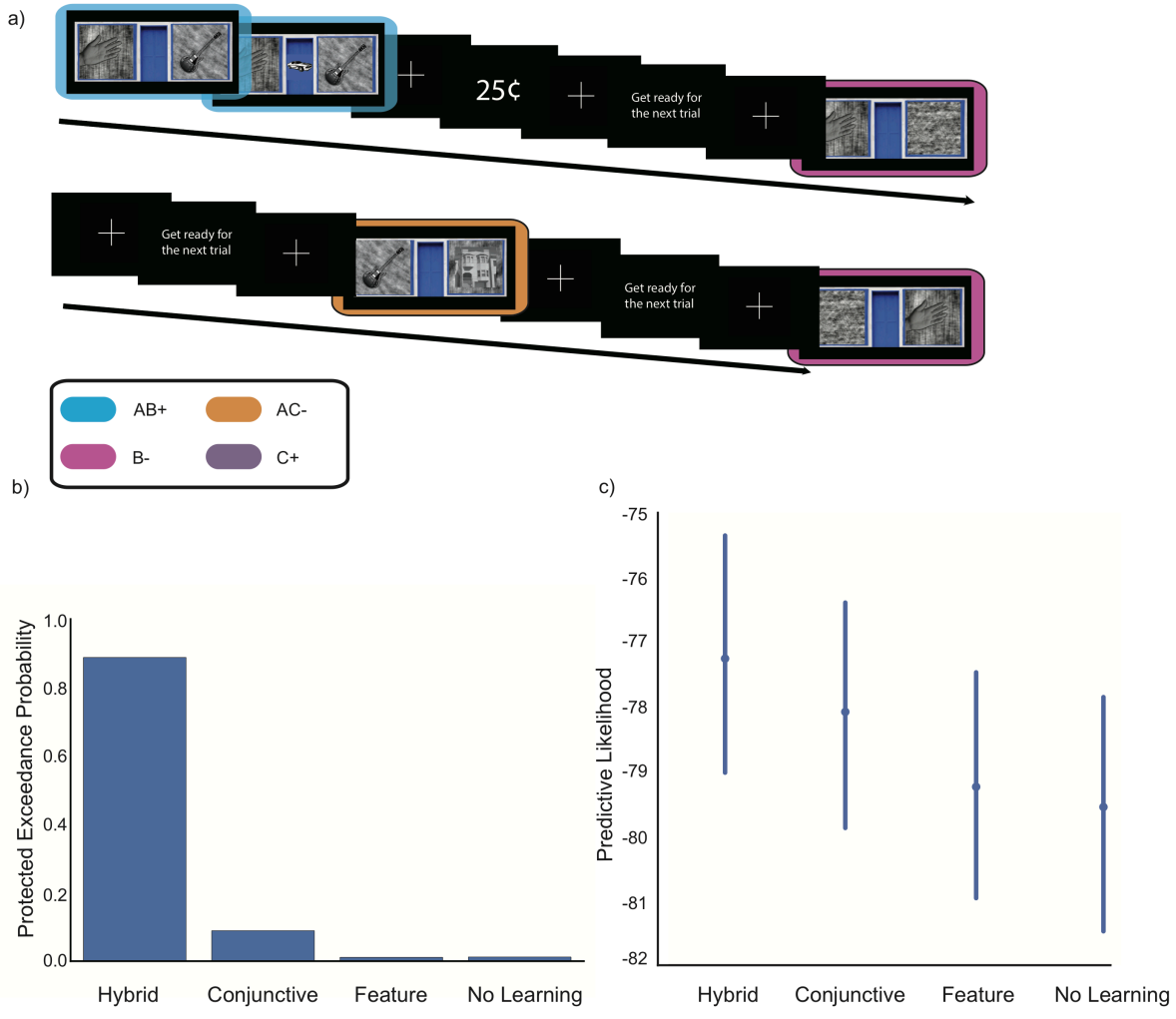

Figure 1. Task design and behavior.

a) AB+, B- and AC- trials. The target appeared at fixation 600 ms after stimuli onset. Stimuli were always presented for 2000 ms. Feedback indicated whether subjects responded quickly enough to earn a reward. Note that stimuli were faces, places and houses. The face image has been replace with a guitar to conform with BioArxiv requirements.
b) Bayesian random effects model comparison showed the Hybrid RL model most likely accounted for behavior. Protected exceedance probabilities sum to 1 across models and because they express a random-effects measure, there are no error bars.
c) Cross-validation model comparison showed the Hybrid RL model best predicted unseen data. Log predictive likelihoods closer to 0 indicate better performance.

