Supplementary Materials for

Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions

G. Elliott Wimmer and Daphna Shohamy*

*To whom correspondence should be addressed. E-mail: shohamy@psych.columbia.edu

Published 12 October 2012, *Science* **338**, 270 (2012)
DOI: 10.1126/science.1223252

This PDF file includes:

Materials and Methods

Participants

Thirty-one fluent English speakers with normal or corrected-to-normal vision participated in the study. All subjects were free of neurological or psychiatric disorders and consented to participate. Informed consent was obtained in a manner approved by the Columbia University Institutional Review Board. Three subjects' data were excluded because they did not show evidence of simple reward learning for the S_2 stimuli as measured in the Decision phase (less than 75% preference for all three reward-associated S_2 stimuli). Behavioral and functional imaging data are presented from the remaining twenty-eight subjects (17 female; mean age, 23 years; range, 18-32). Subjects were paid $20 per hour for the approximately 2-hour duration of participation, plus one-fifth of the nominal rewards they earned in the experiment.

Procedure

The experimental paradigm was a newly developed variant of a "sensory preconditioning" paradigm (*4, 14, 35, 36*). Functional magnetic resonance imaging (fMRI) data were collected during three phases: Association, Reward, and Decision (**Figure 1**). Additional behavioral data were collected before and after the task, as detailed below.

Pre-task stimulus liking ratings

Stimuli for each subject were selected based on pre-task liking ratings. Subjects rated 20 stimuli in each of the three S_1 picture categories (faces, scenes, and body parts) and 20 patterned circle S_2 stimuli using a graded scale anchored with "strong dislike" and "strong like".

Subject's liking ratings were used to select approximately neutral stimuli that were also closely matched in liking across categories; this procedure ensured that subjects did not have strong pre-existing preferences for any of the experimental stimuli. The selection algorithm picked two neutrally rated S_1

pictures from each of the face, scene, and body part categories and six neutrally rated S_2 circle stimuli. From these S_1 and S_2 stimuli, six pairs were constructed. With three S_1 categories and two pairs per category, the experiment contained three pair sets of stimuli. (See supplementary text for descriptive statistics of subjects' pre-experiment stimulus ratings.) Ratings data were collected during the high-resolution anatomical scan using an MRI-compatible trackball response unit; a subset of four subjects instead completed ratings on a laptop computer before entering the MRI scanner.

Liking ratings were collected again at the end of the experiment, allowing an assessment of changes in liking for the stimuli as a consequence of reward learning and decision bias, as detailed below.

Association phase

In the Association phase, subjects were incidentally exposed to temporal pairings of stimuli (**Figure 1A**). On each trial, a picture (S_1; face, scene, or body part) preceded a patterned circle (S_2). A particular S_1 stimulus always preceded its paired S_2 stimulus. Subjects were not informed of these pair relationships or of the trial structure. To distract subjects from becoming explicitly aware of the associations, they were given a cover task that instructed them to respond to occasional inverted target face, scene, or body pictures. Targets were followed by a circle stimulus; the inverted target picture and the associated circle were not part of the critical stimulus pairs. Subjects responded to the S_1 and S_2 stimuli with a button press and to inverted target stimuli with an alternate button press. Each pair was presented 10 times in pseudo-random order, intermixed with 18 target trials.

An Association phase trial consisted of the presentation of an S_1 stimulus for 1.75 s, an inter-stimulus interval (ISI) for 1.75 s, an S_2 stimulus for 1.75 s, and an inter-trial-interval (ITI) for 3.5 s. To improve the implicit pairing of the stimuli, the ITI was twice as long as the ISI (following *13, 15*). FMRI data were collected in two blocks of approximately 6 min duration each.

Reward phase

In the Reward phase, subjects underwent a Pavlovian conditioning procedure in which S_2 stimuli were predictive of either reward or neutral outcomes (**Figure 1B**). Only S_2 stimuli were used as conditioned stimuli; no S_1 stimuli appeared during this phase. Three of the S_2 circle stimuli were paired with a reward (S_2+) and the remaining three were paired with a neutral outcome (S_2-). One S_2 stimulus previously associated with each category of S_1 stimulus (face, scene, and body part) was paired with reward and one with neutral feedback, giving an S_2+ and S_2- stimulus for each category. Each S_2 stimulus was presented on 16 Reward phase trials in a pseudo-random order. For S_2+ stimuli, the reward outcome appeared with a probability of 81% (13/16 trials); for S_2- stimuli, the neutral outcome appeared on all trials.

To maintain continuity between phases, subject instructions were similar to the target detection instructions in the preceding Association phase. Subjects were instructed to respond to a target reward stimulus (a picture of a one-dollar bill) with a button press and the other stimuli (S_2 stimuli and the neutral grey square stimulus) with an alternate button press. Subjects were instructed that when they correctly responded to the reward stimulus, the reward would be added to their earnings. Subjects were also informed that they might notice predictive associations between particular circle stimuli and reward or neutral outcomes.

A Reward phase trial consisted of the presentation of an S_2 stimulus for 2 s, a fixation ISI for 2 s, a reward or neutral outcome for 2 s, and a variable ITI fixation of mean 2 s (range: 0.5-10.5 s). The duration and distribution of null events was optimized for the estimation of rapid event-related fMRI responses as calculated using Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/). FMRI data were collected in two blocks of approximately 6.5 min duration each.

Decision phase

In the Decision phase, we assessed subject's preferences for the S_2 and S_1 stimuli. To assess reward learning, S_2 choices were between a reward-

associated S_2+ circle and a neutral-associated S_2- circle. To assess decision bias, S_1 choices were between S_1+ and S_1- stimuli, i.e., between an S_1 stimulus from an Association phase S_1-S_2 pair where the incidentally associated S_2 was later rewarded (S_1+) and an S_1 where the associated S_2 was not rewarded (S_1-) (**Figure 1C**). Choices between S_1 stimuli were always within-category (e.g. between two face pictures). Subjects were instructed to choose the option that they thought was more likely to lead to winning $1. Subjects were informed that the outcome of their choice would not be presented during the Decision phase and that they would receive a percentage of their earnings at the end of the experiment.

On each trial, two stimuli were presented (randomly permuted on the left and right side). Subjects selected the left or right picture with a corresponding left or right button response during a 2.5 s response period. From the time the subject made a selection until the end of the choice presentation period, the selected option was framed in blue, followed by a jittered ITI with a mean of 2 s (range: 0.5-10.5 s). If no choice was recorded during the choice period, the options remained on the screen until the end of the trial. Each critical choice between reward- and neutral-associated stimuli was presented four times, in pseudo-random order, yielding 24 trials. FMRI data were collected in one block of approximately 5 min duration.

Post-task stimulus ratings

After the completion of the Decision phase and while still in the scanner, subjects again rated their liking for the stimuli to test whether the experiment led to changes in liking. Subjects were presented with each of the S_1 and S_2 stimuli from the experiment and were instructed to indicate how much they currently liked each stimulus.

Explicit memory test

Outside of the scanner, subjects completed a questionnaire on a laptop computer that probed memory for the S_1-S_2 pair relationships from the initial

Association phase of the experiment. A single S_1 face, scene, or body part picture was shown above two S_2 circle stimuli options: one that had been incidentally paired with the presented S_1 stimulus in the initial Association phase of the experiment and a lure that had been paired with a different S_1 stimulus. Subjects were instructed to select the circle (S_2) that seemed "related to" the picture (S_1).

