Psychopharmacology (2019) 236:2425–2435
https://doi.org/10.1007/s00213-019-05322-z

ORIGINAL INVESTIGATION

Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning

Nina Rouhani [1,2] · Yael Niv [1,2]

Received: 22 March 2019 / Accepted: 30 June 2019 / Published online: 26 July 2019

Abstract

Rationale Depression is a disorder characterized by sustained negative affect and blunted positive affect, suggesting potential abnormalities in reward learning and its interaction with episodic memory.

Objectives This study investigated how reward prediction errors experienced during learning modulate memory for rewarding events in individuals with depressive and non-depressive symptoms.

Methods Across three experiments, participants learned the average values of two scene categories in two learning contexts. Each learning context had either high or low outcome variance, allowing us to test the effects of small and large prediction errors on learning and memory. Participants were later tested for their memory of trial-unique scenes that appeared alongside outcomes. We compared learning and memory performance of individuals with self-reported depressive symptoms ($N = 101$) to those without ($N = 184$).

Results Although there were no overall differences in reward learning between the depressive and non-depressive group, depression severity within the depressive group predicted greater error in estimating the values of the scene categories. Similarly, there were no overall differences in memory performance. However, in depressive participants, negative prediction errors enhanced episodic memory more so than did positive prediction errors, and vice versa for non-depressive participants who showed a larger effect of positive prediction errors on memory. These results reflected differences in memory both within group and across groups.

Conclusions Individuals with self-reported depressive symptoms showed relatively intact reinforcement learning, but demonstrated a bias for encoding events that accompanied surprising negative outcomes versus surprising positive ones. We discuss a potential neural mechanism supporting these effects, which may underlie or contribute to the excessive negative affect observed in depression.

Keywords Depression · Memory · Reinforcement learning · Reward · Predictions errors · Surprise

Introduction

Memories help guide future behavior, but which experiences from the past are prioritized? In reinforcement learning, the value of an option is computed by a weighted average over all experienced outcomes, suggesting we integrate across multiple memories when making a decision. In contrast, episodic memories represent single events and allow for rapid, one-shot learning of relations between stimuli and outcomes. Interactions between reinforcement learning, supported by the striatum, and episodic memory, supported by the hippocampus, predict decision-making in a variety of behavioral and neural paradigms (Bornstein et al. 2017; Duncan and Shohamy 2016; Gershman and Daw 2017; Wimmer and Shohamy 2012). Given putative dopaminergic innervation linking the striatum to the hippocampus (Shohamy and Adcock 2010), special attention has been given to reward prediction errors in reinforcement learning and their effect on memory. Reward prediction errors are phasic dopaminergic

This article belongs to a *Special Issue on Translational Computational Psychopharmacology*

✉ Nina Rouhani
ninarouhani@gmail.com

[1] Department of Psychology, Princeton University, Princeton, NJ 08544, USA

[2] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA

signals that increase when outcomes are better than expected and decrease when they are worse than expected, and are thought to update the value of the events that led to the outcomes (Barto 1995; Read Montague et al. 1996).

In previous work, we investigated the interaction between reward prediction errors and episodic memory and found that unsigned (absolute) prediction errors increase memory for a rewarding event, thereby prioritizing more surprising events in memory (Rouhani et al. 2018). Consistent with this behavioral finding, recent work has shown that the locus coeruleus (LC), a region modulated by unsigned prediction errors (for a review, see Sara 2009), co-releases dopamine with norepinephrine, leading to dopamine-dependent plasticity in the hippocampus (Kempadoo et al. 2016; Takeuchi et al. 2016). This work highlights a new source of dopamine other than the ventral tegmental area and substantia nigra pars compacta (where the signed reward prediction error originates), which leads to new predictions of how events that modulate LC activity, such as unsigned reward prediction errors, might boost hippocampal memories (Duszkiewicz et al. 2018).

Importantly, we found that both positive and negative prediction errors enhanced memory. It is unclear, however, how disorders marked by blunted positive and excessive negative affect, such as depression, may bias these effects on memory. To this end, we collected depression scores from all participants in our original sample (Rouhani et al. 2018) and tested for effects of depressive symptoms on reward learning, recognition memory, and the modulation of memory by prediction errors.