4.2 Striatal Prediction Error Analysis

Our behavioral analyses suggested subjects used a reinforcement learning strategy to acquire stimulus-outcome relationships, with learning best described by a model that spread value updates among stimuli sharing features. Because striatal BOLD responses track reward PEs (23), we predicted these BOLD responses would co-vary with PEs derived from the Hybrid model. The key feature of this model is the spread of learning between stimuli sharing features, suggesting that subjects learn jointly about conjunctions and features. We sought to distinguish the contribution of Conjunctive learning, which learns independently about each stimulus, from that of Feature learning, which causes learning to spread across stimuli that share features. A feature PE regressor was constructed from the Feature RL model. A conjunctive PE regressor was constructed by computing the difference between the feature regressor and PEs computed from the Conjunctive RL model. This regressor captures unique variance associated with PEs derived from a model that learns values for conjunctions. We found robust Feature PE responses in the bilateral medial caudate (whole-brain corrected threshold $p < .05$; Figure 2a). We next extracted parameter estimates from an anatomical striatal mask (24) crossed with a statistically-independent functional mask of Feature PE activation and observed these same voxels also showed evidence of a Conjunction PE response ($t(31) = 4.1$; $p < .001$; $d_z = 0.72$; Figure 2b), confirming striatal BOLD tracked reinforcement learning PEs that mixed learning about conjunctions and features.

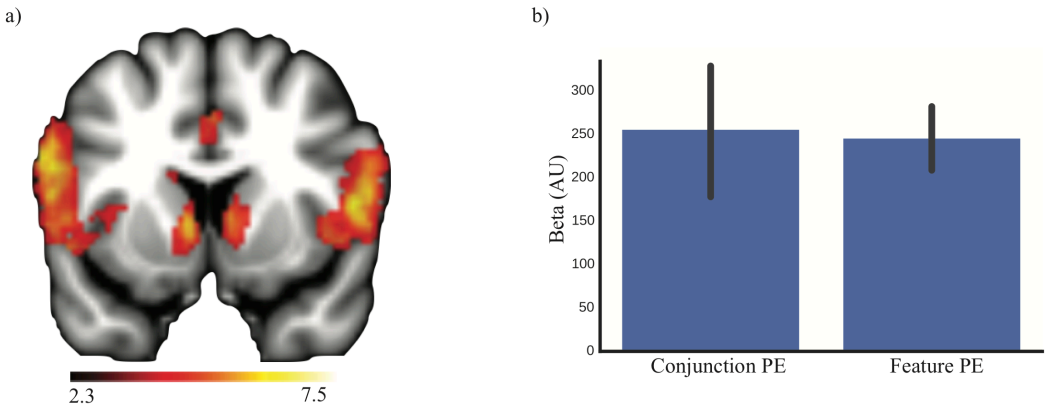

Figure 2. Striatal error response.

a) Regions responsive to PEs from Feature RL model (whole-brain analysis; $p < .05$).

b) An ROI analysis of striatum showed that voxels with responses that scaled with PEs from the Feature RL model also scaled with PEs from the Conjunctive RL model. The Feature PE bar is a statistically independent depiction of the striatal response in (a). The Conjunction PE bar shows that errors from a conjunctive learning system correlated with striatal BOLD above and beyond errors from a feature learning system.

4.3 Pattern Similarity Analysis

We hypothesized the hippocampus formed conjunctive representations of task stimuli, which served as inputs to striatum for reinforcement learning. We used a pattern similarity analysis (PSA) to probe the representational content of hippocampus (25). The PSA compares the similarity of activity patterns among different trials as a function of experimental variables of interest. We computed similarity matrices from the hippocampus, IFS, PRc, and PHc.

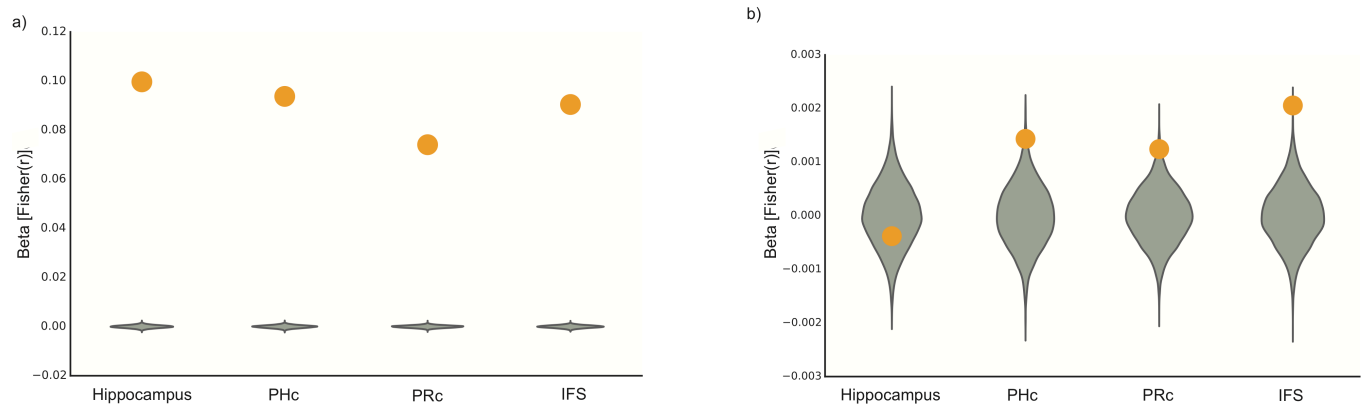

Figure 3. Pattern similarity analysis.

a) A regression analysis on PSA matrices showed strong within-stimulus coding in all ROIs, and within-stimulus coding was significantly stronger in hippocampus relative to other regions. The y-axis shows regression weights from a within-stimulus regressor on the PSA matrix of each ROI.

b) PHc, PRc, and IFS showed increased similarity for pairs of stimuli that shared features and significantly more similarity for these pairs than the hippocampus, consistent with pattern-separated representations in the hippocampus. Green violins show the null distributions of regression coefficients from 10,000 randomly permuted PSA matrices. The y-axis shows regression weights from an overlapping-versus-non-overlapping stimuli regressor on the between-stimuli correlations from the PSA matrix of each ROI.