Next, subjects completed a computerized questionnaire that instructed them to rate the likelihood that the S_2 and S_1 stimuli would be associated with winning money. Subjects then answered a series of written questions that assessed both memory for and awareness of patterns of presentation during the Association phase and Reward phase, and choice strategies for the Decision phase (see supplementary text). Finally, subjects were paid for their participation and were given one-fifth of their winnings from the Reward and Decision phases of the experiment.

Stimuli

All phases of the task were presented using the Matlab (Natick, Massachusetts) and the Psychophysics Toolbox (*37*). Stimuli were projected onto a mirror above the subject's eyes in the MRI system. The face stimuli were selected from the Stanford Face Database and the CVL Face database (Peter Peer, http://www.lrv.fri.uni-lj.si/facedb.html). Scene stimuli were selected from an internal database. Body part stimuli were selected via a search of publicly available images on the internet.

Imaging procedure

Whole-brain imaging was conducted on a 3.0T Phillips MRI system at Columbia University's Program for Imaging and Cognitive Sciences, using a SENSE head coil. Head padding was used to minimize head motion; no subject's motion exceeded 2 mm in any direction from one volume acquisition to the next. Structural images were collected using a high-resolution T1-weighted MPRAGE pulse sequence (1 X 1 X 1 mm voxel size). Functional images were collected

using a gradient echo T2*-weighted echoplanar (EPI) sequence with blood oxygenation level-dependent (BOLD) contrast (TR = 2000 ms, TE = 20 ms, flip angle = 72, 2 X 2 X 3 mm voxel size; 38 contiguous axial slices). For each functional scanning run, five discarded volumes were collected prior to the first trial to allow for magnetic field equilibration.

Behavioral analysis

We analyzed subjects' responses during the Decision phase of the task and during the post-task tests (stimulus liking and explicit memory). We used subjects' tendency to choose S_2+ over S_2- stimuli as a measure of how well they learned the association between S_2 stimuli and reward or neutral feedback in the Reward phase. Decision bias was operationalized as subjects' tendency to choose S_1+ stimuli over S_1- stimuli. A single preference score for each of the three reward-associated S_2+ and the three associated S_1+ stimuli was derived by averaging subject's responses over the four presentations of each choice. Since any spreading of value to S_1+ stimuli could be limited by the individual subject's maximal preference for the S_2+ stimuli, we also calculated a relative score of decision bias on a per-subject basis. This was computed by dividing S_1+ choices by each subject's average preference for the S_2+ stimuli; if S_1 decision bias was greater than S_2 stimulus preference, the relative decision bias was set to 100%.

Imaging analysis

Preprocessing and data analysis was performed using AFNI (*38*) and Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were co-registered manually using AFNI. In SPM, images were realigned to correct for subject motion and then spatially normalized by estimating a warping to template space from each subject's anatomical image and applying the resulting transformation to the EPIs. Images were resampled to 2 mm cubic voxels, smoothed with an 8 mm FWHM Gaussian kernel, and filtered with a 128 s high-pass filter. SPM was used to estimate general linear models (GLMs) and

psychophysiological interaction (PPI) analyses. Reactivation and mediation functional connectivity analyses were completed using AFNI and custom routines in Matlab.

For analysis of all phases of the experiment, fMRI model regressors were convolved with the canonical hemodynamic response function and entered into a general linear model (GLM) of each subject's fMRI data. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the GLM to account for residual effects of subject movement. Linear contrasts of the resulting SPMs were taken to a group-level (random-effects) analysis. We report results corrected for family-wise error (FWE) due to multiple comparisons (*39*); this approach assesses the strength of activations defined by an initial and arbitrary uncorrected threshold, which we take as $P < 0.005$ for all analyses. Accordingly, for display purposes, we render all SVC significant activations at this threshold. We conduct this correction at the peak-level within small volumes for which we had an a priori hypothesis or at the whole-brain cluster level. For regions of interest in the striatum and in the MTL (including both the hippocampus and parahippocampal cortex) we used anatomically defined masks derived from the AAL atlas (*40*). Additionally, for the anterior hippocampus, a 6 mm diameter spherical region of interest was drawn at the coordinates reported previously in a related temporal association learning paradigm (28, -10, -22) (*6*). All voxel locations are reported in Montreal Neurological Institute (MNI) coordinates; results are displayed overlaid on the average of all subjects' normalized high-resolution structural images.

Decision bias-related activation

Our main fMRI analysis focused on detecting activation directly related to decision bias during the Reward phase. The experimental design involved three different S_1 picture categories (with an S_1+ and S_1- pair per category), allowing for three decision bias values per subject. Subjects often showed variability in decision bias across the face, scene, and body part S_1 stimulus categories (**Figure 2B** and supplementary text). This variability allowed us to contrast the

BOLD signal (hereafter referred to as “activation”) related to high versus low decision bias for each subject. This analysis included all subjects who exhibited variability in bias, excluding three subjects with 100% bias for all categories and two subjects with 0% bias for all categories. Behavioral decision bias, as a percent of Decision phase choices, was transformed within-subject into a rank measure, yielding a 1 to 3 low- to high-bias scale. A GLM was estimated using ranked decision bias as a parametric regressor (6 s duration) during the Reward phase. A similar GLM was estimated for the Association phase and the Decision phase to test the specificity of Reward phase results.

Reactivation of associations and decision bias

Next, we explored within-subject correlates of decision bias that may be evident in Reward phase reactivation of paired S_1 stimuli. This analysis tested whether decision bias is predicted by reactivation of associations. During reward learning, such reactivation may allow the binding of reward and neutral outcome value to both the presented S_2 stimulus and reactivated S_1 stimulus. Our paradigm utilized three categories of S_1 stimuli (face, scene, and body part pictures), which have been found to activate distinct regions of the visual cortex: fusiform face area (FFA), parahippocampal place area (PPA), and extrastriate body area (EBA) (*41-43*). This specificity allowed us to test if activation during the Reward phase, where only S_2 stimuli were presented, also elicited activation of the category-specific S_1 stimuli incidentally associated with individual S_2 stimuli during the preceding Association phase.

To measure reactivation, we derived stimulus category masks for each subject from the Association phase, applied these masks to contrasts of S_2 stimuli in the Reward phase, and sorted the resulting values by subsequent decision bias. In detail, first, regions of interest were derived from activation to actual S_1 stimuli presented in the Association phase. A GLM was estimated with separate regressors for S_1 face, scene, and body part stimuli as well as “upside-down” target trials (all for 1.75 s duration). Contrasts were constructed to estimate responses to specific S_1 categories: [face - (scene + body)], [scene -

(face + body)], [body - (face + scene)]. The resulting individual contrasts were masked to include only the top 1% of voxels for each contrast which also fell within a group mask (thresholded at $P < 0.001$, uncorrected) based on activation to category-specific stimuli in the Association phase. For face stimuli, we used a mask of the ventral visual cortex, as the group mask did not capture individual face-responsive regions in the ventral occipital cortex (**Figures S4-S5**). The resulting summed subject masks for face, scene, and body part stimuli resemble expected FFA, PPA, and EBA activation patterns, respectively, based on prior literature (*41-43*) (**Figures S4-S5**).