Previous work characterizing reinforcement learning in major depressive disorder (MDD) has demonstrated decreased sensitivity to rewards (Huys et al. 2013) as well as hypoactivation of reward-related responses in the striatum (for reviews, see Admon and Pizzagalli 2015; Pizzagalli 2014). Accordingly, attenuated reward prediction error signals are reported in MDD (Gradin et al. 2011; Kumar et al. 2018), although these signals were intact in a task that did not require learning (Rutledge et al. 2017). Moreover, behavioral differences in reinforcement learning in MDD have been mixed. Many studies have shown similar learning performance between MDD patients and controls (e.g., Ubl et al. 2015) with differences modulated by individual levels of anhedonia (the inability to feel pleasure) independent of depression severity (Admon and Pizzagalli 2015; Chase et al. 2010). In our nonclinical sample, we therefore did not expect to see large differences in reward learning between those experiencing depressive symptoms and those that do not.

In addition to blunted reward processing, sustained negative affect in depression has led to work showing an asymmetry in processing negative over positive events. Namely, MDD patients show an attentional bias for negative stimuli, displaying difficulty in disengaging from and ignoring negative distractors (for reviews, see Gotlib and Joormann 2010; Joormann and Quinn 2014). In reinforcement learning, neuroimaging studies bolster evidence for this asymmetry by showing hyperactivation of cortico-striatal learning networks for punishment versus reward (Admon and Pizzagalli 2015; Kumar et al. 2018; Ubl et al. 2015), including stronger prediction error signals for punishment (Kumar et al. 2018; Ubl et al. 2015). Of note, in depression, connectivity between the striatum and anterior cingulate cortex, a region associated with unsigned prediction errors (Roesch et al. 2012), is blunted in reward learning (Whitton et al. 2016) and enhanced in punishment learning (Admon et al. 2015).

These results suggest that in depressed individuals, high-magnitude negative prediction errors may have greater influence on learning and memory than do positive prediction errors. In line with this, depressed individuals exhibit a bias for negative versus positive memories (Gaddy and Ingram 2014; Matt et al. 1992). This better memory for negative events in depression is thought to be modulated by the amygdala—a region associated with emotional memories as well as surprising events—and its functional connectivity with the hippocampus (Dillon and Pizzagalli 2018; Leal et al. 2014; Sacchet et al. 2017; Young et al. 2017). Healthy individuals, on the other hand, exhibit a bias for positive versus negative memories, whereas depressed individuals additionally demonstrate an attenuated memory for positive events (Burt et al. 1995), which is linked to reduced activation in the dopaminergic midbrain and medial temporal lobes (Dillon et al. 2014). The literature therefore offers two mechanisms by which depressed individuals' memory may be biased compared with healthy individuals—better memory for negative events and worse memory for positive events. What remains to be elucidated is whether reward prediction error signals modulate this asymmetry in memory in depressed individuals.

To test this, across three experiments, participants learned by trial and error the values of two scene categories (indoor and outdoor scenes) in two learning contexts. Each learning context was associated with a high or low level of outcome variance, allowing us to compare the effects of high-magnitude and low-magnitude reward prediction errors on learning and memory. We compared individuals reporting symptoms of depression (the "depressive" group) with those that reported no such symptoms ("non-depressive" group) in terms of their learning of the average values of the two scene categories, their trial-by-trial prediction errors and learning rates, and their recognition memory for those rewarding events.

Methods

In three experiments, we tested how reward prediction errors experienced during reinforcement learning interact with episodic memory. In each experiment, participants learned by

trial and error the values of two types of scenes, indoor and outdoor images, in two different contexts ("rooms"). One room was associated with low outcome variance ("low-risk room") while the other room was associated with high outcome variance ("high-risk room"; order counterbalanced). The average value of the high- and low-value scene categories was matched across contexts. Critically, on each trial, participants viewed trial-unique indoor or outdoor images, allowing us to test for memory of rewarding events at a later stage in the experiment.

Participants

Across three experiments run on Amazon Mechanical Turk, 500 participants initiated the study (Exp 1—200, Exp 2—200, Exp 3—100), 408 completed the study (Exp 1—174, Exp 2—148, Exp 3—86), and after exclusions (see below), 383 participants are represented in our sample (Exp 1—164, Exp 2—136, Exp 3—83). Participants completed informed consent online and were required to correctly answer questions checking for their understanding of the task before proceeding; procedures were approved by Princeton University's Institutional Review Board. Participants were excluded from analysis if they (1) missed more than three trials during learning or (2) had a memory score (A': sensitivity index in signal detection based on their hit rate and false alarm rate for item recognition memory; Pollack and Norman 1964) of less than 0.5.