We first tested whether representations of stimuli in the hippocampus remained stable across trials because to be useful for learning, a region must have consistent representations across presentations of a stimulus. We ran a regression analysis on PSA matrices to assess the similarity among representations from different presentations of a stimulus. We tested the significance of each effect by permuting the PSA matrices 10,000 times to build a null distribution of regression coefficients. All ROIs had significantly higher similarity for repetitions of the same stimulus (within-stimulus similarity) than for pairs of different stimuli (between-stimulus similarity; all $p < .001$, FDR corrected). Across-region comparisons showed the hippocampus had stronger within-stimulus coding than PRc ($p < .001$), PHc ($p < .001$), and IFS, ($p < .001$, FDR corrected, Figure 3a), indicating the hippocampus had the most stable representations of task stimuli.

Our central hypothesis was that the hippocampus, not PRc, PHc, or IFS, would form conjunctive representations of stimuli. Representations of stimuli that shared features (e.g., AB and B) should be pattern separated, and therefore less correlated, in hippocampus but more similar in cortical regions like PRc and PHc that provide inputs to the hippocampus. Because our task biased subjects away from a rule-based strategy, we predicted that the IFS should not have pattern-separated representations of conjunctions. We tested whether the representational structure in each ROI was more similar for stimuli sharing common features than for stimuli that lacked feature overlap. All control ROIs showed a significant effect of overlap (PRc: p = .008, PHc: p = .008, IFS: p < .001, FDR corrected, Figure 3b) but the hippocampus did not (p > .3). Critically, the hippocampus showed significantly lower representational overlap than PRc (p = .026), PHc (p = .026), and IFS (p = .002, all FDR corrected). Control analyses ruled out potential confounds arising from feature hemifield and reproduced these findings using a parametric mixed-effects model (Supplemental Information). Relative to the control ROIs, the hippocampus formed more pattern-separated conjunctive representations of stimuli.

Hippocampal representations of conjunctions could serve as inputs to the striatal reinforcement learning system. Variability in the formation of pattern-separated conjunctive representations in hippocampus should correlate with striatal learning about conjunctions. When the hippocampus demonstrates relatively more pattern-separated representations, the striatal error signal should more strongly track PEs estimated from the Conjunctive RL model. To examine this relationship, we fit a mixed effects model of the conjunctive component of the striatal PE, with subject as a random intercept and hippocampal overlap as a random slope. The hippocampal overlap term was negatively related to the striatal conjunctive PE ($t(31)$ = -3.43, p = .001, d_z = -0.62, Figure S1). We observed similar relationships in our medial temporal lobe (MTL) cortical ROIs (Supplemental Information) and a positive relationship between the strength of within-stimulus similarity in the hippocampus and striatal conjunctive PE, although this last result depended on the exclusion of an outlier subject (Figures S1). In sum, the more the hippocampus established pattern-separated representations of stimuli, the more striatal error signals tracked the true conjunctive state space.

4.4 Representational Content Analysis

The previous analyses show that the hippocampus has the most distinct representations of stimuli that share features among our regions of interest. However, the null effect of stimulus overlap in the hippocampus cannot confirm that this area contains pattern-separated representations. To directly test this hypothesis, we probed the content of its representations using estimates of categorical feature coding acquired from independent localizer data. If hippocampal conjunctive representations are pattern separated from their constituent features, then they are not composed of mixtures of representations of those features (26, 27) (Figure 4a). Unlike high-level sensory cortex, the hippocampal representation of {face and house} would not be a mixture of the representation of {face} and {house} (26). We predicted the hippocampal representations of two-feature stimuli ({face and house} trials) in our learning task should be dissimilar from representations of faces and houses in the localizer. Hippocampal representations of one-feature trials ({face} trials), which are less conjunctive because they contain only one task-relevant feature, should be more similar to representations of the same one-feature category (e.g., faces) in the localizer. In contrast, cortical representations of both two-feature and single-feature trials should be similar to representations of their corresponding features in the localizer (Figure 4a). We predicted hippocampal representations would be less similar to feature templates than cortical ROIs, and only the hippocampus would show would less similarity for two-feature than single-feature trials.