Next, a Reward phase GLM was estimated similar to the model of the Association phase described above. Here, instead of modeling the presentation of S_1 face, scene, and body part stimuli, the GLM modeled the presentation of S_2 circle stimuli (2 s duration) that were incidentally paired with the face, scene, and body part S_1 stimuli in the Association phase. Contrasts were between S_2 stimuli that differed only in their S_1 associations, e.g. [S_2 face - (S_2 scene + S_2 body)]. The category-specific masks from the Association phase were applied to the resulting contrasts of the Reward phase GLM. In the Reward phase, beta values in voxels falling within the mask were averaged to produce one value per S_2-associated category per subject. These values were sorted within-subject according to subsequent decision bias. Across-subject values for high, medium, and low bias were created from the sorted within-subject values. Statistical tests were performed on the resulting data for high- versus low-bias. The majority of subjects had two equivalent high- or low-bias pairs; these were averaged together. The medium bias condition only existed within a subset of 8 subjects and was not included in subsequent analyses due to lack of statistical power.

The Association phase masks were also used to compute mean levels of activation in the Association phase contrasts themselves. These values were used to establish the specificity of Reward phase reactivation in two ways. First, they were verified to not differ according to later decision bias. Second, they were used to normalize Reward phase reactivation estimates by controlling for baseline levels of responding to the actual picture stimuli in the Association

phase. It could be argued that the use of the Association phase for voxel selection induces a confound, given the inclusion of both the S_1 and associated S_2 stimulus on each trial. However, any bias induced by the Association phase will be minimized as the Association phase activation was itself used to normalize the Reward phase reactivation estimates. Additionally, when a similar reactivation procedure was used with masks derived from the Association phase S_2 presentation, Reward phase activation did not correlate with decision bias.

Reward phase parametric learning signals

We additionally tested for BOLD activity related to learning of direct S_2-reward associations during the Reward phase. FMRI data in the Reward phase were analyzed using event onsets that modeled the whole trial (6 s) as well as the onsets for the S_2 stimulus (0.5 s) and the reward outcome (0.5 s).

To model reward learning, we utilized a standard temporal difference reinforcement learning model to regress trial-by-trial estimates of stimulus value and reward prediction error against BOLD activity (*44-46*). The Reward phase associations between S_2 stimuli and outcomes were Pavlovian, with reward appearance independent of behavioral responses, and thus we could not fit the parameters of a reinforcement learning model to behavior. Instead, the learning rate α was set at 0.25 such that stimulus values for S_2 stimuli, initialized at zero, reached asymptote for S_2+ stimuli (approximately 0.7) by the end of the phase. Results are robust to different learning rate settings.

To model stimulus-stimulus association learning, we adapted a temporal difference model to regress trial-by-trial estimates of stimulus “surprise” due to the unpredicted presentation of lone S_2 stimuli against BOLD activity. During the preceding Association phase of the experiment, specific S_1-S_2 temporal associations were incidentally experienced, but in the Reward phase, these associations were violated by the presentation of lone S_2 stimuli. The unexpected omission of the S_1 and the unexpected lone presentation of the S_2 stimuli in the Reward phase may lead to a “stimulus prediction error”, a response previously attributed to the hippocampus and midbrain dopamine system (*47-49*). In this

model, prediction errors are highest at the beginning of reward learning when expectations of joint S_1–S_2 presentation are the highest. As the Reward phase proceeds, expectations are updated via the prediction error signal, leading to decreasing surprise. The prediction error signal thus captures stimulus-specific activation that exponentially decreases with repeated S_2 presentations during Reward phase trials. The learning rate for this model was set to $\alpha = 0.25$, similar to the reward learning rate.

Functional connectivity analysis

Two methods were used to explore functional connectivity related to decision bias. First, we performed a psychophysiological-interaction (PPI) analysis (*50*). Second, we performed a mediation analysis to explore mediators of decision bias-related functional connectivity, using estimates of trial-by-trial evoked activation (*51*). These analyses focused on the first half of the Reward phase, where we found evidence for significant reactivation.

In the PPI analysis, a seed region in the hippocampus was defined from the GLM contrast of high versus low decision bias during Reward learning (26, -34, -12, 6 mm region of interest; **Figure 4**). The PPI analysis tested for regions showing greater Reward phase functional correlation with the hippocampus (the physiological variable) for high versus low decision bias stimuli (the psychological variable). For the PPI, the timecourse of activation from the hippocampus was extracted and deconvolved. This timecourse was multiplied by the decision bias indicator (0.5 s duration) and then convolved with the hemodynamic response function (HRF). The model included the hippocampus timecourse by decision bias interaction regressor, the decision bias regressor, and the unmodulated hippocampus timecourse regressor (*50*). We focused on potential hippocampal interactions with a striatal region that responded to reward prediction errors (6, 6, 12; 6mm radius sphere; **Figure S3**). As a control, we analyzed data from the Association phase in a similar manner.

Finally, we tested whether activation in visual regions, defined per-subject, mediated decision bias-related connectivity between the hippocampus and

striatum. While the visual cortex pathway lacks direct connectivity with the other regions, functional correlations could arise via intermediate pathways. This analysis was conducted using estimates of trial-evoked activation (*51*). This method is conceptually similar to an approach that estimates separate regressors for each trial in a GLM and combines the resulting estimates to compute between-region correlations (*52*). Stimulus-evoked activation to S_2 stimuli for each trial was estimated by extracting normalized activity from 4-6 s after trial onset, minus a baseline response taken from -2 to 0 s before trial onset. The response timepoint selection was visually verified to overlap with peak trial-evoked responses independent of condition. Each subject's activation values were sorted into high- and low-bias bins.

Regions of interest in the hippocampus and striatum were as in the PPI analysis. For the visual cortex, we extracted activity from subject-specific ROI masks derived from face, scene, and body part Association phase responses as used in the reactivation analysis. As a control, we extracted data from an ROI in the temporal-parietal junction (62, -42, 22) that exhibited a significant response to Reward phase trials overall but which would not be expected to contribute to decision bias.

To test whether visual regions mediate connectivity between the hippocampus and striatum, we conducted a formal mediation analysis (*53*). A multi-level mediation was estimated using a custom Matlab toolbox (Tor Wager, http://wagerlab.colorado.edu/tools). We estimated per-subject path coefficients for the hippocampus to striatum (Path c), hippocampus to visual cortex (Path a), and visual cortex to striatum (Path b). The mediation tests whether the path via the visual cortex accounts for significant covariance in the hippocampus-striatum path. At the second-level, the analysis uses bias-corrected bootstrap significance testing to derive a sensitive measure of path and mediation effects (randomly sampling 100,000 observations within each path at the subject level). This analysis provides path coefficients, standard errors, and two-tailed, uncorrected p-values for each path and the mediation effect.

Association and Decision phases

Where appropriate, parallel GLM and functional connectivity analyses were conducted on BOLD activation from the Association and Decision phases of the experiment to compare activation related to high versus low decision bias.

Supplementary Text

Pre-experiment liking ratings

Stimuli were selected based on individual ratings of 20 stimuli in each category to match median liking across stimulus types. The mean liking for selected S_1 stimuli and S_2 stimuli was almost exactly matched (56.14% ± 1.73 and 56.11% ± 1.66, respectively (mean ± SEM)). Within S_1-S_2 pairings, stimuli were also closely matched in liking: the absolute difference in liking between stimuli in a pair across subjects was near zero (mean participant rating deviation, 1.78%; maximum, 6.75%). Residual initial liking rating differences between S_1 stimuli did not predict subsequent preferences during the Decision phase. Liking ratings for subsequently chosen high-bias stimuli were no different than zero (S_1+ versus S_1-: -0.05% ± 0.61). Liking ratings were numerically lower for low-bias stimuli (S_1+ versus S_1-: -1.66% ± 0.92); such a pre-experiment difference in liking ratings would, if anything, work against a subsequent decision bias.