Task procedure

Learning In each experiment, participants learned the values of two scene categories (indoor or outdoor images) in two "rooms" defined by distinct background colors. The average values of the high- and low-value scene categories were matched across rooms, but the reward variance of the "high-risk room" was approximately double that of the "low-risk room." Participants first saw all the images in one room and then the other (i.e., presentation was blocked, order randomized). Participants were instructed that in each room, one type of scene category was more valuable than the other, and after viewing all of the images within a room, they were asked to indicate the "winner" (i.e., the high-value scene category). Participants were not explicitly informed about the reward values of the scene categories nor that the two rooms would have different levels of outcome variance.

On every trial, participants first saw a trial-unique image (either an indoor or outdoor scene) for 2 s. They were then asked to estimate how much that type of scene is worth on average in that room, from 1 to 100 cents. Participants had a maximum of 5 s to respond, explicitly providing their expected value for that scene category. The image was then presented again along with its true associated reward for 3 s (see Fig. 1). Participants were told that their payment for participating in the experiment would be contingent on the true reward amounts (they received approximately 1/4 of the rewards), not on the accuracy of their estimates.

The three experiments differed in the reward distributions assigned to the scene categories as well as the number of trials in the experiment. In experiment 1, participants experienced 8 trials of each scene category in each of two rooms (16 trials of each category total). In experiments 2 and 3, participants experienced 15 trials of each scene category in each of two rooms (30 trials of each category total). Experiments 2 and 3 were therefore twice as long (60 trials) as experiment 1 (32 trials).

In experiments 1 and 2, the "low-value scene" was worth 40¢ on average (Exp 1 high-risk-low-value rewards—0¢, 20¢, 60¢, 80¢; Exp 1 low-risk-low-value rewards—25¢, 35¢, 45¢, 55¢, twice each; Exp 2 high-risk-low-value rewards—0¢, 20¢, 40¢, 60¢, 80¢; Exp 2 low-risk-low-value rewards—20¢, 30¢, 40¢, 50¢, 60¢, three times each), and the "high-value scene" was worth 60¢ (Exp 1 high-risk-high-value rewards—20¢, 40¢, 80¢, 100¢; Exp 1 low-risk-high-value rewards—45¢, 55¢, 65¢, 75¢, twice each; Exp 2 high-risk-high-value rewards—20¢, 40¢, 60¢, 80¢, 100¢; Exp 2 low-risk-high-value rewards—40¢, 50¢, 60¢, 70¢, 80¢, three times each), whereas in experiment 3, the reward distributions of the scene categories were further apart and not overlapping: the low-value scene was worth 20¢ on average (high-risk-low-value rewards—0¢, 10¢, 20¢, 30¢, 40¢; low-risk-low-value rewards—10¢, 15¢, 20¢, 25¢, 30¢, three times each), and the high-value scene was worth 80¢ (high-risk-high-value rewards—60¢, 70¢, 80¢, 90¢, 100¢; low-risk-high-value rewards—70¢, 75¢, 80¢, 85¢, 90¢, three times each). The value of the scene category (high versus low value) was randomly assigned to scene category type (indoor or outdoor images).

Risk assessment After learning, participants completed a risk attitude questionnaire (DOSPERT, Weber et al. 2002) that created a 5–10-min delay between the learning and memory blocks of the experiment.

Memory Participants were presented with a surprise memory test, in which they judged whether different images were "old" (they had seen during learning) or "new" (they had not seen) as well as their confidence for that memory judgment (from 1 "guessing" to 4 "completely certain"). Half of the test trials were old, and half were new.

Depression measure At the end of the experiment, participants completed the Inventory of Depressive Symptomatology (IDS; Rush et al. 1996).