We correlated the patterns in each ROI with the corresponding localizer feature templates (Methods) but were unable to detect reliable feature responses from IFS in the localizer data (c.f. 28). We found significant similarity among task patterns and feature templates for all conditions ($p < .001$, FDR corrected) except for hippocampal responses to conjunctive stimuli ($p = .116$). The hippocampus had lower similarity to feature templates than PRc ($p < .001$) and PHc ($p < .001$). This effect was not likely driven by regional signal quality differences, as the hippocampus had the strongest within-stimulus coding (Figure 3a). The hippocampal feature template was more similar to the response to a single-feature stimulus than to a two-feature stimulus ($p < .001$), consistent with a gradient in pattern separation as the number of task-relevant features increased. This effect was larger in the hippocampus than either PRc ($p < .001$) or PHc ($p < .001$). We confirmed these results using a parametric mixed effects analysis and also performed a control analysis to verify the results were not driven by stimulus-general activation (Figure S4). Unexpectedly, we observed that similarity in PRc and PHc was stronger for two-

feature than for single-feature stimuli (both $p < .001$); in the mixed effects model, this result was marginal in PRc and nonsignificant in PHc and should be interpreted with caution.

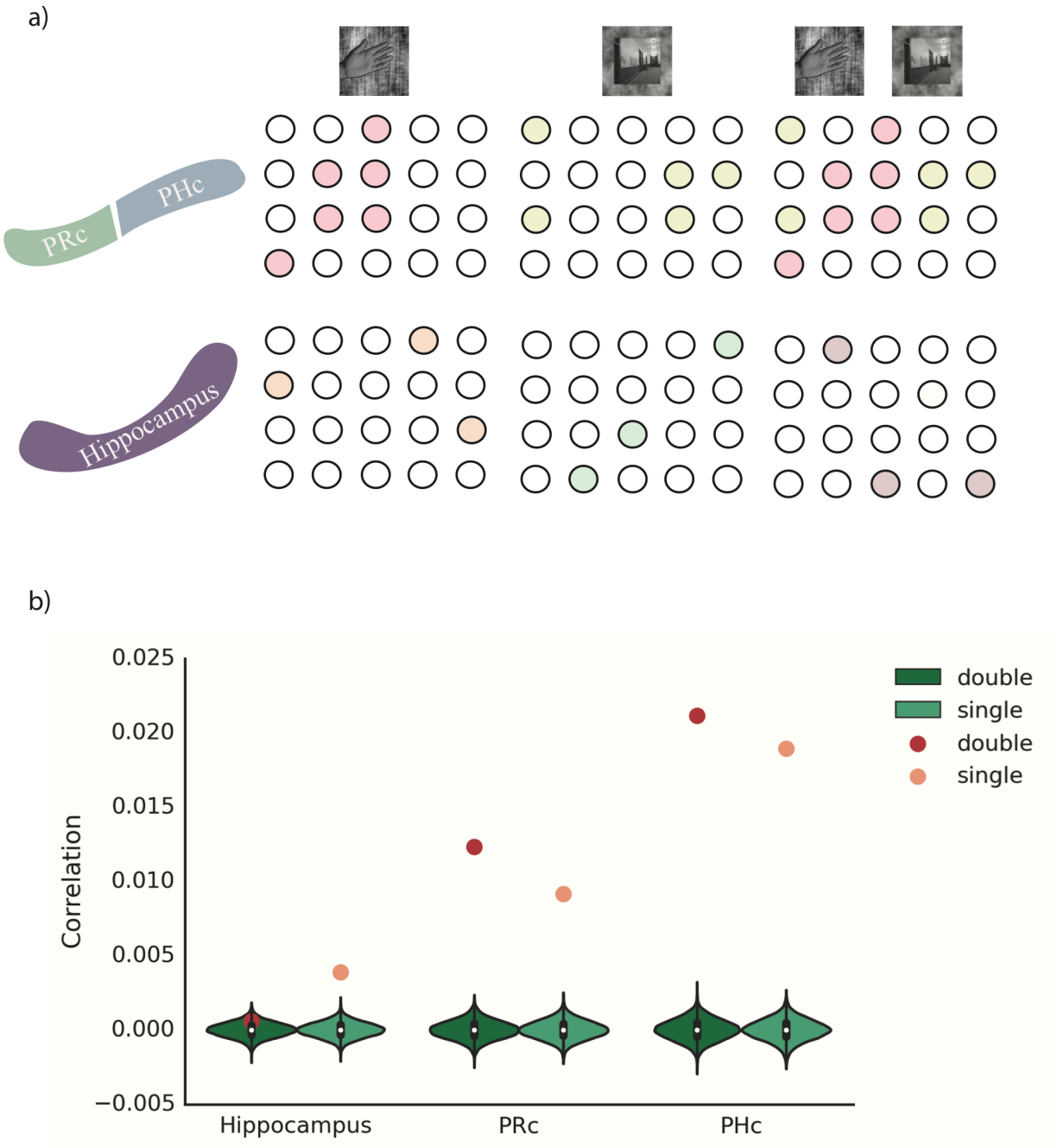

Figure 4. Representational content analysis.

a) Neural predictions: top panel is putative neural ensembles in high-level sensory cortex (parahippocampal, PHc; perirhinal, PRc) for task stimuli (29). Two-feature stimulus should be

represented as union of responses to component features. The lower panel shows putative neural ensembles in the hippocampus; the neural representation of two-feature conjunctions should be orthogonal to responses to its component features.

b) Hippocampal representations were less similar to feature templates than PRc and PHc representations, consistent with increased conjunctive coding. In the hippocampus, representational similarity to templates was higher for single-feature than two-feature stimuli, consistent with increased pattern separation for stimuli with multiple task-relevant features. PRc and PHc showed increased similarity for two-feature relative to one-feature stimuli.