Association phase

During the Association phase, performance on the target-detection task was near ceiling (2.78% ± 1.01 incorrect responses; 4.70% ± 1.22 missed responses). Reaction times to S_1 or S_2 stimuli did not differ according to later pair memory or decision bias.

Reward phase

During the Reward phase, responses to the reward were rarely missed or in error (misses, 1.44% ± 0.77, errors, 0.92% ± 0.35 of trials). Reaction times to the S_2 stimuli were 778.3 ms ± 29.0. Reaction times to S_2+ and S_2- stimuli did not differ; both showed a decrease over time, but this change only reached significance for S_2+ stimuli (from 796.7 ms ± 30.2 to 759.9 ms ± 30.6; $t_{(27)} = 2.44$, $P < 0.05$). Reaction times to S_2 stimuli did not differ according to later decision bias.

Decision phase

Reward-associated S_2 stimulus (S_2+) preference overall was 79.4% ± 3.9. Broken down by associated S_1 stimulus picture type, mean preference values were 79.5% ± 6.7 for face-associated S_2 stimuli, 81.3% ± 5.5 for scene-associated S_2 stimuli, and 76.8% ± 7.0 for body part-associated S_2 stimuli.

Decision bias overall, as measured by S_1+ versus S_1- preference, was 53.8% ± 5.4 (computed relative to S_2 preference). Broken down by picture category, mean preference values were 54.2% ± 9.4 for face S_1 stimuli (*n.s.*), 42.0% ± 8.9 for scene S_1 stimuli (*n.s.*), and 66.4% ± 8.2 for body S_1 stimuli ($t_{(26)}$ = 1.99, $P < 0.10$). Within-participant, ranking the three stimulus category pair sets by decision bias illustrates variability that was subsequently used to probe neural correlates of bias (86.3% high, 50.6% medium, 20.8% low decision bias; **Figure 2B**). Each category type only allowed choices versus one alternative S_1 stimulus in the Decision phase, and because participants were frequently consistent across repeated choices, decision bias for an S_1 stimulus category was often at 100% or 0%. 68% (19/28) of participants exhibited 100% bias for at least one category and 82% (23/28) of participants exhibited greater than 50% preference for at least one category. Conversely, for at least one category, 75% (21/28) of participants exhibited 0% bias.

Decision phase reaction times were 1228.8 ms ± 4.6 for S_1 choices and 1164.3 ms ± 4.7 for S_2 choices. Reaction times for S_1 and S_2 choices did not significantly differ ($t_{(27)}$ = 1.37,$P > 0.10$). Within-participants, reaction times did not differ between high versus low decision bias S_1 choices.

In the Reward learning and Decision phases of the experiment, participants earned an average of $14.00 ± 1.41.

Post-experiment liking ratings

After the Decision phase, participants rated their current liking for the S_1 and S_2 stimuli. (Liking data for one participant is missing due to an error in data recording.) While this measure may not be fully independent of Decision phase preferences (as choices themselves may influence liking), results from this phase

nevertheless support the preferences exhibited in the Decision phase. Liking for S_2 stimuli showed a strong effect of reward association in the Reward phase (change in S_2+ minus change in S_2-, 15.17% ± 3.63; $t_{(26)} = 4.18$, $P < 0.001$). Liking overall did not increase over the experiment for S_1+ versus S_1- stimuli (1.26 % ± 1.65, *n.s.*). However, liking for S_1 stimuli when sorted by decision bias exhibited a significant difference for high versus low decision bias stimuli (change in S_1+ minus change in S_1-; high = 8.95% ± 3.46; low = -3.26% ± 2.33; $t_{(22)} = 5.08$, $P < 0.01$; **Figure S1A**).

A final measure further supports the Decision phase measure of decision bias. When outside of the scanner, participants were asked to rate, for each S_1 and S_2 stimulus, how likely they thought it would be to be associated with winning money. Participants rated stimuli on a scale anchored by “0% reward” and “100% reward”. Reward likelihood ratings for S_2 stimuli reflected reward associations in the Reward phase. Reward likelihood ratings for S_1 stimuli when sorted by decision bias exhibited a significant difference for high versus low decision bias stimuli (S_1+ minus S_1-; high = 17.15% ± 4.85; low = -11.59% ± 5.53; $t_{(22)} = 4.46$, $P < 0.001$), consistent with the post-experiment liking ratings.

Testing explicit awareness of associations

After leaving the MRI scanner, participants completed a matching probe and a written questionnaire to determine memory and post-experiment awareness for the incidental S_1-S_2 associations from the Association phase of the experiment (see Methods for details). (Data from one participant is missing due to an error in data recording.) Pair memory accuracy for S_1-S_2 associations was not different than chance (53.7% ± 3.25 correct; only one participant exceeded 4/6 correct). Pair memory averaged across picture categories did not correlate with mean decision bias, nor did pair memory within-category correlate with decision bias within-category. When pair memory accuracy was sorted by high- versus low-bias, the difference was not significant (high 59.38% ± 7.3 versus low 45.7% ± 7.5; $t_{(22)} = 1.19$, $n = 23$, $P > 0.2$; **Figure S1B**).

As noted in the main text, exploratory analyses of brain activation related to association memory in the Association and Reward phases revealed no activation in regions of interest to correctly-associated versus incorrectly-associated stimulus pairs, even at a liberal uncorrected threshold of $P < 0.01$.

On the final paper questionnaire, we asked participants if they noticed any regularities or pairings in the first (Association) phase of the experiment. For text of questions, see section at end of supplement. No participants exhibited evidence of awareness of the pairings. Finally, in an additional question, no participants exhibited knowledge of the hypothesis of the study. Combined with participants' chance level of pair memory accuracy, these data indicate that participants' Decision phase choices as well as Reward phase activation in the hippocampus were not driven by explicit awareness of stimulus associations.

Association memory was probed at the conclusion of the fMRI study, and this leaves open the possibility that a mid-experiment memory probe may have revealed higher levels of memory. While we cannot probe memory in the middle of the experiment without drawing attention to the manipulation of interest, we conducted a separate behavioral study to address this question. In this experiment, instead of probing memory for pairings at the end of the experiment after the Reward and Decision phases, pair memory was probed immediately following the Association phase. Significant memory performance and pair association awareness at this point could indicate that participants in the fMRI study also had access to stronger association memories during learning and decision making.

In the supplemental study, the procedures for the pre-experiment stimulus rating, Association phase, and pair memory probe were identical to the procedures of the fMRI study. Immediately after the Association phase, participants (n = 19, 11 female, age range 18-26 years) completed the pair memory probe that instructed participants to match an S_1 stimulus at the top of the screen with one of two S_2 stimuli below. Replicating the finding from the main fMRI experiment as well as those from a prior behavioral study (*13*), pair memory performance was not different from chance (56.1% ± 5.4; $t_{(18)} = 1.13$, $P > 0.20$),

Further, in an extended written questionnaire, no participants reported noticing the regular S_1-S_2 pair presentations. When informed about the pairs and then asked if they noticed the regular pair presentations, the majority of participants responded in the negative (11 responded “no”, 4 responded “a little”, and 4 responded “yes”). Memory performance in the subgroup that reported having noticed the pair presentations was lower than in the full group (41.5%, n = 4). After being informed about the Association phase pairs, participants were asked which type of picture (circle, face, scene, or body part) was always present in a pair. 75% of participants (of 16 who selected an answer) correctly responded that the circle was always present. We then asked participants to identify whether the circle came first or second in the sequence. Only half of those who had responded correctly that the circle was always present answered correctly that the circle came second. Participants’ inability to recognize the temporal position of the common stimulus, immediately subsequent to an Association phase with 10 repetitions of pairings, indicates a low level of awareness for the association structure.