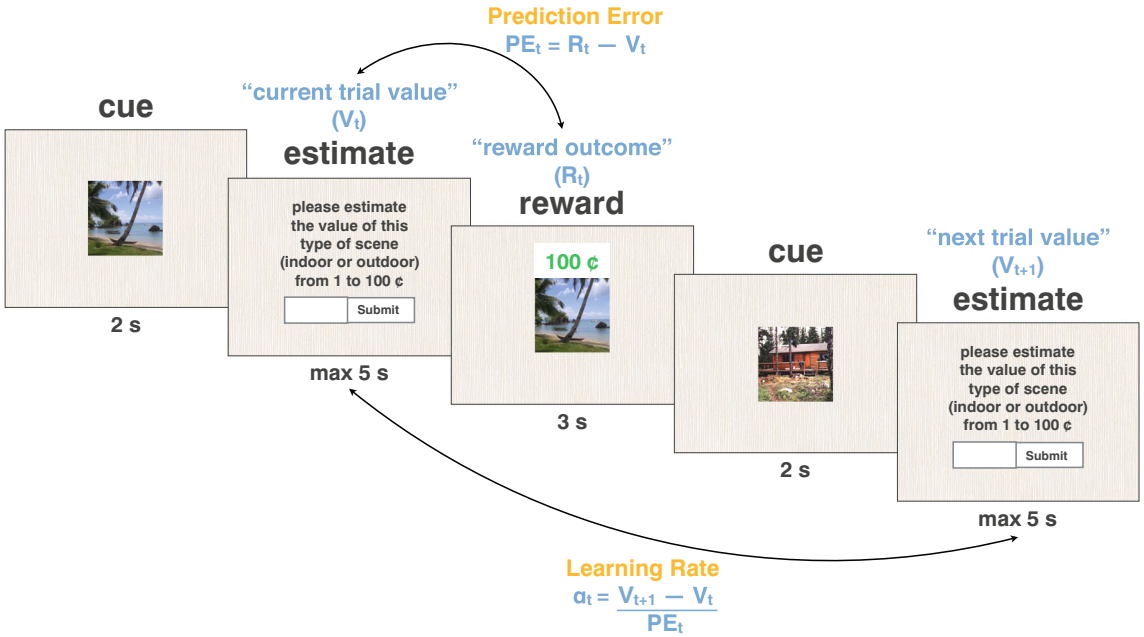

Fig. 1 Trial schematic during learning (from Rouhani et al. 2018). Each trial begins with a trial-unique indoor or outdoor image (cue), and participants are asked to estimate how much that scene category is worth on average (estimate). After entering their estimate, they see the image again with its outcome (reward). The prediction error is calculated by subtracting the value estimate of that scene category with the reward received on that trial. The learning rate is calculated over two consecutive trials of the same scene category (i.e., we assumed that separate values were learned for each scene category)

Statistical analyses

We investigated whether depression modulated learning and memory performance across all experiments. To do this, we compared participants who scored from moderate to very severe on the IDS (score—26–84, which we refer to as "depressive," $N = 101$) with participants who reported low or no depressive symptoms (score—0–13, which we refer to as "non-depressive," $N = 184$). Participants with an intermediate "mild" IDS score (14–25, $N = 98$) were excluded from the analysis (for categorization of scores, see www.ids-qids.org).

All comparisons were conducted using the linear or generalized mixed-effects models (R lme4 package, Bates et al. 2015), with experiment as a random effect and subject as a nested random effect within experiment (for both intercept and slope) and trial-unique scene image as a random effect (for intercept). We used depression category (depressive or not) as a fixed or interacting effect to predict the below learning and memory measures. We additionally tested whether depression severity predicted the effects under question within the depressive group.

If depression was a significant predictor, to confirm group differences, we ran a simplified regression model (not including depression as an effect) separately within each experiment and depression group and extracted subject-level intercepts and slopes. We then ran an ANOVA on the average difference in intercept and slope estimates between the "depressive" and "non-depressive" participants across all experiments. Finally, we corrected for multiple comparisons using Bonferroni correction.

Learning As a measure of learning, we took the absolute deviation of participants' trial-by-trial estimates from the true average values of the scene categories (40¢ or 60¢ in Exp 1–2; 20¢ or 80¢ in Exp 3). This deviation should decrease as participants learn the average values of the scene categories. In other words, with every trial, the learner should be estimating closer to the true mean of that scene category, and so a significant effect of trial number in decreasing this measure reflects learning. We ran two models testing (1) whether depression predicted overall deviation from the true means and (2) whether depression interacted with trial number, indicating that depressed participants learned differently than non-depressed participants.

Prediction errors Trial-by-trial prediction errors were calculated by subtracting participants' value estimates from the reward outcome experienced on that trial (see Fig. 1). We ran two models testing (1) whether depression predicted the average prediction error experienced during learning and (2) whether depression interacted with our previously reported finding that prediction errors are modulated by an interaction between risk context and scene category value, leading to greater underestimation of the high-value scene category and greater overestimation of the low-value scene category in the high-risk room.

Learning rates We calculated trial-by-trial learning rates as the proportion of the prediction error experienced on one trial that was then applied to update the value estimate on the next trial involving the same scene category (see Fig. 1). We ran four models testing (1) whether depression modulated the average learning rate applied during learning, (2) whether depression interacted with our previous findings that unsigned prediction errors increase learning rate, (3) whether depression interacted with our previous finding that a lower risk context leads to higher learning rates, and finally, (4) whether depression more specifically modulated an interaction between learning rate, unsigned prediction error, and the valence of the prediction error (positive or negative) in a 3-way interaction; for example, surprising negative (versus positive) events could lead to higher learning rates (i.e., more value updating) in participants with depression.