5 DISCUSSION

We tested whether pattern separation in hippocampus enabled learning stimulus-outcome relationships over multifeatural stimuli. We used a novel reinforcement learning task that required learning over non-spatial conjunctions of features. The hippocampus encoded stable representations across repetitions of a stimulus, and conjunctive representations were distinct from the representations of composite features. The hippocampus showed stronger evidence for pattern-separated conjunctive representations than PRc, PHc, and IFS. Hippocampal coding was also related to PE coding in the striatum. Our results suggest that the hippocampus provides a pattern-separated state space that supports the learning of outcomes associated with conjunctive codes.

There is increased interest in a potential role of the hippocampus and MTL systems in reinforcement learning. Deep convolutional neural networks trained on natural images produce patterns of responses strikingly similar to the inferotemporal (IT) processing hierarchy (30), and when combined with reinforcement learning, these networks can learn to achieve human-level performance on video games (31). This model has been augmented with an MTL-cortex-like system that retrieves related memories during learning, and the model can learn some games nearly as quickly as humans (32). Such work illustrates the importance of similarity-based recall, a function associated with MTL cortical computation (14, 33), for effective reinforcement learning. Our data suggest future models that exploit the computational and representational properties of hippocampus will benefit from an enhanced ability to distinguish between similar stimuli and environments with different associated outcomes.

Lesion studies showing dissociations between the hippocampus and striatum in learning (34) along with some imaging studies demonstrating a negative relationship between hippocampal and striatal learning signals (35) have led to the hypothesis that these regions compete during learning. By contrast, other evidence document cooperative interactions. For example, neurons in the striatum represent spatial information derived from hippocampal inputs (36) and contextual information in the hippocampus drives the formation of conditioned place preferences via its connection to ventral striatum (37, 38). These findings support a model in which hippocampal information about spatial contexts, location, or conjunctions serve as inputs for striatal associative learning.

We contend the hippocampus forms representations of conjunctions of features that are reinforced via dopamine release on hippocampal-striatal synapses, but the hippocampus could form a representation of the temporal sequence of task events (39). In AB+ trials, AB would trigger a representation of the target in the hippocampus, and this target representation could then feed into the striatum or prefrontal cortex to drive responses. This model is similar to the idea that the hippocampus encodes a "successor representation" for reinforcement learning (40) in which the target representation occurs in proportion to the probability of each stimulus preceding the target. Both explanations differ in mechanism but require a conjunctive representation in the hippocampus. Future work should directly test the role of hippocampal sequence representation in reinforcement learning.

The circuit properties of the hippocampus allow it to rapidly bind distributed cortical representations of features into orthogonalized conjunctive representations. Hippocampal pattern completion, triggered by partial cues, along with recurrent outputs back to sensory cortex allow the hippocampus to reactivate the ensemble of event features that constitute the retrieval of an episodic memory (2, 41). Dense inputs to the striatum suggest hippocampal representations could also form the basis for associative learning over conjunctive codes. Our results extend the role of the hippocampus to include building conjunctive representations that are useful for striatal outcome and value learning.

6 *METHODS*

6.1 Data and Software Availability

All MRI and behavioral data will be made available at OpenfMRI, and all analysis code will be made available on GitHub prior to publication. Key resources are listed in Table S2.

6.2 Experimental Model and Subject Details

The study design and methods were approved by the Stanford Institutional Review Board. Forty subjects provided written informed consent. Data from eight subjects were excluded from analyses: One ended the scan early due to claustrophobia; three had scanner-related issues that prevented reconstruction or transfer of their data; two had repeated extreme (>2mm) head movements across most runs; and two subjects demonstrated extremely poor performance, as indexed by less than $2.50 of earnings (see below for payment details). Note that our task is calibrated to the individual subjects' practice data in such a way that a simple target detection strategy would be expected to earn $7.50, and any effort to learn the task should improve on these earnings. This left 32 subjects in the analysis cohort, 19 females, mean age 22.1 yrs, SD 3.14, range 18 to 29. Due to an error, behavioral data for one subject were lost; thus, while her imaging data were included in fMRI analyses, all behavioral analyses were conducted with a sample of 31 subjects.

