

Signed Reward Prediction Errors in the Ventral Striatum Drive Episodic Memory

 Cristian B. Calderon,* Esther De Loof,* Kate Ergo, Anna Snoeck, Carsten N. Boehler, and  Tom Verguts

Department of Experimental Psychology, Ghent University, B-9000 Ghent, Belgium

Recent behavioral evidence implicates reward prediction errors (RPEs) as a key factor in the acquisition of episodic memory. Yet, important neural predictions related to the role of RPEs in episodic memory acquisition remain to be tested. Humans (both sexes) performed a novel variable-choice task where we experimentally manipulated RPEs and found support for key neural predictions with fMRI. Our results show that in line with previous behavioral observations, episodic memory accuracy increases with the magnitude of signed (i.e., better/worse-than-expected) RPEs (SRPEs). Neurally, we observe that SRPEs are encoded in the ventral striatum (VS). Crucially, we demonstrate through mediation analysis that activation in the VS mediates the experimental manipulation of SRPEs on episodic memory accuracy. In particular, SRPE-based responses in the VS (during learning) predict the strength of subsequent episodic memory (during recollection). Furthermore, functional connectivity between task-relevant processing areas (i.e., face-selective areas) and hippocampus and ventral striatum increased as a function of RPE value (during learning), suggesting a central role of these areas in episodic memory formation. Our results consolidate reinforcement learning theory and striatal RPEs as key factors subtending the formation of episodic memory.

Key words: episodic memory; fMRI; reward prediction error; ventral striatum

Significance Statement

Recent behavioral research has shown that reward prediction errors (RPEs), a key concept of reinforcement learning theory, are crucial to the formation of episodic memories. In this study, we reveal the neural underpinnings of this process. Using fMRI, we show that signed RPEs (SRPEs) are encoded in the ventral striatum (VS), and crucially, that SRPE VS activity is responsible for the subsequent recollection accuracy of one-shot learned episodic memory associations.

Introduction

When meeting a new person, being able to remember his/her name from a single encounter is essential. Referred to as episodic memory (Tulving, 1993), this information can, for instance, be recalled to start a conversation when running into that person later on.

Several studies investigated the mechanisms by which such one-shot episodic memories are formed. In particular, previous work identified an important role of reward. Compared with unrewarded contexts, items memorized within rewarding contexts are associated with better recognition performance in old-new item decisions (Shneyer and Mendelsohn, 2018). Neurally,

this beneficial effect of reward on episodic memory is ascribed to the following increased activity of the striatum: rewarded to-be-remembered items eliciting stronger striatal activation are more likely to be subsequently remembered (Wittmann et al., 2005; Adcock et al., 2006; Miendlarzewska et al., 2016).

However, learning based solely on reward has limited computational power. Reinforcement learning (RL) theory instead highlights the importance of reward prediction errors (RPEs) for learning. RPEs capture to what extent choice outcomes deviate from their predictions, thereby giving them a computational advantage. Specifically, RPEs can be used for learning parameter settings based on the computational principle of gradient descent (in a way that a pure reward signal cannot), allowing for more efficient learning (Sutton and Barto, 2018).

In procedural learning, the role of RPEs is well established (Schultz et al., 1997). The idea that RPEs also play a role in episodic memory has recently gained empirical support (for review, see Ergo et al., 2020). Behaviorally, memory encoding improves linearly with better-than-expected rewards and decreases linearly with worse-than-expected rewards, called signed RPEs (SRPEs; De Loof et al., 2018; Jang et al., 2019). Neurally, high- β (20–30 Hz) and high- α (10–17 Hz) oscillatory SRPE signatures mirror behavioral SRPE effects (Ergo et al., 2019). However, memory also improves when agents experience different-than-expected

Received July 10, 2020; revised Dec. 4, 2020; accepted Dec. 7, 2020.

Author contributions: C.B.C., E.D.L., K.E., and T.V. designed research; C.B.C. and A.S. performed research; C.B.C. and E.D.L. analyzed data; C.B.C., E.D.L., K.E., A.S., C.N.B., and T.V. wrote the paper.

The current work was supported by Research Foundation—Flanders (FWO)/Fonds National de la Recherche Scientifique EOS Grant G0F3818N. C.B.C. is supported by FWO Grant 1207719N. K.E. is supported by FWO Grant 1153420N. E.D.L. and T.V. were supported by FWO Grant BOF17-GOA-004. We thank Carlos González-García, Clay Holroyd, and Jacqueline Janowich for helpful comments.

*C.B.C. and E.D.L. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Cristian B. Calderon at cristian.buccalderon@ugent.be.

<https://doi.org/10.1523/JNEUROSCI.1785-20.2020>

Copyright © 2021 the authors

rewards and scales with unsigned RPEs (URPEs; Fazio and Marsh, 2009; Metcalfe, 2017; Rouhani et al., 2018; Stanek et al., 2019).

Although significant evidence points toward an RPE-based acquisition of episodic memory, key neural RPE predictions remain to be tested. Here, we test these predictions using a novel face–word memory association task. This task builds on earlier variable-choice tasks (De Loof et al., 2018; Ergo et al., 2019), where humans learned to associate faces with unknown words, while SRPE size was manipulated by having different numbers of choices available.

First, based on work showing an effect of SRPE on word–word memory association (e. × g., De Loof et al., 2018), we hypothesized that this SRPE effect would generalize to a novel face–word memory association task: Face–word memory would increase with SRPE values (prediction 1).

Second, RPEs are implemented via dopaminergic activity bursts, stemming from midbrain nuclei (i.e., ventral tegmental area and substantia nigra), and predominantly broadcast to ventral striatum (VS) and other cortical/subcortical areas (Watabe-Uchida et al., 2017). Indeed, earlier work already established a prediction error (PE) effect in VS during declarative learning (Tricomi and Fiez, 2008). Thus, following prediction 1, we hypothesized that this behavioral pattern would be mirrored in VS activity [Pine et al., 2018; Ripollés et al., 2018; i.e., VS activity would increase as a function of SRPE values (prediction 2)].

Third, crucially, if RPEs drive episodic memory, then following prediction 2, striatal activity should mediate the relation between experimentally manipulated SRPE and memory performance on a within-subject basis (prediction 3).

Fourth, previous research has shown that the behavioral SRPE effect on episodic memory correlates with functional connectivity between hippocampus and striatum (Davidow et al., 2016). SRPE may thus modulate the functional connectivity between stimulus-processing areas and VS and hippocampus. Specifically, connectivity strength would increase with SRPE values (prediction 4).

We addressed these RPE predictions on episodic memory encoding using fMRI.

Materials and Methods

In this section, we provide all technical, methodological, and analytical details. We start by describing participant demographics, followed by a full description of the experimental design. We then explain behavioral and fMRI data acquisition and analyses. All tasks and analysis codes, as well as the face, word, and house stimuli used in our experiment are available on OSF (Open Science Framework; https://osf.io/6vkwmm/?view_only=8b1364c4cb6b41d4b0e96ad615827d33).

Participants

Thirty right-handed participants (26 females: mean age = 22 years, SD = 6.62; 4 males: mean age = 30 years, SD = 11.15) with normal vision participated in this study approved by the local ethics committee (Ghent University Hospital, Ghent, Belgium). Participants received monetary compensation (30 €) and provided written informed consent before the experiment. Our sample size was motivated by previous studies in our laboratory, which robustly showed the sought after behavioral effect using $N = 20$ (for immediate testing condition, see De Loof et al., 2018; Ergo et al., 2019).

Experimental design

Participants underwent four tasks (total duration, ~80 min) in the following order: celebrity knowledge task (outside the scanner, ~15 min); variable-choice and functional localizer tasks (inside the scanner,

~50 min); and memory test (outside the scanner, ~15 min; Fig. 1A). All tasks were programmed with PsychoPy2.

Celebrity knowledge task. In the celebrity knowledge task, participants were shown a celebrity face alongside four potential celebrity names. Participants had to select the correct celebrity name by pressing the keys “d,” “c,” “n,” or “j” to, respectively, select the name in the top left, bottom left, bottom right, or top right corner of the display (Fig. 1B, top). Following their choice, participants gave a certainty rating on a 4-point scale ranging from “completely unsure,” “rather unsure,” and “rather sure” to “completely sure.” They gave their answer using the same keys as above. Participants had no time restriction on these choices. The celebrity knowledge task was performed in one block of 140 trials. Once completed, we selected a subset of 70 accurately recognized faces, which had been rated as completely sure. When participants did not reach that level of accuracy for 70 faces, a random draw of faces was selected to complete the subset of 70 trials. This subset of 70 faces composed the stimulus set subsequently used in the variable-choice task for that subject (see below). On top of this, we randomly selected 6 of the remaining faces for the variable-choice task training, and a set of 60 of the remaining faces (not used in the variable-choice task) for the functional localizer task. This task was run on a Dell Latitude E5550 laptop.

Note that in piloting work, the pairing between a face and a house could not be learned because the encoding failed at the earlier stage of differentiating individual faces and houses in a large novel stimulus set. Thus, in order not to overload the learning capabilities of the participants (i.e., how many stimuli they could tell apart after a single exposure), we opted for a design using Swahili words (in line with our previous research) and famous faces that participants should be more familiar with. The current stimulus material was sufficient not to overload the learning capacity, and participants completed the task successfully.