Memory We evaluated whether depression influenced item recognition by running the following mixed-effects logistic regressions predicting a "hit" or a "miss" during the memory test. We tested (1) whether depression affected overall memory, (2) whether depression interacted with the valence of the prediction error to influence memory; for example, by promoting negative prediction error memories over positive ones, (3) whether depression interacted with our previously reported finding that unsigned prediction errors increase memory, and (4) whether depression more specifically modulated an interaction between memory, the valence of the prediction error, and absolute prediction error; for example, by selectively enhancing surprising negative events in memory over surprising positive ones.

Results

Sample Across all experiments, 184 participants scored within the "non-depressive" category (Exp 1—69, Exp 2—68, Exp 3—47), and 101 participants scored within the "depressive" category (Exp 1—51, Exp 2—32, Exp 3—18).

Learning The absolute deviation of participants' estimates from the true averages of the two scene categories decreased as a function of trial number, indicating learning of the values of the two scene categories within each room (model 1: $\beta = -0.05$, $t = -2.93$, $p = 0.004$). Depression did not predict participant estimates on average (model 1: $\beta = -0.05$, $t = -1.04$, $p = 0.30$) nor did it interact with learning (model 2: $\beta = -0.02$, $t = -0.61$, $p = 0.54$); see Fig. 2a–c. However, depression severity did predict an overall increase in estimation error within the depressive group (model 1: $\beta = 0.01$, $t = 3.44$, $p < 0.001$; Fig. 2d).

Prediction errors Depression did not predict participant prediction errors (model 1: $\beta = -0.00063$, $t = -0.033$, $p = 0.97$) nor did it interact with the effect of risk and scene category value on prediction errors (model 2: $\beta = -0.0062$, $t = -0.35$, $p = 0.73$); see Fig. 3.

Learning rates Trial-by-trial learning rates were similarly not predicted by depression (model 1: $\beta = -0.046$, $t = -1.53$, $p = 0.13$); depression did not interact with the increase in learning rate with unsigned prediction error (model 2: $\beta = 0.034$, $t = 1.053$, $p = 0.29$) nor did it interact with the effect of risk context on learning rate (model 3: $\beta = -0.020$, $t = -0.92$, $p = 0.36$). Finally, there was no effect of depression in an interaction between the unsigned value and the valence of the prediction error on learning rate (model 4: $\beta = 0.034$, $t = 1.053$, $p = 0.29$); see Fig. 4.

Memory Depression did not affect average recognition memory (model 1: $\beta = -0.019$, $z = -0.14$, $p = 0.89$). It did not interact with an effect of prediction error valence on memory in general (model 2: $\beta = 0.07$, $z = 0.53$, $p = 0.60$) nor with the effect of unsigned prediction error on memory (model 3: $\beta = 0.07$, $z = 1.02$, $p = 0.31$). However, we did find that depression modulated the interaction between the unsigned value and the valence of the prediction error on memory. In particular, "non-depressive" participants were more likely to remember more surprising positive events, while "depressive" participants were more likely to remember more surprising negative events, as predicted (model 4: $\beta = 0.31$, $z = 2.05$, $p = 0.040$; Fig. 5).

To confirm and further illustrate this effect (see "Statistical analyses" above for details), we found that an interaction between prediction error valence and depression group predicted the slope of the effect of unsigned prediction errors on memory ($F(1,283) = 16.95$, $p < 0.0001$). This interaction passed Bonferroni adjusted levels of $p = 0.004$ (alpha = 0.05/14 comparisons). Following up on this interaction, we tested for across and within-group differences. We found that depressive participants had higher slopes for negative prediction error events than non-depressive participants, $t(274.35) = 2.79$, $p = 0.0057$, whereas non-depressive participants had higher slopes for positive prediction error events than depressive participants, $t(139.66) = -4.46$, $p < 0.001$. Within the depressive group, there were significantly higher slopes for the negative prediction error events than positive ones, $t(100) = -4.04$, $p < 0.001$, and within the non-depressive group, the opposite was true, $t(183) = 2.04$, $p = 0.043$. We did not find the interaction to predict the intercept of this model ($F(1,283) = 1.25$, $p = 0.26$); see Fig. 6.