6.3 Task

Subjects performed a target detection task in which performance could be improved by learning predictive relationships between visually presented stimuli and the target. The target appeared 70% of the time for two-feature stimulus AB as well as the single-feature stimulus C, and 30% of the time for two-feature stimulus AC as well as the single-feature stimulus B. The task bears strong similarities to the "ambiguous feature discrimination task" used to study rodent learning (19). Subjects were instructed that they would earn 25¢ for each correct response, lose 25¢ for each incorrect response, or no money for withheld responses. Response time (RT) thresholds were calibrated for each subject during training so that responses initiated by perception of target onset would lead to success on 50% of trials. Earning rewards on more than

50% of trials therefore required anticipating target onset based on the preceding stimulus (A, B, AB, or AC). During the experiment, RT thresholds were slowly adjusted to account for gradual speeding of responses over the course of the task. Subjects were instructed that in order to earn the most money, they should learn which stimuli predicted the target and respond as quickly as possible, even if the target has not yet appeared. These instruction were meant to bias subjects towards an instrumental learning strategy, rather than an explicit rule-based learning strategy (42).

Subjects performed one practice run and were instructed on the types of relationships they might observe. During fMRI scanning, subjects performed three runs of the task. Each run consisted of 10 trials for each stimulus (AB+, AC-, B-, C+), resulting in 40 total trials per run. Features A, B, and C were mapped to a specific house, face, and body part image for the duration of the run. The category-to-stimulus mapping was counter-balanced across runs, resulting in each visual category being associated with each feature type (A, B or C) over the course of three runs. Further, different participants saw different stimuli within each category. Each subject encountered the same pseudo-random trial sequence of both stimuli and targets, which facilitated group modeling of parametric prediction error effects. Features could appear on either the left or the right of a fixation cross, with the assignment varying randomly on each trial. For single-feature stimuli, the contralateral location was filled with a phase-scrambled image designed to match the house/face/body part features on low-level visual properties. Inter-trial intervals and the interval between the stimulus/stimulus+target and feedback were taken from a Poisson distribution with a mean of 5 s, truncated to have a minimum of 2 s and maximum of 12 s. Visual localizer task details are described in *SI Methods*.

6.4 Behavioral Analysis

We used reaction time data from subjects to infer learning in the task, an approach that has been used successfully in a serial reaction time task (43). Log-transformed reaction times were fit with linear regression. For the three RL models, we modeled reaction time with a value regressor taken from a simple reinforcement learning model:

$$V(s)^{t+1} = V(s)^{t} + \alpha[R_t - V(s)^{t}]$$

where R_t is an indicator on whether or not the target appeared, α is the learning rate, and *s* is the state. These values represent the strength of association between a stimulus and a target/outcome, and are not updated based on the reward feedback, which also depends on whether the subject responded quickly enough. The values we measured are more relevant for learning because they correspond to the probability that the subject should respond to the stimulus. For the Conjunctive and Hybrid models, the state corresponded to the current stimulus $s \in \{B, C, AB, AC\}$. For the Feature model, the states were single features $s \in \{A, B, C\}$ and in two-feature trials (e.g., AB), the value was computed as the sum of the individual feature values (e.g., V(AB) = V(A) + V(B)), and both feature values were updated after feedback.

The Hybrid RL model was a variant of the Conjunctive model in which a portion of the value update blends onto the overlapping stimulus:

$$V(s')^{t+1} = V(s')^{t} + \alpha[R_t - V(s)^{t}] * \omega O(s, s'), \forall\, s' \neq s$$

where $O(s, s')$ is an indicator function that is equal to 1 when the two stimuli share a feature and 0 otherwise, and ω is a spread parameter that controls the magnitude of the spread. For example, if the current stimulus is AB, a proportion of the value update for AB would spread to B and to AC. We allowed value to spread between any stimuli sharing features (e.g., an AB trial would lead to updates of both AC and B). This approach reflects the fact that not only will conjunctions activate feature representations in cortex, but features can activate conjunctive representations, a property that has been extensively studied in transitive inference tasks (44). We fit the Conjunctive RL model, the Feature RL model, and the Hybrid RL model using Scipy's minimize function. The linear regression weights were fit together with the parameters of the learning model. The fitted regression weights for the Hybrid model were significantly less than 0, $T = 64$, $p < .001$, Wilcoxon test, indicating that stimuli more strongly associated with the target are associated with faster reaction times. Finally, we fit a null No Learning model with no value regressor (and therefore fits to only mean RT). Model comparison procedures are described in detail in *SI Methods*.

6.5 fMRI Modeling

fMRI acquisition and preprocessing as well as ROI selection procedures are described in

detail in *SI Methods*. Analysis was conducted using an event-related model. Separate experimental effects were modeled as finite impulse responses (convolved with a standard hemodynamic response function). We created separate GLMs for whole brain and for pattern similarity analyses (PSA). For the whole brain analysis, we modeled the 1) stimulus period as an epoch with a duration of 2 s (which encompasses the stimulus, target and response) and 2) the feedback period as a stick function. In addition, we included a parametric regressor of trial-specific prediction errors extracted from a Feature reinforcement learning model that learns about A, B, and C features independently, without any capacity for learning conjunctions. In addition, we included a regressor that was computed by taking the difference between the Feature RL errors and the Conjunctive RL model errors. This regressor captured variance that was better explained by Conjunctive RL prediction errors than by Feature RL prediction errors. Both parametric regressors were z-scored and were identical across all subjects. We included nuisance regressors for each slice artifact and the first principal component of the deep white matter, which captures residual nuisance components of the whole-brain signal. Following standard procedure when using ICA denoising, we did not include motion regressors as our ICA approach is designed to remove motion-related sources of noise. GLMs constructed for PSA had two important differences. First, we did not include parametric prediction error regressors. Second, we created separate stimulus regressors for each of the 40 trials. Models were fit separately for individual runs and volumes of parameter estimates were shifted into the space of the first run. Fixed effects analyses were run in the native space of each subject. We estimated a nonlinear transformation from each subject's T1 anatomical image to the MNI template using ANTs. We then concatenated the functional-to-structural and structural-to-MNI transformations to map the fixed effects parameter estimates into MNI space. Group analyses were run using FSL's FLAME tool for mixed effects.