Variable-choice task. Participants underwent the variable-choice task in the scanner (Fig. 1C). They first observed a fixation cross (0.5 s), followed by the presentation of a celebrity face on top of the screen together with four village pseudo-names. After exploring the display for 4 s, one, two, or four names were framed. The participant’s task was to guess which village name was associated with the celebrity face. Participants were constrained to choose between the framed names. By varying the number of framed village names, we were able to manipulate the SRPE on each trial. We can compute the SRPE as $r - p$, where r is the observed reward (1 and 0 for correct and incorrect guesses, respectively), and p is the probability of making a correct guess. This probability is 1, 0.5, or 0.25, respectively, for the one-, two-, or four-frame conditions. Hence, SRPE could take on the values -0.5 , -0.25 , 0 , 0.5 , and 0.75 . Participants made their choice (no time restrictions) using Cedrus Lumina LS-Pair MRI compatible ergonomic response pads. To select the top left, bottom left, bottom right, or top right option, participants pressed, respectively, with their left middle finger, left index, right index, or right middle finger. Following their choice, a fixation cross was shown for a jittered amount of time, drawn from a Poisson distribution (mean = 4 s) truncated between 1 and 7 s. Subsequently, during the choice feedback (5 s), participants were shown the celebrity face associated with the correct village name, either framed in green (if they guessed correctly) or in red (otherwise). Choice feedback was followed by another jittered fixation cross, drawn from a Poisson distribution (mean = 4 s) truncated between 1 and 4 s. Participants were instructed to use the 5 s that the correct word pair is on the screen to remember these pairs for the follow-up memory test. The trial ended by a monetary update (2 s) indicating that the money earned for that trial (0.70 € for correct guesses and 0 € otherwise), as well as a total tally. As an incentive to memorize the face–name associations, participants were informed that they could receive an additional monetary reward based on their performance on the memory test at the end of the experiment, if they had the best memory score of all participants. Both fixation cross jitters were optimized to maximize design efficiency using the DesignDiagnostics toolbox (<https://montilab.psych.ucla.edu/fmri-wiki/>).

The variable-choice task was performed in one block of 70 trials. Figure 1D displays the distribution of trials in each condition with the associated SRPE values. The one-frame, two-frame, and four-frame conditions comprised 10, 20, and 40 trials, respectively, of which 10 trials (in each case) led to a correct guess (i.e., rewarded choice).

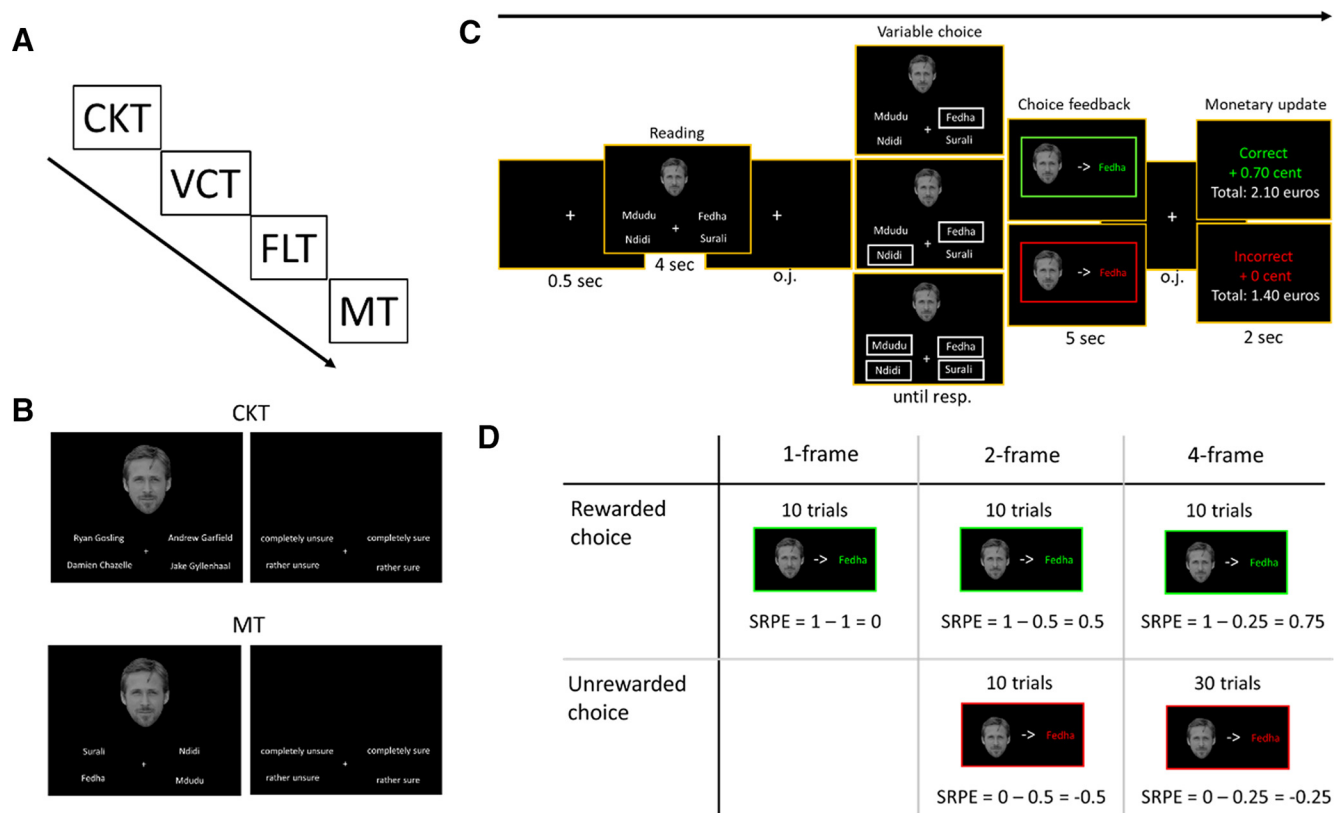


Figure 1. Experimental design overview. **A**, Schedule. Participants underwent four tasks sequentially: the celebrity knowledge task (CKT), the variable-choice task (VCT), the functional localizer task (FLT), and the memory task (MT). The VCT and FLT were performed in the scanner. **B**, CKT and MT. Participants were first shown a celebrity face alongside four competing celebrity (CKT) or village (MT) names. No time limit was imposed to respond with the keyboard (keys d, c, n, and j, respectively, for top left, bottom left, bottom right, and top right names). Participants then (using the same keys) indicated their choice certainty. **C**, VCT. Following a fixation cross (0.5 s), participants saw a face alongside four words (4 s). Subsequently, one, two, or four names were framed, indicating the options participants could choose from. Following an optimized jittered (o.j.) interval with just a fixation cross, participants were shown the to-be-learned face–word association framed in green/red for correct/incorrect choices; this choice feedback evoked an SRPE. Subsequently, they were shown a monetary update. **D**, Number of trials and associated SRPE for each condition.

Before the variable-choice task, participants underwent six training trials outside the scanner; two of each condition (i.e., one, two, and four frames) leading to correct and incorrect guesses (except for the one-frame condition always leading to correct guesses). Therefore, participants had experienced all the SRPEs before performing the task inside the scanner. The training phase was run on a Dell Latitude E5550 laptop, and participants gave their response with the d, c, n, or j key (as in the celebrity knowledge task).

Note that to investigate the effect of reward and prediction errors on declarative learning, most earlier studies combined two learning tasks (Wimmer et al., 2014; Davidow et al., 2016). One is an incremental learning task to induce reward and prediction errors; the other is the declarative memory task, in which the effect of the rewards and prediction errors generated during the incremental learning task are measured.

Our design allows us to induce reward prediction errors that are as clearly interpretable as possible. Instead of an incremental learning inducer, in the variable-choice task participants can explicitly calculate RPEs on each trial. Here, the inducer (choosing a word) is almost trivial, but this is intentional: it allows us to calculate reward prediction errors based on simple probability theory. Earlier behavioral work demonstrated that this paradigm generates very robust prediction error effects on the subsequent declarative memory diagnostic task (De Loof et al., 2018; Ergo et al., 2019).

Functional localizer task. Immediately following the variable-choice task, participants underwent the functional localizer task. In this task, participants alternately observed a centrally presented celebrity face or a house (1.5 s) and a fixation cross (0.5 s). The task of the participants was to respond with the right index finger, whenever the presentation of a face/house would repeat (i.e., one-back task). Note that contrasting

blocks of a one-back task on face versus house stimuli has been proven efficient to functionally reveal fusiform face area activation (Berman et al., 2010). House pictures were taken from earlier work (Schiffer et al., 2014), and faces were randomly selected from the subset of 60 faces that was not presented during the variable-choice task. Participants performed 16 blocks (8 blocks with faces and 8 blocks with houses, in random order) of 18 trials. Each stimulus had a 0.2 repetition probability (but could not repeat twice in a row).