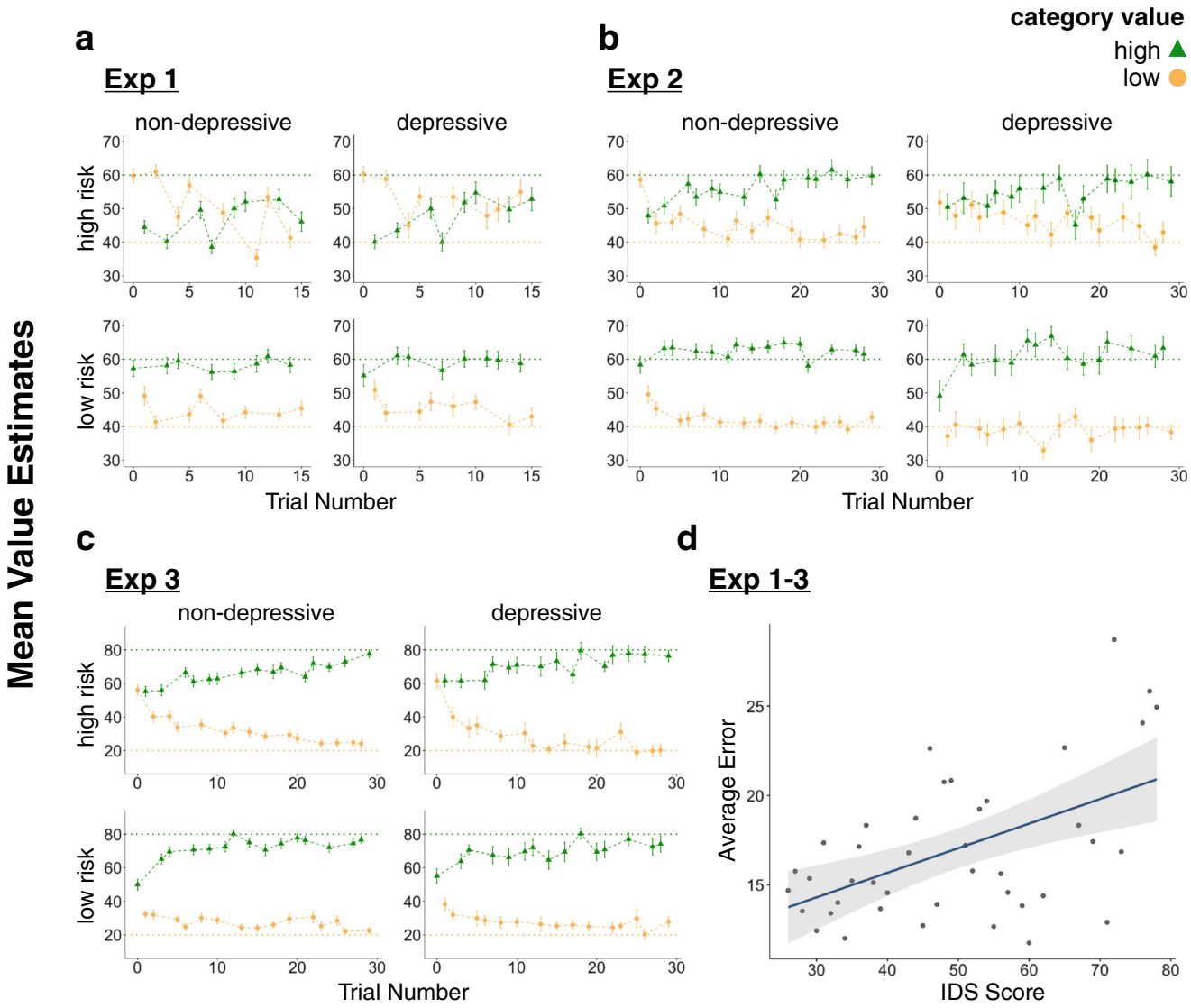

Fig. 2 **a**–**c** Learning. Average value estimates for high- and low-value scene categories as a function of trial number, within high- and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. We did not find any significant differences in value learning between the depressive and non-depressive groups. Note that "Trial 0" represents participant estimation at the beginning of each room and without having received any feedback. Error bars represent SEM. **d** Average estimation error during learning as a function of IDS score in the depressive group. Depression severity predicted greater average error during learning. Each dot represents a participant; shaded regions represent 95% confidence intervals

Discussion

Depressive symptoms include a diminished ability to feel pleasure (anhedonia) as well as excessive negative affect, thereby suggesting abnormalities in learning from rewards as well as their effect on memory. In a non-clinical sample, we tested for differences in reward learning and memory performance in individuals with and without depressive symptoms. We did not find gross differences in learning performance between the two groups: subjects did not differ in how they learned the average values of two scene categories, as measured by their trial-by-trial estimates, prediction errors, and learning rates throughout the task. However, we did find that depression severity predicted greater estimation error (i.e., difference between the estimated value and the true mean value of the scene categories) in the depressive group, which suggests impaired reinforcement learning in individuals with more severe depression. Nevertheless, we found that both groups increased their learning rates after high-magnitude prediction errors, and neither group showed a bias towards updating expectations more after a positive or negative prediction error event. Together, these results suggest that dopaminergic prediction error signaling was relatively intact throughout our non-clinical sample.