6.6 PSA Analysis

We were interested in distinguishing the effect of different experimental factors on the representational similarity matrices (PSM). PSM preprocessing is described in *SI Methods.* We constructed linear models of each subject's PSM. We included main regressors of interest as well as several important regressors that controlled for similarities arising from task structure. The linear model included: 1) a nuisance "response" regressor that was coded as 1 for stimuli that

shared a response (both target or both non-target) and -1 otherwise, 2) a nuisance "target" regressor that was coded as 1 for stimuli that both had a target, -1 for both non-target, and 0 otherwise, 3) a "within-stimulus similarity" regressor that was 1 for pairs of stimuli that were identical and 0 otherwise, 4) an "overlap" regressor that was coded as 1 for pairs of stimuli that shared features, -1 for those that did not, and 0 for pairs of the same stimuli, 5) a "prediction error" regressor that was computed as the absolute value of the difference in trial-specific prediction errors, extracted from the Hybrid reinforcement learning model, between the two stimuli (Figure S3), 6) a "value" regressor that was computed as the absolute value of the difference in trial-specific updated values, extracted from the Hybrid reinforcement learning model, between the two stimuli (Figures S3), 7) nuisance regressors for the mean of the runs, and 8) two "time" regressors that accounted for the linear time elapsed between the pair of stimuli and its interaction with the within-stimulus similarity regressor. We included this last interaction because within-stimulus similarity effects were by far the most prominent feature of the PSA, and temporal effects were therefore more likely to have larger effects on this portion of the PSMs. We included prediction error (5) and value (6) regressors because we were interested in exploratory analyses of these effects based on theoretical work suggesting that the hippocampus pattern separates stimuli based on the outcomes they predict (39). Both the value and the prediction error regressors were orthogonalized against the response and target regressors, thereby assigning shared variance to the regressors modeling outcomes. All regressors were z-scored so that their beta weights could be meaningfully compared. Correlations, our dependent variable, were Fisher transformed so that they followed a normal distribution. To assess the significance of the regression weights as well as differences between regions, we compared empirical regression weights (or differences between them) to a null distribution of regression weights (or differences between them) generated by shuffling the values of the PSA matrices 10,000 times. In addition, we fit a linear mixed effects model using R with subject as a random intercept, ROI as a dummy code with hippocampus as the reference, and ROI by task interactions for each of the above regressors. Using random slopes resulted in convergence errors, and so we did not include them. By using both a parametric and a nonparametric approach to assessing our data, we gained confidence that our results are robust to differences in power between different statistical analysis techniques due to outliers or violations of distribution assumptions.

6.7 Representational Content Analysis

Data from the localizer task were preprocessed and analyzed in the same manner as the main task data. GLMs were constructed for each run and included a boxcar regressor for every miniblock with 4-s width, as well as a nuisance regressor for the targets, each slice artifact and the first principal component of the deep white matter. To compute template images, we computed the mean across repetitions of each stimulus class (face, place, character, object, body part). For the representational content analysis depicted in Figure 4, we computed the correlations as follows: Assume that A is a face, B is a house, and C is a body part. For "single-feature" stimuli, we computed the similarity of B trials with the house template and C trials with the body part template. For "two-feature" stimuli, we computed the similarity of AB trials with the house template and AC trials with the body part template. Therefore, within each run, the task-template correlations for AB and B (and AC and C) were computed with respect to the same feature template. This means that any differences between AB and B (or AC and C) correlations reflect differences in the task representations, rather than potential differences in the localizer representations. We repeated this for each run's stimulus category mappings.

7 *AUTHOR CONTRIBUTIONS*

Conceptualization, ICB and SMM; Methodology, ICB; Software, ICB; Formal Analysis, ICB; Data Curation, ICB; Writing Original Draft, ICB; Writing, Review & Editing, SMM and ADW.; Funding Acquisition, ICB and SMM; Supervision, SMM and ADW.

8 *REFERENCES*

1. Davis T, Xue G, Love BC, Preston AR, Poldrack RA (2014) Global neural pattern similarity as a common basis for categorization and recognition memory. *J Neurosci* 34(22):7472–7484.

2. McClelland JL, McNaughton BL, O'Reilly RC (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(3):419.

3. Marr D (1971) Simple memory: a theory for archicortex. *Philos Trans R Soc Lond, B, Biol Sci* 262(841):23–81.

4. Leutgeb JK, Leutgeb S, Moser M-B, Moser EI (2007) Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus. *Science* 315(5814):961–966.