Memory test. Upon completion of the functional localizer task, participants performed the memory test (Fig. 1B, bottom). During this test, participants again observed the 70 faces alongside the same four competing village names (shuffled relative to their previous positions on the display during the variable-choice task). In order to make the time between learning and testing similar for all face–word pairs, we chunked the 70 trials presented during the variable-choice task in seven chunks of 10 trials; the first 10 trials forming chunk one, the following 10 trials forming chunk two, and so on. The trials of each chunk were randomly shuffled, and the chunks were represented in sequential order (from chunk one to seven). As in the celebrity knowledge task, participants pressed keys d, c, n, or j to, respectively, select the village names at the top left, bottom right, or top right, and they were instructed to provide a certainty rating on their choice after each trial. We imposed no time restrictions on either task.

Behavioral data analysis

To test the behavioral effect of SRPE on recognition performance (prediction 1), we applied a (generalized in case of binary outcomes, using a logit link function) linear mixed-effects model with a random intercept per subject and centered predictors. These purely behavioral analyses

were performed on trial-level data. For all analyses (including the mediation analyses reported below), none of the centered predictors had a significant random slope, nor did the random slopes interact with the fixed effects; these were hence dropped from the reported models. We report the χ^2 statistics from the ANOVA type III tests. All analyses were performed in R (RRID:SCR_000432).

fMRI data acquisition and analysis

fMRI data acquisition. fMRI data were collected with a 3 T Magnetom Trio MRI scanner system (Siemens Medical Systems), with a 32-channel radio-frequency head coil. A 3D high-resolution anatomic image was obtained using a T1-weighted MPRAGE sequence [TR = 2200 ms; TE = 2.51 ms; TI = 900 ms; flip angle (FA) = 8°; FOV = 250 × 250 mm; matrix size = 176 × 256 × 256; interpolated voxel size = 0.9 × 0.9 × 0.9 mm]. In addition, we acquired a field map per participant, to correct for magnetic field inhomogeneities (TR = 520 ms; TE1 = 4.92 ms; TE2 = 7.38 ms; image matrix = 80 × 80 × 50; FOV = 192 mm; FA = 60°; slice thickness = 2.5 mm; voxel size = 2.5 × 2.5 × 2.5 mm; distance factor = 20%). Whole-brain functional images were acquired using a T2*-weighted EPI sequence (TR = 1730 ms; TE = 30 ms; FA = 66°; multi-band acceleration factor = 2; matrix size = 50 × 84 × 84; voxel size = 2.5 × 2.5 × 2.5 mm).

fMRI data analysis. SPM 12 (<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>; RRID:SCR_007037) or MATLAB (R2016b; RRID:SCR_001622) was used for preprocessing and fMRI data analysis. For each participant, the first five volumes were discarded to avoid transient spin saturation effects; and BIDS (Brain Imaging Data Structure; <https://bids.neuroimaging.io/>)-formatted raw data were defaced, realigned, and unwrapped, slice time corrected, normalized, and smoothed using a Gaussian kernel of 8 mm full-width at half-maximum. We performed coregistration of the structural image with the mean realigned functional image. Additionally, time series were inspected for excessive movement using the Artifact Detection Toolbox (ART; https://www.nitrc.org/projects/artifact_detect/; RRID:SCR_005994), allowing us to regress out functional volumes displaying excessive motion/spikes (see below).

Each participant's data were modeled with a general linear model (GLM) using an event-related design. To test our different empirical neural predictions, we performed two distinct GLMs. Common to all GLMs, regressors were convolved with the canonical HRF, and the cut-off period for high-pass filtering was 128 s. Six movement parameters, derived from spatial realignment, as well as spikes detected using ART were included as covariates of noninterest. First-level summary statistic images were entered in a second-level analysis in which subjects were treated as random effects. Our neural predictions are based on choice feedback related activity (Fig. 1A). We created two GLMs to test them. The SRPE-GLM tested both predictions (2, 3) associated with the SRPE effect on memory accuracy. The psychophysiological interaction (PPI)-GLM tested the PPI between SRPE (psychological factor) and subject-specific face-selective area (FSA) time series (physiological factor), and the hypothesis that functional connectivity between FSA and VS and hippocampus would be modulated by SRPE value (prediction 4). Below, we describe in detail our GLMs.

SRPE memory analyses. First, to reveal areas encoding SRPE, we constructed the SRPE-GLM. In the SRPE-GLM, we modeled the reading phase (i.e., the joint onset of faces and names) and the choice feedback period as boxcar functions (4 and 5 s, respectively); and choice onset, response presses, and monetary update as delta functions. We added a parametric modulator (PM) to the choice onset (note that here and in all subsequent analyses PMs are mean centered). This PM consisted of expected value updates (value PE) at the moment of choice onset given the structure of the experiment. Indeed, participants may have computed value PEs for each condition (i.e., one-, two-, and four-framed words) before experiencing the choice feedback RPE. Modeling this PM allows us therefore to control for potential spills of the predicted choice onset value PE onto the experienced choice feedback RPE. To compute the magnitude of these value PEs, we subtracted the overall probability of reward (i.e., 30 of 70 = 0.43 in our design) from the probability of being rewarded (i.e., from 1, 0.5, and 0.25 for our one-, two-, and four-framed conditions, respectively).

For our primary analysis, we added five regressors of choice feedback for each SRPE value (i.e., for levels -0.5 , -0.25 , 0 , 0.5 , and 0.75). To reveal the area encoding SRPE, we tested at the individual level for a mean-centered SRPE contrast [-0.6 , -0.35 , -0.1 , 0.40 , and 0.65] across the five regressors of interest (i.e., the choice-feedback events associated with our five SRPE values). Any area sensitive to this contrast would display increased activity as the SRPE value increases and would therefore encode SRPE. Given previous results (Davidow et al., 2016; Pine et al., 2018), we predicted that VS would encode SRPE (although statistical correction was at whole-brain level; prediction 2). Furthermore, we conducted three exploratory analyses based on the SRPE-GLM. In a first exploratory analysis, we assessed the SRPE contrast by extracting mean β weights for each of the five SRPE regressors in four distinct empirically derived regions of interest (ROIs) previously reported in episodic memory studies (Davidow et al., 2016; Ripollés et al., 2016; Pine et al., 2018). Our three first ROIs were defined following the same procedure. We used Neurosynth (<http://neurosynth.org/>) and searched for the terms “ventral striatum,” “ventral tegmental,” and “hippocampus,” and applied respectively a threshold of 18, 18, and 12 on z scores of the activation maps, thereby yielding our VS, VTA and hippocampus ROIs. The threshold values were chosen to match anatomic locations (i.e., to avoid activation outside the anatomic ROI). Our fourth FSA ROI was defined using subject-specific face contrasts (see below). In a second and third exploratory analysis, we respectively tested for an URPE and an inverse SRPE (iSRPE) over the five SRPE levels in the SRPE-GLM. The URPE levels are calculated by taking the absolute value of the SRPE levels followed by centering, to form the URPE contrast [0.1 , -0.15 , -0.4 , 0.1 , 0.35]. Such contrast would reveal any potential areas encoding for surprise, regardless of whether surprise has a positive or negative valence. The iSRPE contrast is computed simply by reversing the value signs in the SRPE contrast. As a sanity check, we tested for a memory-encoding effect instead of using the [1 , 1 , 1 , 1 , 1] contrast over the five regressors of interest in the SRPE-GLM. This memory check contrast would reveal any area reflecting a main effect of memory encoding and is expected to reveal bilateral medial temporal lobe (i.e., hippocampus) activation.

To test for a potential mediation effect of VS activation on the SRPE-memory link (prediction 3), we combined the behavioral data with the fMRI results. Trial characteristics (e.g., accuracy and neural activation) were, however, averaged per SRPE level for each participant because trial-level β weights (i.e., neural activation) cannot be estimated reliably. Hence, any analysis related to predictions 2 and 3 involving neural activation estimates is performed on condition-level data. As for the behavioral analyses, we report the χ^2 statistics from the ANOVA type III tests.

SRPE connectivity analyses. We conducted two standard PPI analysis (Gitelman et al., 2003; PPI-GLM-A and PPI-GLM-B). PPI can reveal which areas show an increase in functional connectivity with the seed (stimulus-relevant) area, as a function of SRPE. In PPI-GLM-A, three regressors were added (on top of the nuisance regressors described in the SRPE-GLM). The first regressor consisted of the BOLD signal extracted from a 3 mm sphere (Davidow et al., 2016) around the peak value from a subject-specific functionally localized face-selective area (FSA seed, see below; see Table 2, MNI coordinates of subject-specific peak values). The second regressor is the psychological vector consisting of the mean-centered SRPE contrast [-0.6 , -0.35 , -0.1 , 0.40 , 0.65] as described in the SRPE-GLM. The third regressor is the seed-by-condition interaction. A significant seed-by-condition interaction regressor will identify areas that show a coupling with the seed, and that coupling increases as a linear function of SRPE value (prediction 4). To control for reward, in PPI-GLM-B we performed the analysis only for rewarded trials. For that purpose, we performed the PPI analysis using the following contrast [-1 , 0 , 1] over the three SRPE levels of 0 , 0.5 , and 0.75 , respectively (i.e., only rewarded trials). Finally, to gain a deeper understanding of the functional connectivity among our ROIs, we also performed the PPI-GLM-A analysis using our hippocampus, VTA, and VS ROIs (defined above) as seeds of interest. We therefore estimated four PPI-GLM-A models in total, one for each seed ROI (i.e., FSA, VS, hippocampus, VTA), following the same method as mentioned above.