The lack of significant pair memory in the fMRI study and supplemental behavioral study supports the hypothesis that the kind of association learning engaged during the task does not tend to lead to explicit, declarative knowledge of associations. Thus, with respect to mechanisms underlying decision bias, these null memory findings suggest that strategic inferential reasoning about S_1-S_2-reward associations, operating during the Reward phase or the Decision phase, is unlikely to account for shifts in S_1 preferences.

Nevertheless, it is possible that the memory probes we employed did not reveal the full level of subject’s awareness of the pair associations. Some studies have demonstrated that awareness during learning may not be reflected in memory measures (c.f. *54*). Thus, determining the degree of implicitness underlying decision bias is an important area for future study.

Decision bias predictors

During the Reward phase, a whole-brain FWE-corrected analysis of the decision bias contrast did not show any additional regions predictive of bias. In the striatum, we did not find any regions predictive of bias, even at a liberal uncorrected threshold of $P < 0.01$.

Reactivation predictors of decision bias

In the first half of the Reward phase we found significant reactivation of associations for stimuli in pairs that later showed high versus low decision bias, as described in the main text (**Figure 3**). We found that over all categories, reactivation was greater for high versus low decision bias in 74% of participants ($t(22) = 2.55$, $P < 0.05$). The reactivation effect was present numerically across the entire Reward phase (high = 0.121 ± 0.039, low = -0.020 ± 0.081; $t_{(22)} = 1.13$, $P < 0.20$). When limiting our analyses to the second half of the Reward phase a difference was not found (high = -0.006 ± 0.041, low = 0.014 ± 0.044; $t_{(22)} = 0.33$, *n.s.*).

In a follow-up analysis we examined activation to each stimulus category separately (faces, scenes and body parts), comparing individual associations that led to later high versus low decision bias. As this analysis is restricted to a separate test for each category, activation values for high versus low decision bias are necessarily compared between-participants and include a smaller group of participants, both of which lead to weaker statistical power to detect an effect. For face-associated S_2 stimuli, the reactivation measure was greater for high versus low decision bias trials at a trend level (high 0.238 ± 0.110, low -0.034 ± 0.075; $t_{(19)} = 2.08$, $P < 0.06$; $n_{high} = 10$, n $n_{low} = 11$). For scene-associated S_2 stimuli, we detected no difference (high 0.003 ± 0.060, low 0.009 ± 0.065, *n.s.*; $n_{high} = 7$, $n_{low} = 14$). For body-associated S_2 stimuli, the reactivation measure was numerically greater for high versus low decision bias (high 0.056 ± 0.043, low -0.060 ± 0.0132; $t_{(17)} = 1.08$, $P > 0.10$; $n_{high} = 13$, $n_{low} = 6$), while in

As a control, we tested whether reactivation was due to generally higher levels of activity across visual regions of interest rather than in specific visual

regions reflecting S_2 pairing. Mean activation across the three sets of participant-specific masks did not differ for high versus low decision bias Reward phase trials ($P > 0.45$). Also, we verified that the reactivation result was found if category-specific region of interest activation was normalized by activation across all visual regions of interest. This measure exhibited much higher variance (for example, activation to actual visual stimuli in the Association phase was only marginally significant), but nevertheless, the reactivation effect was replicated (69% of participants exhibited greater reactivation for high versus low decision bias).

Finally, as noted in the main text, as a control we also tested whether activation in the Association phase predicted decision bias. While we found activation overall during the Association phase in the hippocampus and surrounding medial temporal lobe (**Figure S2**, masked for the MTL), we did not find significant correlates of decision bias in the MTL. Similarly, during the Association phase, activation to the presentation of S_1 face, scene, and body stimuli did not differ by later bias ($t_{(22)} = -0.78$, *n.s.*). A parallel analysis of reactivation in the Decision phase is not possible as choices present actual face, scene, and body part stimuli.

Reward learning activation

In the Reward phase reward learning GLM, we examined brain activation related to learning associations between S_2 stimuli and reward and neutral outcomes. The reward learning signals we examined were derived from a simple temporal difference model of learning (e.g., *45, 55*). At S_2 stimulus presentation we tested correlates of learned stimulus value, the reward prediction error elicited by stimulus presentation; at reward presentation we tested correlates of reward prediction error, focusing on the striatum (*45*).

Activation in the bilateral striatum significantly correlated with the predicted value of presented stimuli (**Figure S3A**). At reward presentation, a region of the right caudate correlated with reward prediction error (**Figure 3B**; 90 voxels at $P < 0.005$, uncorrected). While this activation did not survive small-volume correction

in the striatal region of interest (P < 0.12), we focus on this region in later analyses given its proximity to the striatal region correlated with value. Additionally, a region of the right ventral putamen significantly correlated with reward prediction error (**Figure S3B**).

In an exploratory analysis, we also found that activation in the posterior hippocampus correlated with reward prediction error (20, -28, -6; z = 3.86, p < 0.05 SVC, correcting for the additional ROI analysis).

Reward phase association learning activation

In this analysis, we found a cluster of activation in the ventral tegmental area/substantia nigra pars compacta (VTA/SNc) that significantly correlated with the stimulus prediction error regressor (z = 3.40 (4, -18, -12)). This cluster overlaps with a midbrain activation reported in a previous study of learning in a related paradigm (*31*). Clusters of activation in the left and right hippocampus were significant at P < 0.001 uncorrected but did not survive small-volume correction. This result was specific to the Reward phase and was not due to generic decreases in activity across the phase.

Functional connectivity analysis

Results of the Reward phase PPI analysis of hippocampal-striatal functional connectivity for high- versus low-bias stimuli are reported in the main text and **Figure 4**. A control analysis of hippocampal connectivity during the Association phase revealed no interaction between decision bias and hippocampal connectivity anywhere in the brain, even at a liberal uncorrected threshold of P < 0.01.

A summary of the Reward phase mediation result is reported in the main text. This analysis tested whether a pathway from hippocampus to striatum via participant-specific visual cortical regions mediated hippocampal-striatal connectivity. We report mediation results from the full group of participants with high-bias stimuli, for the first half of the Reward phase. While this analysis is different from a PPI analysis and uses trial-by-trial estimates of evoked

responses, we found that hippocampal-striatal connectivity replicated that found using PPI. More complete statistics on the paths between regions are as follows: Path a) hippocampus to visual cortex (0.426 ± 0.086, $P < 0.001$); Path b) visual cortex to caudate (0.050 ± 0.015, $P < 0.001$); Path c) hippocampus to caudate (0.214 ± 0.054, $P < 0.001$). For low-bias stimuli, functional connectivity between the hippocampus and visual cortex was not significant and we found no evidence for mediation.