5. O'Reilly RC, McClelland JL (1994) Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus* 4(6):661–682.

6. Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. *Nat Rev Neurosci* 7(6):464–476.

7. Reynolds JNJ, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. *Nature* 413(6851):67–70.

8. Schultz W (1997) A Neural Substrate of Prediction and Reward. *Science* 275(5306):1593–1599.

9. Stuber GD, et al. (2008) Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* 321(5896):1690–1692.

10. Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. *Nature* 413(6851):67–70.

11. Bayley PJ, Frascino JC, Squire LR (2005) Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* 436(7050):550–553.

12. Finch DM (1996) Neurophysiology of converging synaptic inputs from the rat prefrontal cortex, amygdala, midline thalamus, and hippocampal formation onto single neurons of the caudate/putamen and nucleus accumbens. *Hippocampus* 6(5):495–512.

13. Goto Y, Grace AA (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* 8(6):805–812.

14. Norman KA, O'Reilly RC (2003) Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychol Rev* 110(4):611–646.

15. Davachi L, Mitchell JP, Wagner AD (2003) Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences* 100(4):2157–2162.

16. Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences* 7(9):415–423.

17. Waskom ML, Frank MC, Wagner AD (2017) Adaptive Engagement of Cognitive Control in Context-Dependent Decision Making. *Cereb Cortex* 27(2):1270–1284.

18. Ballard I, Miller EM, Piantadosi ST, Goodman ND, McClure SM (2017) Beyond Reward Prediction Errors: Human Striatum Updates Rule Values During Learning. *Cerebral Cortex* 19(3):1–11.

19. Gallagher M, Holland PC (1992) Preserved configural learning and spatial learning impairment in rats with hippocampal damage. *Hippocampus* 2(1):81–88.

20. Niv Y, et al. (2015) Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *J Neurosci* 35(21):8145–8157.

21. Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies - revisited. *Neuroimage* 84:971–985.

22. Wimmer GE, Shohamy D (2012) Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. *Science* 338(6104):270–273.

23. McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38(2):339–346.

24. Tziortzi AC, et al. (2013) Connectivity-Based Functional Analysis of Dopamine Release in the Striatum Using Diffusion-Weighted MRI and Positron Emission Tomography. *Cereb Cortex* 24(5):bhs397–1177.

25. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proceedings of the National Academy of Sciences* 103(10):3863–3868.

26. O'Reilly RC, Rudy JW (2001) Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol Rev* 108(2):311.

27. Rudy JW, Sutherland RJ (1995) Configural association theory and the hippocampal formation: an appraisal and reconfiguration. *Hippocampus* 5(5):375–389.

28. Kuhl BA, Rissman J, Wagner AD (2012) Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia* 50(4):458–469.

29. Liang JC, Wagner AD, Preston AR (2013) Content Representation in the Human Medial Temporal Lobe. *Cereb Cortex* 23(1):80–96.

30. Yamins DLK, et al. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111(23):8619–8624.

31. Mnih V, et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518(7540):529–533.

32. Pritzel A, Uria B, Srinivasan S, Puigdomènech A (2017) Neural Episodic Control. *arXiv*.

33. LaRocque KF, et al. (2013) Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. *J Neurosci* 33(13):5466–5474.

34. Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol Learn Mem* 65(1):65–72.

35. Poldrack RA, et al. (2001) Interactive memory systems in the human brain. *Nature* 414(6863):546.

36. Mulder AB, Tabuchi E, Wiener SI (2004) Neurons in hippocampal afferent zones of rat striatum parse routes into multi-pace segments during maze navigation. *European Journal of Neuroscience* 19(7):1923–1932.

37. Pennartz CMA, Ito R, Verschure PFMJ, Battaglia FP, Robbins TW (2011) The hippocampal–striatal axis in learning, prediction and goal-directed behavior. *Trends in Neurosciences* 34(10):548–559.

38. Ito R, Robbins TW, Pennartz CM, Everitt BJ (2008) Functional Interaction between the Hippocampus and Nucleus Accumbens Shell Is Necessary for the Acquisition of Appetitive Spatial Context Conditioning. *J Neurosci* 28(27):6950–6959.

39. Gluck MA, Myers CE (1993) Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus* 3(4):491–516.

40. Stachenfeld KL, Botvinick MM, Gershman SJ (2017) The hippocampus as a predictive map. *Nature Publishing Group* 20(11):1643–1653.

41. Gordon AM, Rissman J, Kiani R, Wagner AD (2014) Cortical Reinstatement Mediates the Relationship Between Content-Specific Encoding Activity and Subsequent Recollection Decisions. *Cereb Cortex* 24(12):3350–3364.

42. Sternberg DA, McClelland JL (2012) Two Mechanisms of Human Contingency Learning. *Psychological Science* 23(1):59–68.

43. Bornstein AM, Daw ND (2012) Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience* 35(7):1011–1023.

44. Kumaran D, McClelland JL (2012) Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychol Rev* 119(3):573–616.