Our functional localizer task was modeled using a block design. Each block of the one-back task was modeled using a boxcar function over the

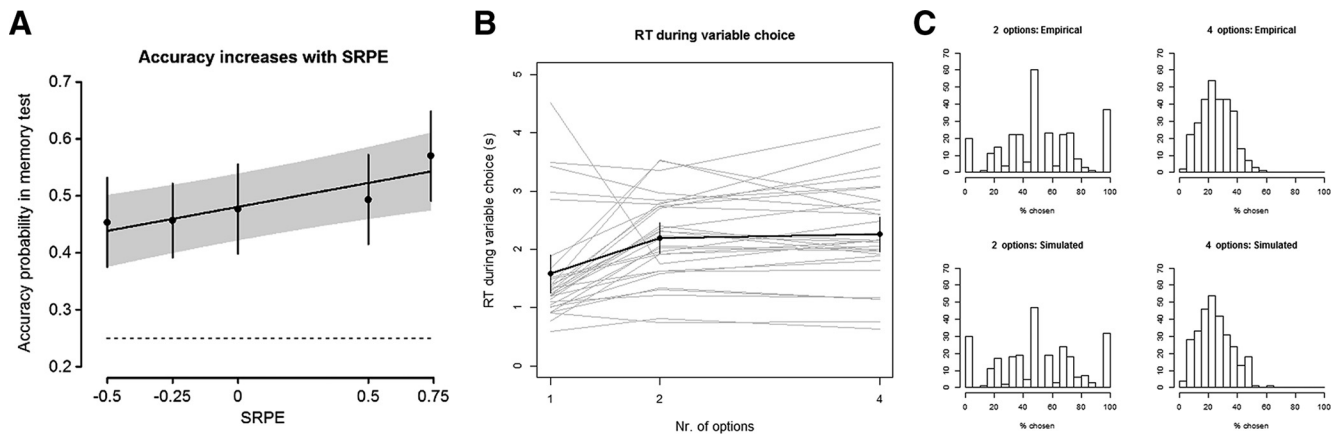


Figure 2. Behavioral results. **A**, A positive linear relationship between SRPE and accuracy on the subsequent memory test was observed (i.e., recognition accuracy increased as SRPE values increased). The regression line and 95% confidence band depict the effect of SRPE as a continuous regressor; the point estimates and 95% confidence intervals depict the effect of SRPE as a factor. The dashed line indicates chance performance (probability, 0.25). **B**, RT during variable choice (vertical axis) as a function of the number of options (horizontal axis). The black dots indicate the point estimates for the RTs, and the error bars indicate their 95% confidence intervals. For each of the 30 participants, their average RTs are plotted with a gray line. **C**, Empirical (top panels) and simulated (bottom panels) distributions for the probabilities that a particular village name was chosen during the choice period.

entire block duration. Nuisance regressors were modeled as in the SRPE-GLM. To reveal subject-dependent functional FSA, we contrasted face versus house blocks ($p < 0.0001$, uncorrected). Using the resulting statistical maps and MRICron (RRID:SCR_002403), we created an overlap image statistic displaying where and to what extent FSAs overlap with one another.

Results

Episodic memory improves with learning-evoked SRPE

Confirming prediction 1, and as expected from previous work (De Loof et al., 2018; Ergo et al., 2019; Jang et al., 2019), memory performance (i.e., trial-by-trial recognition accuracy) increased linearly with SRPE [$\chi^2(1, N = 30) = 11.72$, $\beta = 0.374$, $p = 0.00062$; Fig. 2A; point estimates of the accuracy probabilities for the five SRPE levels in increasing order are 0.453, 0.456, 0.477, 0.493, and 0.570]. This SRPE effect is statistically significant even when simultaneously including reward in the model [SRPE effect: $\chi^2(1, N = 30) = 4.32$, $\beta = 0.435$, $p = 0.038$; reward effect: $\chi^2(1, N = 30) = 0.12$, $\beta = -0.0615$, $p = 0.73$, note that the reward predictor is not significant]. Therefore, this SRPE-based linear increase in accuracy is unlikely to be reducible to a mere reward effect.

Furthermore, we refuted additional possible alternative explanations for our behavioral data. First, time-on-task does not linearly increase with the number of options (i.e., participants are not spending more time in conditions where they have to choose between more options; Fig. 2B). Also, there is some variability of the pattern across participants, as depicted in Figure 2B by the gray lines representing the average reaction times (RTs) for each participant. When fitting a linear mixed-effects model to predict the RTs based on the number of options (as a factor), this results in a significant effect of the number of options [$\chi^2(2, N = 30) = 15.06$, $p = 0.00054$]. Follow-up tests reveal that RTs are higher for two- and four-option trials compared with one-option trials (two options: $\chi^2(1, N = 30) = 13.7$, $\beta = 0.610$, $p = 0.00022$; four options: $\chi^2(1, N = 30) = 18.8$, $\beta = 0.678$, $p = 0.00014$). There is no significant difference between the RTs on the two- and four-option trials [$\chi^2(1, N = 30) = 0.79$, $\beta = 0.0680$, $p = 0.38$]. This comes as no surprise as we tried to minimize differences in RTs (at least for the two- and four-options trials) by including the reading period of 4 s before onset of the option frames (Fig. 1C). More broadly, we indeed controlled for

time-on-task in the behavioral analyses. To start, the RT during variable choice was not predictive for the memory accuracy [$\chi^2(1, N = 30) = 1.21$, $\beta = 0.0647$, $p = 0.27$]. We also tested the effect of RT on accuracy for the one-, two-, and four-option trials separately, with none of the analyses reaching significance (all p values > 0.16). Nevertheless, we also tested whether the RT during variable choice interacted with the effect of SRPE on memory accuracy. The effect of SRPE on accuracy remained unchanged [$\chi^2(1, N = 30) = 11.91$, $\beta = 0.378$, $p = 0.00056$], while there was no main effect of RT during variable choice [$\chi^2(1, N = 30) = 0.65$, $\beta = 0.0410$, $p = 0.42$], and no interaction with SRPE [$\chi^2(1, N = 30) = 0.76$, $\beta = -0.101$, $p = 0.38$].

Second, participants did not display choice biases in favor of specific village names. We ran a simulation to see whether the choices of the participants (empirical results) were in line with what would happen when no differences in expected values existed, and all choices are completely at random (simulated results). We calculated how often each particular village name was chosen when it was presented as a valid option in either a two-option trial (chance level, 50%) or a four-option trial (chance level, 25%). In the empirical data, individual village names were presented in two-option trials between 1 and 10 times (e.g., six presentations means that only six participants could choose this particular village name in a two-option trial; none of the other participants could choose this particular word in a two-option trial). By comparison, individual village names were presented in the four-option trials between 9 and 22 times. For all of the 304 village names (70 experiment trials + 6 practice trials), we calculated in what percentage of trials that particular village name was actually chosen by the participants. Figure 2C depicts the empirical and simulated probability of choosing the village name on the horizontal axis, and the frequency of the probabilities is presented on the vertical axis. The simulated histograms closely mimic the empirical results, indicating that it seems likely that there were no consequential differences among the expected values of the village names. If there would have been stable differences in expected values, the empirical histogram would have been biased toward higher percentages compared with the simulated histograms.

Finally, whether or not a celebrity was successfully named did not have any impact on the learning process. For our tests, we operationalized celebrity recognition in the following three

Table 1. Summary of the activation clusters

Contrast area (WFU_PickAtlas)	Local maxima MNI coordinates	Cluster size (no. of voxels)	Peak z	Cluster-level $p_{(FWE-corrected)}$
SRPE contrast (SRPE-GLM)**				
Left pallidum (peak)	−10, 4, −4	42	5.44	0.001
Right pallidum (peak)	12, 6, −6	34	5.03	0.002
PPI contrast (PPI-GLM)				
Cluster 1				
Right lateral front-orbital gyrus (peak)	20, 6, −16	369	4.64	0.017
Right putamen	14, 4, −12		4.57	
Right anterior hippocampus	38, −10, −14		3.90	
Cluster 2				
Right parahippocampal gyrus (peak)	8, −42, 8	11,388	5.94	0.000
Left parahippocampal gyrus	−8, −44, 8		5.53	
Left hippocampus	−22, −30, −4		5.44	
Cluster 3***				
Right hippocampus (peak)	24, −26, −8	115	5.01	0.009
	30, −36, −10		4.27	
Left hippocampus (peak)	−22, −30, −4	163	5.44	0.005
	−28, −34, −12		3.62	

Note that all contrasts survive cluster-level FWE correction ($p < 0.05$). We report peak and local maxima for all contrasts.

*Voxel-level threshold $p = 0.001$ uncorrected; **voxel-level threshold $p_{(FWE\ correction)} = 0.05$; ***SVC.