We also did not find overall differences in memory performance nor in memory for positive versus negative prediction

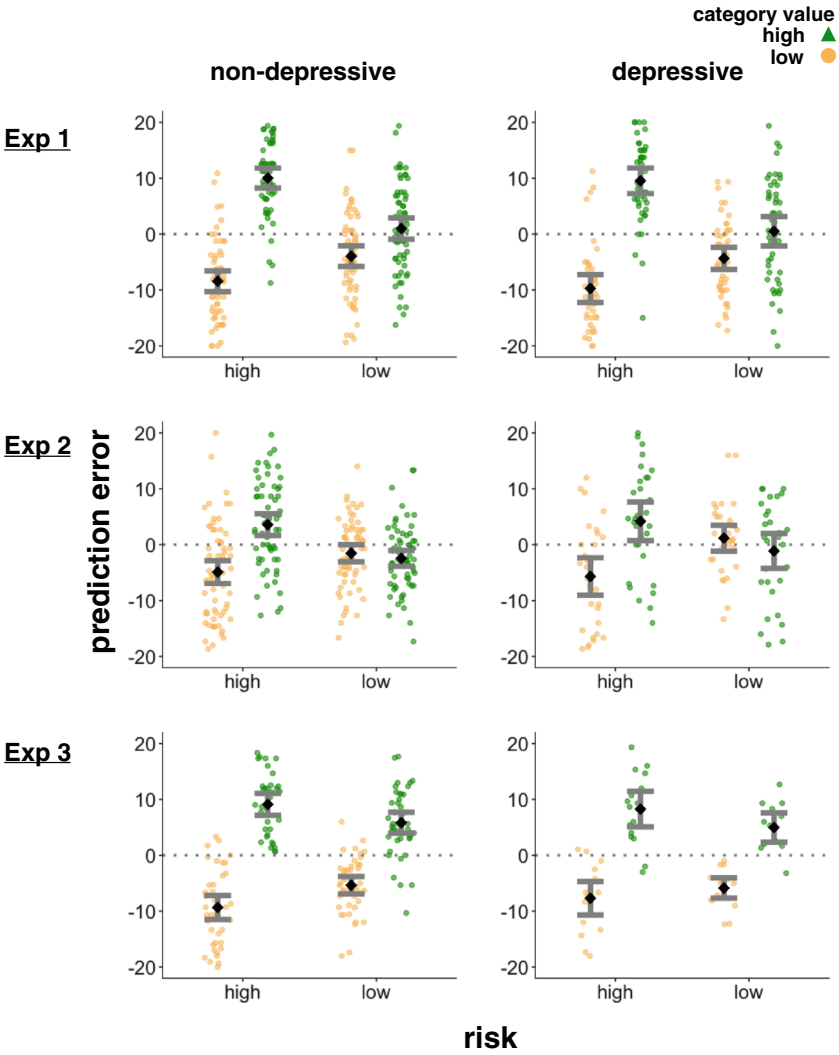

Fig. 3 Prediction errors. Experienced prediction errors for high- and low-value scene categories within high- and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. There is an overall interaction between risk and category value, such that participants are more likely to overestimate the low-value category and underestimate the high-value category in the high-risk room. There were no differences between the depressive and non-depressive groups. Error bars represent SEM

error events, on average. Instead, we found that the modulation of memory by reward prediction errors was differently biased in the two groups such that in individuals with depressive symptoms, large negative prediction errors enhanced memory to a greater extent than did large positive prediction errors and more so than they did in the non-depressive group. The opposite was true for non-depressive individuals: here, large positive prediction errors enhanced memories more than large negative prediction errors and more so than they did in the depressive group.