To verify the specificity of the functional connectivity result between the hippocampus and the caudate in predicting decision bias in the Reward phase, we conducted several control analyses, as mentioned in the main text. First, a control analysis tested connectivity between regions of interest and a task-activated region of the temporal-parietal junction during the Reward phase that was not expected to show differential connectivity in the temporal-parietal junction. This analysis revealed that connectivity between the hippocampus and the control ROI did not differ between trials that led to high versus low decision bias.

As another control, we analyzed Association phase connectivity between regions of interest for high- and low-bias stimuli. We found no significant connectivity during trials that led to high bias for either the hippocampus to visual cortex pathway or the pathway between visual cortex and striatum. These results confirm the selective role of functional connectivity during the Reward phase between the hippocampus, visual cortex, and caudate in predicting subsequent decision bias.

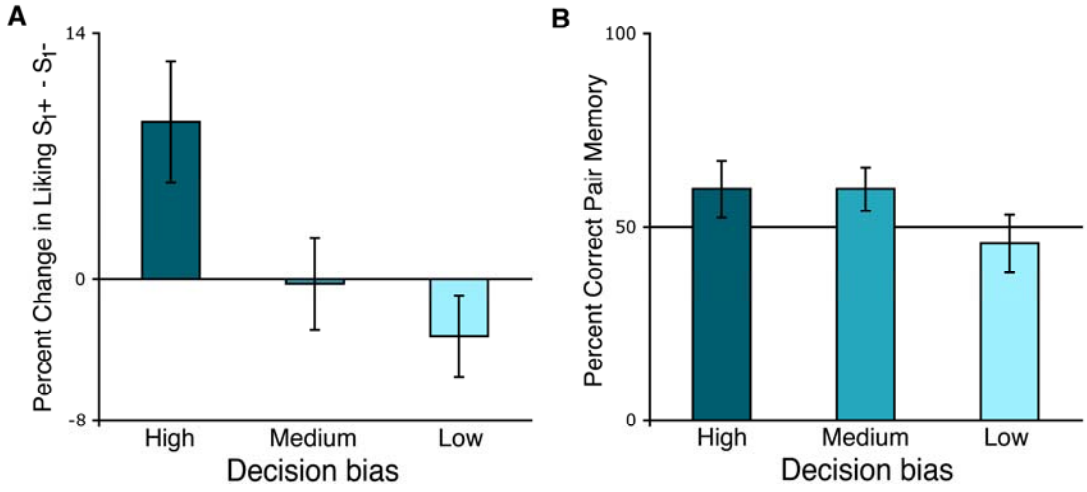

Fig. S1.

Liking ratings and pair association memory. **A**, Change in liking ratings over the experiment for the three stimulus categories, sorted by participants' level of decision bias. **B**, Post-experiment pair association memory accuracy (S_1-S_2 associations) sorted by participants' level of decision bias.

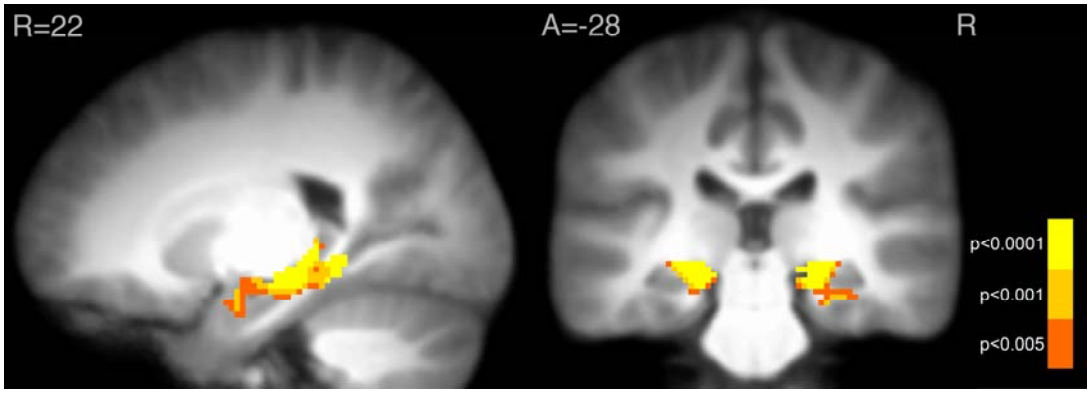

Fig. S2

Association phase activation in the hippocampus and medial temporal lobe. Right hippocampus, z = 5.67 (22, -28, -6); left hippocampus, z = 5.32 (-22, -26, -6). (All fMRI results P <0.05 SVC for FWE unless otherwise noted, masked for regions of interest; images thresholded at P <0.005, uncorrected for display.)

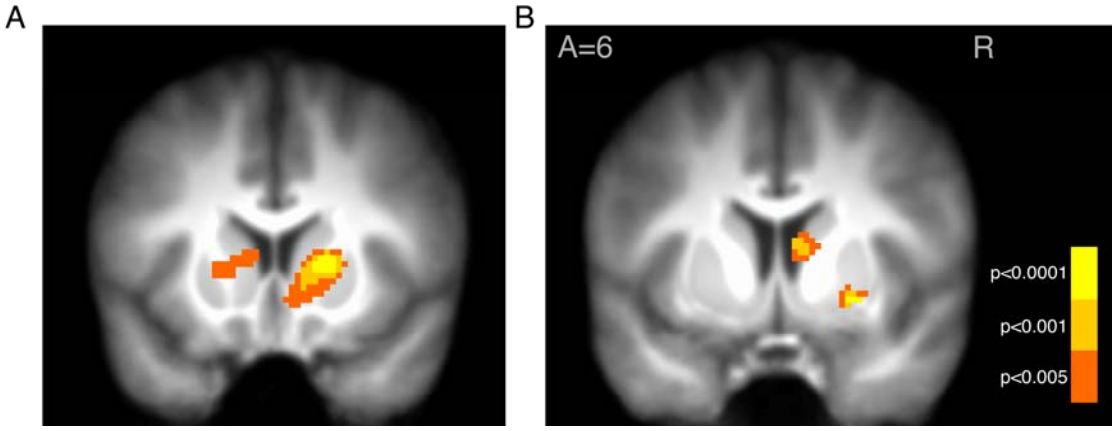

Fig. S3

Reward learning-related activation during the Reward phase. **A**, Activation in the striatum correlated with predicted value. Right striatum peak: caudate-putamen, z = 4.31 (18, 12, 4); left striatum peak: putamen, z = 3.91 (-18, 0, 16). **B**, Activation in the dorsal caudate correlated with reward prediction errors (z = 3.50, P < 0.005 unc., 90 voxels). Right ventral putamen, z = 3.86 (26, 6, -8).