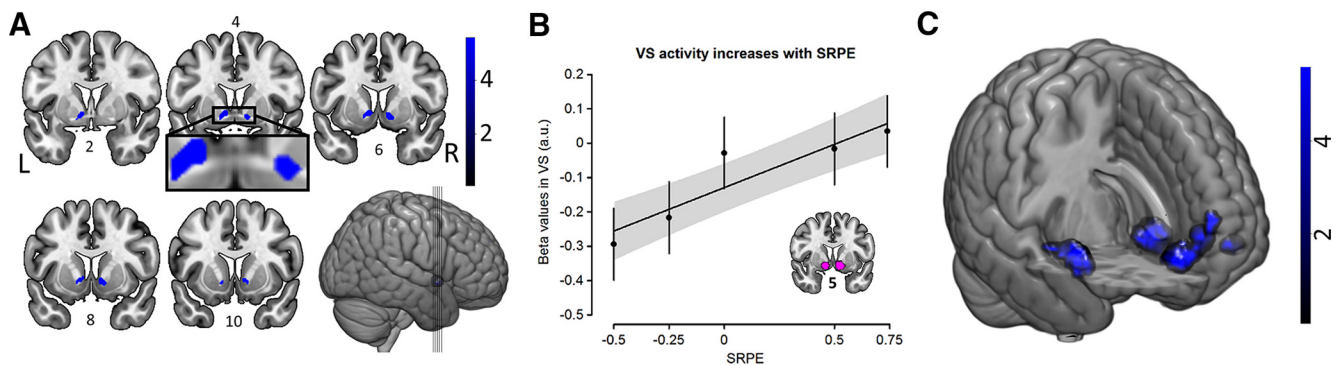


Figure 3. SRPE-GLM results (Table 1, map activation details). **A**, SRPE contrast from SRPE-GLM shows that the SRPE effect is robustly encoded in bilateral VS (blue activation map; left: $z = 5.41$, peak $[-10, 4, -4]$; right: $z = 5.04$, peak $[12, 6, -6]$). Numbers above/below brain slices represent y -dimension MNI coordinates. Color bar indicates z scores. **B**, Positive linear relation between VS activation and SRPE. Mean β weights from the SRPE contrast for each SRPE value were extracted from our Neurosynth-derived VS ROI (violet area in bottom right brain; number below brain slice represents the y -dimension MNI coordinate). The regression line and 95% confidence band depict the effect of SRPE as a continuous regressor; the point estimates and 95% confidence intervals depict the effect of SRPE as a factor. **C**, The memory-encoding contrast reveals robust bilateral hippocampal (left: $z = 5.81$, peak $[-24, -12, -16]$; right: $z = 6.19$, peak $[30, 159 -16, -14]$); vmPFC: $z = 5.41$, peak $[-6, 38, -10]$; and left IFC: $z = 5.95$, peak $[-42, 34, -12]$) activations ($p_{(FWE-corrected)} < 0.05$, voxel-level threshold). Color bar indicates z scores.

different ways: (1) a binary recognition score based on whether the celebrity was correctly named or not (regardless of certainty); (2) a binary recognition score based on whether the celebrity was correctly named in a confident way (“rather sure” or “completely sure”) or not (all other trials); and (3) a continuous score that ranges from -2 for faces that were misnamed with complete (undue) certainty to 2 for faces that were correctly named with complete certainty. None of these three measures significantly predicted the correct recognition of the face–word pair in the memory test (no significant random or fixed effects, all p values > 0.59). We also tested whether the three celebrity recognition measures altered the effect of SRPE on face–word pair recognition in the memory test. Compared with the reported effect of SRPE on accuracy [$\chi^2(1, N = 30) = 11.72$, $\beta = 0.374$, $p = 0.00062$] that main effect of SRPE remained stable in all three interaction models (all p values < 0.0012), there was no main effect of celebrity recognition (all p values > 0.67) and no significant interactions (all p values > 0.16).

Ventral striatum encodes SRPE

We next turn to the fMRI data (see Materials and Methods for description of fMRI data acquisition and analyses). We built two

classes of GLMs to test neural predictions described in the Introduction. To test which brain areas were more active as a function of SRPE levels, we built the SRPE-GLM. The PPI-GLM tested for the PPI between SRPE (psychological factor) and subject-specific FSA time series (physiological factor; PPI-GLM-B controlled for reward). All fMRI results, summarized in Table 1, are familywise error (FWE) cluster corrected (Table 1, contrast-specific voxelwise thresholds).

First, we investigated which brain areas encode SRPEs. Based on the SRPE-GLM, we tested at the individual subject level for a mean-centered SRPE contrast across the five regressors of interest (i.e., the choice–feedback events associated with our five SRPE values). This contrast would identify SRPE-sensitive brain areas with increased activity as SRPE value increases. The SRPE contrast revealed robust bilateral VS activations (blue activation: Fig. 3A, Table 1; left VS: FWE $p = 0.001$; right VS: FWE $p = 0.002$). This result supports prediction 2, suggesting a crucial role of the VS in computing SRPE, and aligns with earlier work (Hyman et al., 2006; Lisman et al., 2011; Scimeca et al., 2016; Pine et al., 2018). Notably, no other brain area was identified with this contrast. We ensured that VS activation did not reflect the computation of expected values at the moment of variable-

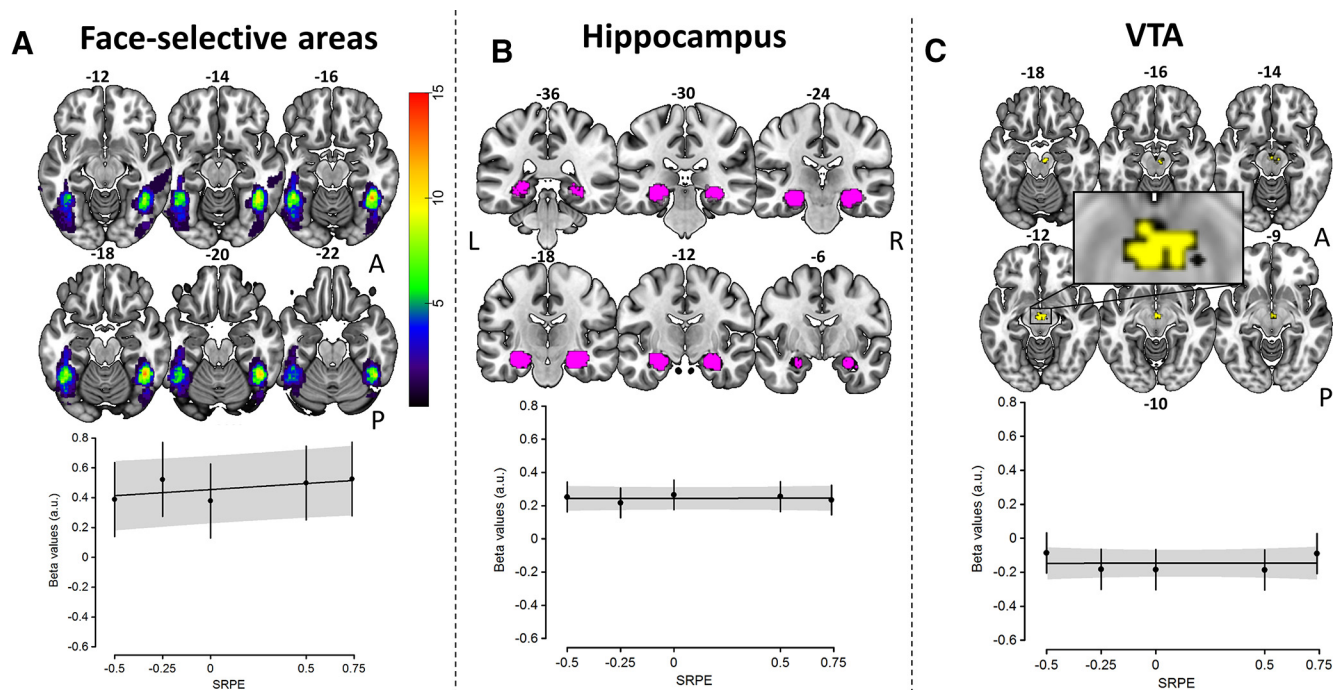


Figure 4. SRPE activation in ROIs. **A**, Face-selective areas. Top, Overlap of subject-specific face contrast activation maps in the inferior temporal lobes. Color bar indicates the number of subjects displaying activation in each voxel; numbers on top of slices indicate z-dimension MNI coordinates. Bottom, No linear relationship was observed between SRPE values and mean activation extracted from FSAs. **B**, Hippocampus. Top, Bilateral hippocampus ROI (violet) extracted from Neurosynth (see Materials and Methods). Numbers on top of slices indicate y-dimension MNI coordinates. Bottom, No linear relationship was observed between SRPE values and mean activation extracted from the hippocampus. **C**, VTA. Top, VTA ROI (yellow) extracted from Neurosynth (see Materials and Methods). Numbers above/under slices indicate z-dimension MNI coordinates. Bottom, No linear relationship was observed between SRPE values and mean activation extracted from VTA.

choice onset (see Materials and Methods). Specifically, we included a choice-onset regressor in the SRPE-GLM and added a PM to this regressor coding for choice onset expected value. When testing for this PM, no striatal activity was observed. Hence, the observed SRPE effect during a choice–feedback event is not an artifact of choice-onset regressor in the SRPE-GLM and added a parametric modulator (PM).

To further explore the striatal SRPE signature, we extracted the β weights from an independent Neurosynth-derived VS ROI (see Materials and Methods for ROI definition) for each SRPE level and subject. These condition-level data displayed a positive relationship with SRPE values [$\chi^2(1, N=30)=37.48$, $\beta = 0.117$, $p=9.3e-10$; Fig. 3B]. To validate that the SRPE effect is not reducible to a mere reward effect, we performed the following analysis: for each participant, we averaged the slope for the negative RPEs (i.e., -0.5 to -0.25) with the slope for the positive RPEs (i.e., 0.5 – 0.75), resulting in the average VS activation increase as the number of options is increased from two to four options (hence, controlling for the unbalanced design; Fig. 1D). We compared the slope of this activation increase against 0 (one-sample t test, one sided). Results show a significant increase ($t_{(29)} = 1.84$, $d = 0.335$, $p = 0.038$), suggesting an effect of SRPE, which is also determined by the number of options and is not reducible to a pure reward effect.

We further explored whether the linear effect of SRPE on activation was present in our other ROIs. We did not observe such an effect in FSAs (see below), hippocampus, or VTA (all $p > 0.09$; Fig. 4). We then tested whether specific brain areas would encode URPE and iSRPE effects (see Materials and Methods). No brain areas were active based on these contrasts.

Finally, as a sanity check, to ensure that participants performed the task (i.e., encoded the face–word associations), we tested for a main effect of memory encoding (see Material and

Methods). This effect revealed bilateral hippocampus (i.e., mediolateral temporal lobe; Fig. 3C).

Striatal activation mediates the effect of SRPEs on memory accuracy

Given the positive linear relationship between SRPE values and subsequent memory, and the finding that SRPE values are encoded in the VS, VS activation may mediate the SRPE–memory link (prediction 3). To test this, we performed a mediation analysis (Sobel, 1982; Baron and Kenny, 1986; Padmala and Pessoa, 2011) to investigate whether VS activation extracted from the SRPE-GLM (for all five SRPE values: condition-level data, within subjects) mediates the relation between SRPE and memory accuracy. We already demonstrated that SRPE significantly predicts memory accuracy on trial-level data (Fig. 2); this also holds at the condition level [$\chi^2(1, N=30)=11.18$, $\beta = 0.0382$, $p=0.00083$; Fig. 5A]. We had already established the highly significant relation between SRPEs and VS activity at the condition level [Fig. 3B; $\chi^2(1, N=30)=37.48$, $\beta = 0.117$, $p=9.3e-10$; Fig. 5B]. At the condition level, VS activity is predictive for accuracy [$\chi^2(1, N=30)=13.44$, $\beta = 0.0520$, $p=0.00025$; Fig. 5C]. Crucially, when we predict condition-level accuracy based on a model that includes the main effects of SRPE and VS activation, the highly significant SRPE effect disappears [$\chi^2(1, N=30)=3.48$, $\beta = 0.0240$, $p=0.062$], but the VS activation is still predictive for memory accuracy [$\chi^2(1, N=30)=5.70$, $\beta = 0.0378$, $p=0.017$; Fig. 5D].

Functional connectivity between stimulus-processing areas and VS and hippocampus depends on SRPE value

Next, we tested whether the encoding of episodic memory is reflected by increased functional connectivity between stimulus-

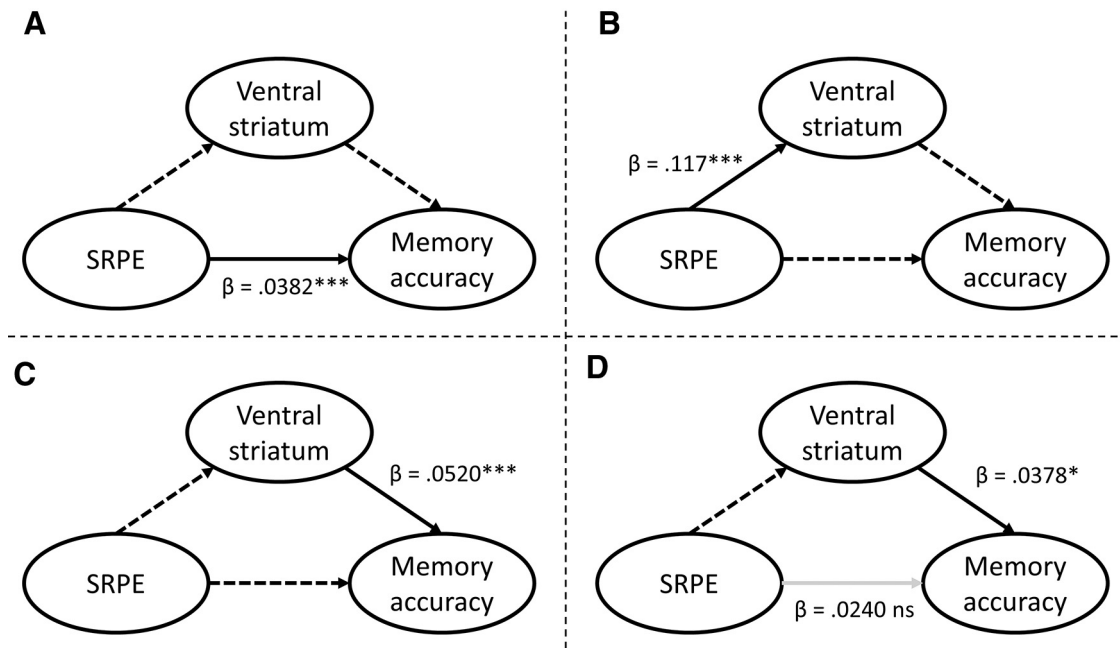


Figure 5. Mediation analysis. **A**, Effect of SRPE on memory accuracy at the condition level. **B**, Effect of SRPE on VS activation as reported in the Results subsection entitled Ventral striatum encodes SRPE. **C**, Effect of VS activation on memory accuracy at the condition level. **D**, When both SRPE and VS activation are jointly added as predictors of memory accuracy, VS activation predicts memory accuracy while SRPE does not. This suggests that the SRPE effect on memory is mediated by VS activation. Full lines define predictors and independent variables in the model, dashed lines are inactive (unmodeled) links. Black and light gray lines represent significant and nonsignificant predictions, respectively (* $p < 0.05$, *** $p < 0.001$).

processing areas (i.e., subject-specific FSAs; Fig. 4A) and VS and hippocampus as a function of SRPE value (prediction 4). Figure 4A shows the across-subject overlap image statistic displaying where and to what extent FSAs overlap, showing bilateral activation maps in the inferior temporal lobes. Table 2 reports the peak activation of face contrast activation maps for each participant.

The PPI analysis (PPI-GLM-A) revealed two clusters surviving whole-brain analysis (Fig. 6A,B, blue activation map; FWE-corrected $p < 0.05$, cluster level). Cluster 1 (Fig. 6A) encompasses activation in the right VS, extending to the right (anterior) hippocampus and right amygdala. Cluster 2 (Fig. 6B) displays activation in left hippocampus, bilateral parahippocampal gyri, and medial and inferior temporal lobe, as well as superior parietal lobe. Note, however, that only bilateral hippocampal gyrus peaks survived cluster-level correction (Table 1). Cluster 3, encompassing hippocampus, survived theoretically motivated small volume correction (SVC). Indeed, hippocampus has been involved in previous episodic memory connectivity results and was therefore a theoretically predicted area (Davidow et al., 2016). Cluster 3 displays activation along bilateral hippocampus (Fig. 6A, blue activation map), and SVC was applied using our Neurosynth-derived hippocampus ROI (see Materials and Methods).

We performed the same PPI analysis only on rewarded trials [PPI-GLM-B; i.e., any PPI results can therefore not be explained by a mere reward (rather than SRPE) effect]. This analysis revealed activation of the left hippocampus (Fig. 6A, yellow activation map). Overall, our PPI results demonstrate that functional connectivity between a stimulus-processing area (FSA) and a network that presumably supports RL of episodic memory (VS, hippocampus; Lisman et al., 2011; Watabe-Uchida et al., 2017) increases as a function of SRPE. Additional exploratory analyses revealed two points worth mentioning. First, with our other three ROIs as seeds (i.e., hippocampus, VS, VTA), no significant results were observed. Second, no significant (across-subjects)

Table 2. MNI Coordinates of subject-specific face contrast activation maps peak value

Subject (n)	MNI coordinates	
	Left temporal lobe	Right temporal lobe
1	-46, -54, -16	44, -50, -20
2	-38, -70, -16	46, -48, -22
3	-36, -54, -20	48, -48, -22
4	-48, -54, -18	38, -56, -8
5	-40, -78, 8	44, -46, 0
6	N/A	52, -76, -4
7	-38, -78, 4	50, -44, 4
8	-40, -68, -16	46, -74, -2
9	-44, -46, -14	44, -44, -16
10	-44, -60, -16	34, -56, -12
11	-38, -52, -16	42, -58, -12
12	-46, -50, -22	52, -60, -18
13	N/A	40, -56, -16
14	-38, -52, -12	42, -46, -12
15	-46, -60, -14	40, -56, -18
16	-44, -72, -14	36, -54, -16
17	-64, -40, -8	N/A
18	-38, -44, -26	42, -44, -16
19	-44, -46, -20	48, -48, -24
20	-38, -48, -18	40, -46, -18
21	-42, -64, -14	42, -60, -12
22	-46, -68, -18	42, -54, -18
23	-44, -48, -18	44, -54, -16
24	-36, -48, -16	40, -52, -22
25	-44, -44, -26	42, -68, -16
26	-38, -62, -18	42, -54, -20
27	-46, -70, 8	46, -52, -16
28	N/A	42, -64, -12
29	-38, -80, -12	48, -74, -8
30	-4, -48, -16	40, -42, -18

N/A, Not applicable. We report the activation peaks of subject-specific face-selective areas in MNI coordinates. Face-selective activation maps can be seen in Figure 4A, and the peaks of these maps are used in the PPI analysis described in the Materials and Methods.