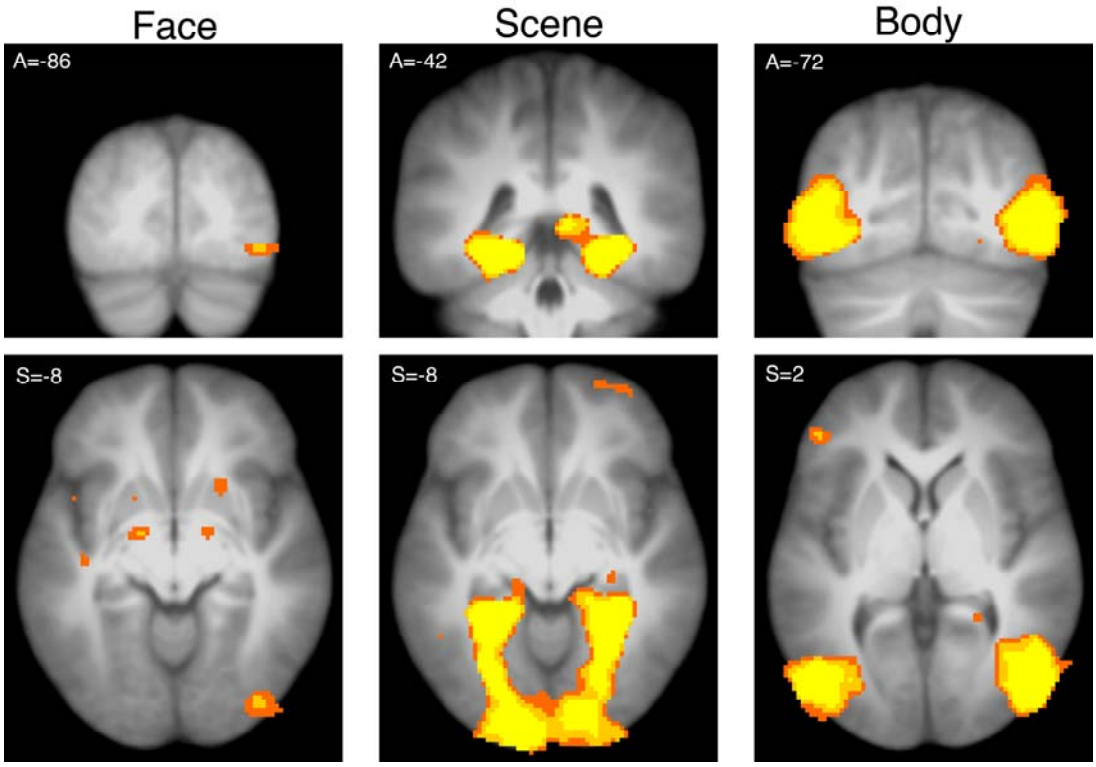

Fig. S4

Activation in the Association phase to face, scene, and body part S_1 stimuli ($P < 0.001$, uncorrected).

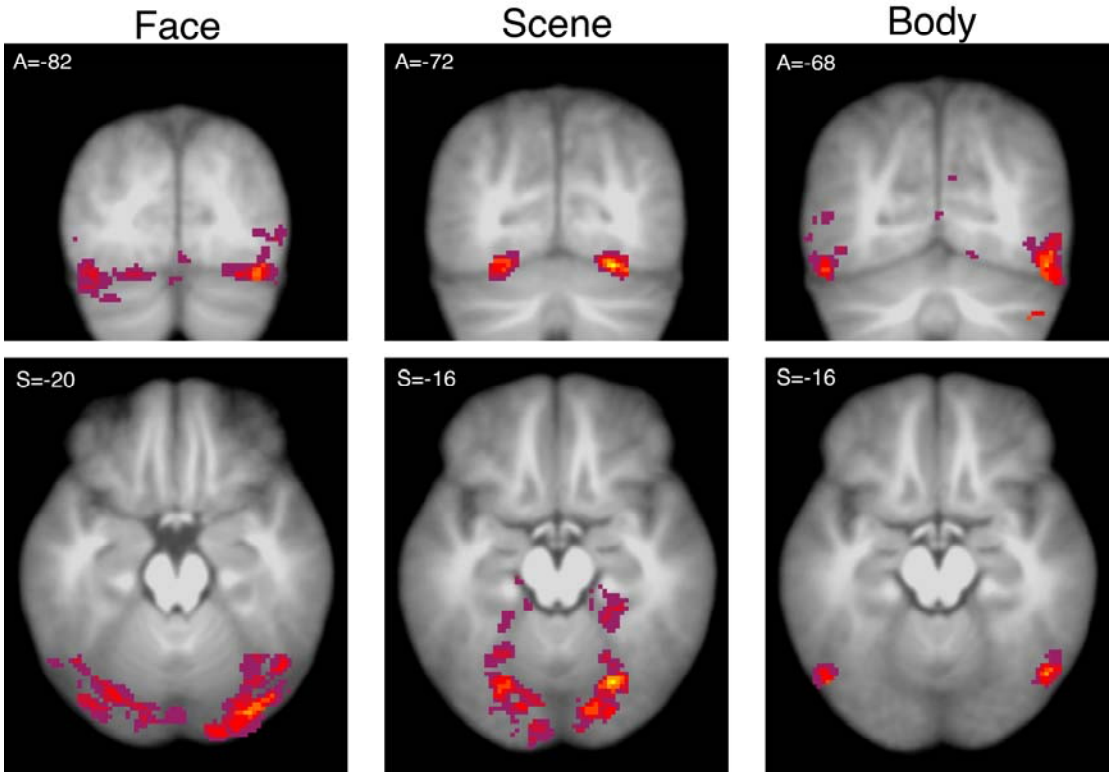

Fig. S5

Sum of participants' category-selective masks that were derived from S_1 stimulus presentation in the Association phase and then used in the Reward phase reactivation analysis. Individual masks represent the peak 1% of voxels falling within group Association phase-derived mask (see Methods). Voxel color represents the number of participants with overlapping masks (of n = 23 participants); maximum n = 8, yellow; minimum n = 1, magenta.

Supplementary post-experiment questionnaire text

Post-experiment association memory test instructions:

“Please try to pair the top picture with the bottom picture that seems more related to the top picture in some way. Follow a hunch or guess if you are not sure.”

Post-experiment written questionnaire examples:

“During the part in the scanner where you chose between two pictures, what strategy (or strategies) did you use when choosing between the two presented scene, face, or body pictures?”

“Did you notice any regularities, sequences, or pairs during the very first part of the task when you were responding to target upside-down pictures (faces, scenes, body pictures), and if so, what?”

Post-experiment questionnaire examples – supplemental behavioral study:

“During the target detection task, pairs of pictures were always presented sequentially together. Did you notice this?”

“In these pairs, one category of picture was always part of the pair. Which category was this?

[Circle] – [Face] – [Scene] – [Body] – [I don’t know]

Was this category of picture always presented first or second?”

References and Notes

1. W. Schultz, Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* **57**, 87 (2006). doi:10.1146/annurev.psych.56.091103.070229 Medline

2. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545 (2008). doi:10.1038/nrn2357 Medline

3. E. C. Tolman, Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189 (1948). doi:10.1037/h0061626 Medline

4. W. J. Brogden, Sensory pre-conditioning. *J. Exp. Psychol.* **25**, 323 (1939). doi:10.1037/h0058944 Medline

5. H. Eichenbaum, A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41 (2000). doi:10.1038/35036213 Medline

6. B. P. Staresina, L. Davachi, Mind the gap: Binding experiences across space and time in the human hippocampus. *Neuron* **63**, 267 (2009). doi:10.1016/j.neuron.2009.06.024 Medline

7. H. Eichenbaum, N. J. Cohen, *From Conditioning to Conscious Recollection: Memory Systems of the Brain* (Oxford Univ. Press, New York, 2001).