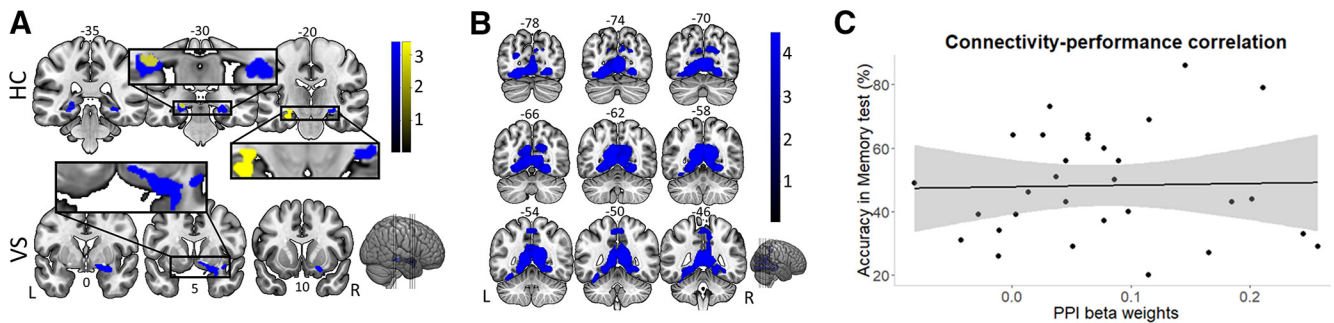


Figure 6. Connectivity results. **A**, Functional connectivity with FSA as a function of SRPE value reveals clusters (PPI-GLM-A, blue activation map) encompassing bilateral hippocampus (left: $z = 5.44$, peak $[-22, -30, -4]$; right: $z = 5.01$, peak $[24, -26, -8]$; and right VS: $z = 4.57$, peak $[14, 4, -12]$). Controlling for reward revealed activation of the left hippocampus (PPI-GLM-B, yellow activation map; $z = 4.30$; peak $[-28, -18, -18]$). **B**, Additional connectivity cluster. For transparency, we display the whole of cluster 2 (PPI-GLM-A), displaying activation in medial and inferior temporal lobes, as well as in the superior parietal lobe. Color bar indicates z scores; numbers on top of slices indicate y -dimension MNI coordinates. **C**, Correlation between connectivity strength and memory accuracy (percentage). No relation was found between connectivity strength and memory accuracy.

correlation was observed between connectivity strength and memory accuracy (Fig. 6C).

Discussion

Using fMRI and our novel variable-choice task, we experimentally manipulated SRPE. In contrast to previous studies (Davidow et al., 2016), our design did not involve incremental learning to induce prediction errors. Rather, by manipulating the number of eligible responses, RPEs are calculated based on probability theory. Using this paradigm, we confirmed key RL theory predictions regarding episodic memory formation. First, we extended the behavioral effect of SRPEs on memory to faces and pseudo-names stimuli; stimuli associated with higher SRPE values induce better recognition test performance. Second, we confirmed that SRPEs are encoded in the VS. Third, we observed that VS activation mediates the relation between SRPE value and episodic memory within subjects. Fourth, we demonstrated that functional connectivity between task-relevant areas and VS and hippocampus is modulated by SRPE.

Signed reward prediction errors influence encoding of episodic information

Our behavioral results replicated and extended work showing that SRPEs drive episodic learning. A stronger better-than-expected reward signal is associated with better performance at a subsequent recognition test (De Loof et al., 2018).

We interpret our results in light of the neoHebbian framework (Lisman et al., 2011). This framework suggests that episodic learning is based on principles similar to those of procedural learning (Watabe-Uchida et al., 2017). In this framework, in our variable-choice task, whenever the obtained reward deviates from the expectation, an RPE mismatch signal is generated in midbrain nuclei. This signal is broadcast to the hippocampus and VS to enhance long-term potentiation of the to-be-remembered face-word association. As a result, the stronger the mismatch (i.e., the higher the RPE value), the stronger the potentiation and the better the memory accuracy.

Alternatively, the SRPE modulation may derive from the RL principle of prioritized experience replay (Mattar and Daw, 2018; Sutton and Barto, 2018). Here, (episodic) events are replayed during offline periods to increase memory consolidation, with events being prioritized depending on the RPE attached to them (the larger the RPE, the higher the priority). Following this interpretation, the SRPE effect would occur in the (in this experiment, relatively short) period between learning and the test. In our

study, we observed an RPE effect on memory accuracy directly after the encoding task. Based on the idea of prioritized replay, future research should investigate whether the relationship between SRPE and memory may be enhanced (i.e., Fig. 2, steeper slope) for delayed versus immediate memory test. De Loof et al. (2018) observed results supporting this prediction (i.e., the SRPE slope is steeper in the delayed versus immediate memory test). Furthermore, several studies observed reward-related effects in delayed (and not immediate) memory tasks (Wittmann et al., 2005; Braun et al., 2018), suggesting that the reward effect (via prioritized replay) on memory benefits from longer consolidation times. Additionally, the SRPE effect on episodic memory can last up to 24 h (Davidow et al., 2016; De Loof et al., 2018; Jang et al., 2019) or a week (Pine et al., 2018). Therefore, the RPE effect operates at time scales supporting learning and memory in the real world.

In our work, we observed a signed RPE effect. However, some studies found support for a URPE effect on memory (Butterfield and Metcalfe, 2001; Fazio and Marsh, 2009; Metcalfe, 2017; Rouhani et al., 2018; Stanek et al., 2019). As discussed in the study by Ergo et al. (2020), these differences may depend on how RPEs are manipulated. For instance, multiple-repetition paradigms (Butterfield and Metcalfe, 2006; Metcalfe et al., 2012), where questions querying memorized information are repeatedly presented and trial-by-trial RPEs are estimated based on the response confidence and feedback, usually yield URPE effects. In contrast, in reward-prediction paradigms (Wimmer et al., 2014) participants must learn declarative information while tracking a reward distribution to maximize their reward; here, trial-by-trial RPEs are derived from estimated parameters obtained through fitting an RL model. These paradigms usually yield an SRPE effect. Future research should investigate why these two classes of paradigms typically display either URPE or SRPE effects. Interestingly, however, a recent study using the variable-choice task observed both SRPE and URPE electrophysiological signatures in distinct frequency bands, respectively high-beta/alpha and theta (Ergo et al., 2019). Thus, both SRPE and URPE may be important for episodic (declarative) learning. In fact, Bayesian (normative) learning models naturally combine both pieces of information. For instance, the Kalman filter model updates its value estimates using SRPE, and its learning rate using URPE (i.e., considered here as a proxy for uncertainty; Dayan et al., 2000).

Striatal activation mediates the effect of SRPEs on subsequent memory

If RPEs are broadcast to the hippocampus via reward-processing areas such as the VS, then VS should mediate the link between

RPE and memory accuracy. Our data support this prediction at the within-subject level. These results nicely complement previous work showing that PEs/RPEs are encoded in the striatum (Tricomi and Fiez, 2008; Gläscher et al., 2010; Daw et al., 2011) and extend them to show their crucial mediating role in episodic memory performance.

The dopaminergic neuromodulatory nature of the RPE is phasic in our task (i.e., computed at the choice feedback event; see also Wittmann et al., 2005). However, several studies investigated the effects of tonic dopaminergic signal manipulations on memory. For instance, reward can act as a motivational incentive cue and lead to better item memory in high- versus low-rewarding contexts (Aberg et al., 2020). Rewarding contexts could therefore induce a tonic, rather than phasic, dopaminergic signal. This would imply that any memoranda being encoded during this tonic signal would display beneficial reward-based memory effects. In line with this idea, novel word learning performance is enhanced when participants are given levodopa (a dopaminergic precursor) compared with risperidone (a dopamine antagonist; Ripollés et al., 2018). Interestingly, Aberg et al. (2020) showed that phasic and tonic dopaminergic signals, via manipulation of immediate or average reward magnitude, influence memory distinctively. While the magnitude of immediate reward (presumably phasic dopaminergic effect) scales linearly, average reward (presumably tonic dopaminergic effect) scales nonlinearly (i.e., inverted U shape) with memory accuracy. This suggests that phasic and tonic dopamine levels may interact (i.e., reward may have a differential effect on memory consolidation), depending on the tonic dopamine level. As a testable prediction for future work, participants with higher dopamine synthesis capacity may benefit more from RPE-based learning. Relatedly, dopamine synthesis capacities have been shown to explain differences in effort-based decision-making, also involving VS (Westbrook et al., 2020).

Our task is designed to test the effect of RPEs during encoding. However, striatal RPEs may also play an important role during retrieval. For instance, in an old–new item decision task, biased positive feedback (i.e., overall more positive feedback regardless of decision accuracy) induces a shift in the decision criterion (toward a more lenient criterion; Scimeca et al., 2016). Thus, striatal RPEs play a role in the strategies used to make memory decisions. Together with ours, these earlier results suggest that RL may support episodic memory both during encoding and retrieval.

SRPE value modulates connectivity strength between stimulus-relevant areas and the episodic memory network

An important feature of our task is that participants remembered face–word associations. Therefore, using an independent localizer, we were able to investigate connectivity profiles between stimulus-processing (i.e., face-selective) areas and a previously reported RPE-based learning network consisting of VS and hippocampus (Davidow et al., 2016). The results confirmed our prediction: connectivity between face-selective areas and the episodic learning network increases with SRPE values.

One potential mechanism underlying the link between RPE and connectivity observed here may be phase locking to neural oscillations in specific frequency bands. Recent evidence suggests that brain areas in theta-phase synchrony communicate and learn efficiently (Fries, 2015), thereby facilitating memory (Backus et al., 2016). Furthermore, episodic memory is enhanced when to-be-remembered audiovisual stimulus associations are synchronously presented in theta phase (Clouter et al., 2017;

Wang et al., 2018). Interestingly, theta-phase locking between dopaminergic midbrain neurons and cortical areas is stronger for remembered stimuli (Kamiński et al., 2018). Thus, RPEs (via neuromodulatory signaling) may enhance theta synchrony, leading to more tightly phase locked presynaptic and postsynaptic activity, and, thus, in line with the neoHebbian framework (Lisman et al., 2011), relevant brain areas would store the episode efficiently (Berens and Horner, 2017).

In line with the prioritized experience replay concept discussed above, SRPE may also modulate postlearning connectivity: consistently, in a study where prelearning versus postlearning changes in the VTA-hippocampus resting-state functional connectivity predicted reward-related memory recognition advantages (Gruber et al., 2016). However, this postlearning connectivity needs to be studied more systematically in the current experimental paradigm.

Conclusions

We suggest that episodic memory encoding is guided by an RPE-based neural (RL) mechanism, as is also the case in procedural learning (Schultz et al., 1997). In episodic memory, bilateral VS may relay midbrain-computed RPEs toward stimulus-processing areas and hippocampus, thereby increasing functional connectivity between hippocampus and cortical areas.

References

- Aberg KC, Kramer EE, Schwartz S (2020) Interplay between midbrain and dorsal anterior cingulate regions arbitrates lingering reward effects on memory encoding. *Nat Commun* 11:1829.
- Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JDE (2006) Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50:507–517.
- Backus AR, Schoffelen JM, Szebenyi S, Hanslmayr S, Doeller CF (2016) Hippocampal-prefrontal theta oscillations support memory integration. *Curr Biol* 26:450–457.
- Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 51:1173–1182.
- Berens SC, Horner AJ (2017) Theta rhythm: temporal glue for episodic memory. *Curr Biol* 27:R1110–R1112.
- Berman MG, Park J, Gonzalez R, Polk TA, Gehrke A, Knaffla S, Jonides J (2010) Evaluating functional localizers: the case of the FFA. *Neuroimage* 50:56–71.
- Braun EK, Wimmer GE, Shohamy D (2018) Retroactive and graded prioritization of memory by reward. *Nat Commun* 9:4886.
- Butterfield B, Metcalfe J (2001) Errors committed with high confidence are hypercorrected. *J Exp Psychol Learn Mem Cogn* 27:1491–1494.
- Butterfield B, Metcalfe J (2006) The correction of errors committed with high confidence. *Metacognition Learning* 1:69–84.
- Clouter A, Shapiro KL, Clouter A, Shapiro KL, Hanslmayr S (2017) Theta Phase Synchronization Is the Glue that Binds Human Associative Memory. *Curr Biol* 27:3143–3148.
- Davidow JY, Foerde K, Galván A, Shohamy D (2016) An upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence. *Neuron* 92:93–99.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–1215.
- Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. *Nat Neurosci* 3:1218–1223.
- De Loof E, Ergo K, Naert L, Janssens C, Talsma D, Van Opstal F, Verguts T (2018) Signed reward prediction errors drive declarative learning. *PLoS One* 13:e0189212.
- Ergo K, De Loof E, Janssens C, Verguts T (2019) Oscillatory signatures of reward prediction errors in declarative learning. *Neuroimage* 186:137–145.
- Ergo K, De Loof E, Verguts T (2020) Reward prediction error and declarative memory. *Trends Cogn Sci* 24:388–397.

- Fazio LK, Marsh EJ (2009) Surprising feedback improves later memory. *Psychon Bull Rev* 16:88–92.
- Fries P (2015) Rhythms for cognition: communication through coherence. *Neuron* 88:220–235.
- Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19:200–207.
- Gläscher J, Daw N, Dayan P, O’Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585–595.
- Gruber MJ, Ritchey M, Wang SF, Doss MK, Ranganath C (2016) Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron* 89:1110–1120.
- Hyman SE, Malenka RC, Nestler EJ (2006) Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci* 29:565–598.
- Jang AI, Nassar MR, Dillon DG, Frank MJ (2019) Positive reward prediction errors during decision-making strengthen memory encoding. *Nat Hum Behav* 3:719–732.
- Kamiński J, Mamelak AN, Birch K, Mosher CP, Tagliati M, Rutishauser U (2018) Novelty-Sensitive Dopaminergic Neurons in the Human Substantia Nigra Predict Success of Declarative Memory Formation. *Curr Biol* 28:1333–1343.e4.
- Lisman J, Grace AA, Duzel E (2011) A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends Neurosci* 34:536–547.
- Mattar MG, Daw ND (2018) Prioritized memory access explains planning and hippocampal replay. *Nat Neurosci* 21:1609–1617.
- Metcalfe J (2017) Learning from errors. *Annu Rev Psychol* 68:465–489.
- Metcalfe J, Butterfield B, Habel C, Stern Y (2012) Neural correlates of people’s hypercorrection of their false beliefs. *J Cogn Neurosci* 24:1571–1583.
- Miendlarzewska EA, Bavelier D, Schwartz S (2016) Influence of reward motivation on human declarative memory. *Neurosci Biobehav Rev* 61:156–176.
- Padmala S, Pessoa L (2011) Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *J Cogn Neurosci* 23:3419–3432.
- Pine A, Sadeh N, Ben-Yakov A, Dudai Y, Mendelsohn A (2018) Knowledge acquisition is governed by striatal prediction errors. *Nat Commun* 9:14.
- Ripollés P, Marco-Pallarés J, Alicart H, Tempelmann C, Rodríguez-Fornells A, Noesselt T (2016) Intrinsic monitoring of learning success facilitates memory encoding via the activation of the SN/VTA-hippocampal loop. *Elife* 5:e17441.
- Ripollés P, Ferreri L, Mas-Herrero E, Alicart H, Gómez-Andrés A, Marco-Pallarés J, Antonijoan RM, Noesselt T, Valle M, Riba J, Rodríguez-Fornells A (2018) Intrinsically regulated learning is modulated by synaptic dopamine signaling. *Elife* 7:e38113.
- Rouhani N, Norman KA, Niv Y (2018) Dissociable effects of surprising rewards on learning and memory. *J Exp Psychol Learn Mem Cogn* 44:1430–1443.
- Schiffer A-M, Muller T, Yeung N, Waszak F (2014) Reward activates stimulus-specific and task-dependent representations in visual association cortices. *J Neurosci* 34:15610–15620.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Scimeca JM, Katzman PL, Badre D (2016) Striatal prediction errors support dynamic control of declarative memory decisions. *Nat Commun* 7:13061.
- Shneyer A, Mendelsohn A (2018) Previously rewarding environments enhance incidental memory formation. *Learn Mem* 25:569–573.
- Sobel ME (1982) Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol* 13:290.
- Stanek JK, Dickerson KC, Chiew KS, Clement NJ, Adcock RA (2019) Expected reward value and reward uncertainty have temporally dissociable effects on memory formation. *J Cogn Neurosci* 31:1443–1454.
- Sutton R, Barto A (2018) Reinforcement learning: an introduction. Cambridge, MA: MIT.
- Tricomi E, Fiez JA (2008) Feedback signals in the caudate reflect goal achievement on a declarative memory task. *Neuroimage* 41:1154–1167.
- Tulving E (1993) What is episodic memory? *Curr Dir Psychol Sci* 2:67–70.
- Wang D, Clouter A, Chen Q, Shapiro KL, Hanslmayr S (2018) Single-trial phase entrainment of theta oscillations in sensory regions predicts human associative memory performance. *J Neurosci* 38:6299–6309.
- Watabe-Uchida M, Eshel N, Uchida N (2017) Neural circuitry of reward prediction error. *Annu Rev Neurosci* 40:373–394.
- Westbrook A, van den Bosch R, Määttä JJ, Hofmans L, Papadopetraki D, Cools R, Frank MJ (2020) Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science* 367:1362–1366.
- Wimmer GE, Braun EK, Daw ND, Shohamy D (2014) Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. *J Neurosci* 34:14901–14912.
- Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, Düzel E (2005) Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron* 45:459–467.