8. D. E. Hannula, C. Ranganath, The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron* **63**, 592 (2009). doi:10.1016/j.neuron.2009.08.025 Medline

9. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963 (2005). doi:10.1126/science.1117645 Medline

10. D. J. Foster, M. A. Wilson, Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680 (2006). doi:10.1038/nature04587 Medline

11. B. A. Kuhl, J. Rissman, M. M. Chun, A. D. Wagner, Fidelity of neural reactivation reveals competition between memories. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5903 (2011). doi:10.1073/pnas.1016939108 Medline

12. B. A. Kuhl, A. T. Shah, S. DuBrow, A. D. Wagner, Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501 (2010). doi:10.1038/nn.2498 Medline

13. G. E. Wimmer, D. Shohamy, in *Decision Making, Affect, and Learning: Attention and Performance XXIII*, M. R. Delgado, E. A. Phelps, T. W. Robbins, Eds. (Oxford Univ. Press, Oxford, 2011), pp. 281–309.

14. H. D. Kimmel, Notes from “Pavlov’s Wednesdays”: Sensory preconditioning. *Am. J. Psychol.* **90**, 319 (1977). doi:10.2307/1422055 Medline

15. E. Walther, Guilty by mere association: evaluative conditioning and the spreading attitude effect. *J. Pers. Soc. Psychol.* **82**, 919 (2002). doi:10.1037/0022-3514.82.6.919 Medline

16. N. B. Turk-Browne, B. J. Scholl, M. M. Chun, M. K. Johnson, Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* **21**, 1934 (2009). doi:10.1162/jocn.2009.21131 Medline

17. A. C. Schapiro, L. V. Kustner, N. B. Turk-Browne, Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622 (2012). doi:10.1016/j.cub.2012.06.056 Medline

18. D. Shohamy, R. A. Adcock, Dopamine and adaptive memory. *Trends Cogn. Sci.* **14**, 464 (2010). doi:10.1016/j.tics.2010.08.002 Medline

19. C. M. Pennartz, R. Ito, P. F. Verschure, F. P. Battaglia, T. W. Robbins, The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends Neurosci.* **34**, 548 (2011). doi:10.1016/j.tins.2011.08.001 Medline

20. Materials and methods are available as supplementary materials on *Science* Online.

21. J. De Houwer, A. Moors, in *Implicit Measures of Attitudes: Procedures and Controversies*, B. Wittenbrink, N. Schwarz, Eds. (Guilford Press, New York, 2007), pp. 179–194.

22. L. Reddy, N. Kanwisher, Coding of visual objects in the ventral stream. *Curr. Opin. Neurobiol.* **16**, 408 (2006). doi:10.1016/j.conb.2006.06.004 Medline

23. D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux; New York; 2011).

24. C. K. Morewedge, D. Kahneman, Associative processes in intuitive judgment. *Trends Cogn. Sci.* **14**, 435 (2010). doi:10.1016/j.tics.2010.07.004 Medline

25. E. U. Weber *et al.*, Asymmetric discounting in intertemporal choice: A query-theory account. *Psychol. Sci.* **18**, 516 (2007). doi:10.1111/j.1467-9280.2007.01932.x Medline

26. T. Gilovich, D. Griffin, D. Kahneman, Eds., *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge Univ. Press, New York, 2002).

27. D. Kahneman, A perspective on judgment and choice: Mapping bounded rationality. *Am. Psychol.* **58**, 697 (2003). doi:10.1037/0003-066X.58.9.697 Medline

28. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction. *Science* **331**, 1279 (2011). doi:10.1126/science.1192788 Medline

29. A. R. Preston, Y. Shrager, N. M. Dudukovic, J. D. Gabrieli, Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* **14**, 148 (2004). doi:10.1002/hipo.20009 Medline

30. A. Johnson, A. D. Redish, Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176 (2007). doi:10.1523/JNEUROSCI.3761-07.2007 Medline

31. D. Shohamy, A. D. Wagner, Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378 (2008). doi:10.1016/j.neuron.2008.09.023 Medline

32. M. A. Gluck, C. E. Myers, Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus* **3**, 491 (1993). doi:10.1002/hipo.450030410 Medline

33. N. D. Daw, Y. Niv, P. Dayan, Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704 (2005). doi:10.1038/nn1560 Medline

34. S. E. Hyman, R. C. Malenka, E. J. Nestler, Neural mechanisms of addiction: The role of reward-related learning and memory. *Annu. Rev. Neurosci.* **29**, 565 (2006). doi:10.1146/annurev.neuro.29.051605.113009 Medline

35. W. J. Brogden, Sensory pre-conditioning of human subjects. *J. Exp. Psychol.* **37**, 527 (1947). doi:10.1037/h0058465 Medline

36. R. L. Port, A. L. Beggs, M. M. Patterson, Hippocampal substrate of sensory associations. *Physiol. Behav.* **39**, 643 (1987). doi:10.1016/0031-9384(87)90167-3 Medline

37. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433 (1997). doi:10.1163/156856897X00357 Medline

38. R. W. Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162 (1996). doi:10.1006/cbmr.1996.0014 Medline

39. K. J. Friston, K. J. Worsley, S. J. Frackowiak, J. C. Mazziotta, A. C. Evans, Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1**, 210 (1993). doi:10.1002/hbm.460010306

40. N. Tzourio-Mazoyer *et al*., Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273 (2002). doi:10.1006/nimg.2001.0978 Medline

41. P. E. Downing, Y. Jiang, M. Shuman, N. Kanwisher, A cortical area selective for visual processing of the human body. *Science* **293**, 2470 (2001). doi:10.1126/science.1063414 Medline

42. R. Epstein, N. Kanwisher, A cortical representation of the local visual environment. *Nature* **392**, 598 (1998). doi:10.1038/33402 Medline

43. N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302 (1997). Medline

44. T. A. Hare, J. O'Doherty, C. F. Camerer, W. Schultz, A. Rangel, Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* **28**, 5623 (2008). doi:10.1523/JNEUROSCI.1309-08.2008 Medline

45. J. O'Doherty *et al*., Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452 (2004). doi:10.1126/science.1094285 Medline

46. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876 (2006). doi:10.1038/nature04766 Medline

47. D. Kumaran, E. A. Maguire, Match mismatch processes underlie human hippocampal responses to associative novelty. *J. Neurosci.* **27**, 8517 (2007). doi:10.1523/JNEUROSCI.1677-07.2007 Medline

48. J. E. Lisman, A. A. Grace, The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron* **46**, 703 (2005). doi:10.1016/j.neuron.2005.05.002 Medline

49. D. Kumaran, E. A. Maguire, An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLoS Biol.* **4**, e424 (2006). doi:10.1371/journal.pbio.0040424 Medline

50. K. J. Friston *et al*., Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218 (1997). doi:10.1006/nimg.1997.0291 Medline

51. R. A. Adcock, A. Thangavel, S. Whitfield-Gabrieli, B. Knutson, J. D. Gabrieli, Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron* **50**, 507 (2006). doi:10.1016/j.neuron.2006.03.036 Medline

52. J. Rissman, A. Gazzaley, M. D'Esposito, Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage* **23**, 752 (2004). doi:10.1016/j.neuroimage.2004.06.035 Medline

53. T. D. Wager *et al*., Brain mediators of cardiovascular responses to social threat: Part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *Neuroimage* **47**, 821 (2009). doi:10.1016/j.neuroimage.2009.05.043 Medline

54. B. Gawronski, E. Walther, What do memory data tell us about the role of contingency awareness in evaluative conditioning? *J. Exp. Soc. Psychol.* **48**, 617 (2012). doi:10.1016/j.jesp.2012.01.002