Relatively intact learning in the depressive group is consistent with several studies that have not found strong behavioral differences between MDD patients and healthy controls in reward learning (Chase et al. 2010; Knutson et al. 2008; Smoski et al. 2009; Ubl et al. 2015). In our sample, however, the positive relationship between depression severity and task error suggests that reinforcement learning is affected in depression, but only in more severe cases, which may in part explain the heterogeneity of results in the literature. Moreover, we implemented a Pavlovian reward learning paradigm that did not involve choices between differently rewarding options. This leaves open the possibility that depression is a greater modulator of instrumental learning than it is of prediction learning. Finally, given the striatal hypoactivity commonly reported in depression (for a review, see Admon and Pizzagalli 2015), it is possible that depressive individuals are not as *affectively* influenced by reward, meaning they may not feel its associated pleasure or impact, even if they are unimpaired in following explicit task goals by using rewards to update the values of their experiences.

On the other hand, the surprise recognition memory test provides a measure unrelated to explicit task goals potentially capturing affect-driven cognitive biases in depression. Here, we did not find a general difference in memory for events associated with positive versus negative prediction errors between the depressive and non-depressive groups. We instead found a bias in the unsigned prediction error modulation of memory. This signal, which increases memory for surprising outcomes (Rouhani et al. 2018), more significantly modulated memory for negative prediction error events in the depressive

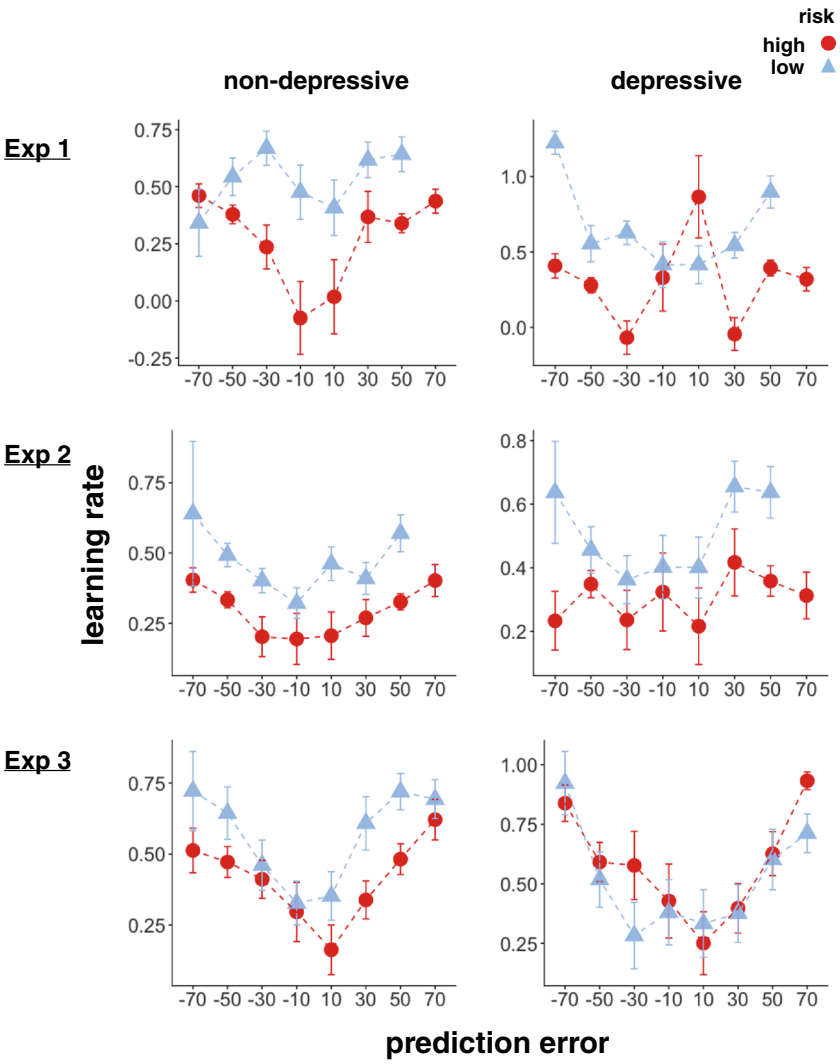

Fig. 4 Learning rates. Average learning rates for outcomes as a function of prediction error in high- and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. High-magnitude prediction errors increase learning rate across all experiments and groups. We did not find any differences between the depressive and non-depressive groups. Error bars represent SEM

group and positive prediction error events in the non-depressive group. In other words, depressive individuals were more likely to remember surprising negative events, whereas healthy individuals were more likely to remember positive ones. Such a bias in memory is in line with the tendency to ruminate on negative events in depression and provides evidence that surprising negative (versus surprising positive) events are indeed prioritized in memory.

There are several mechanisms that could contribute to the better encoding of surprising negative events in depressive individuals. Unsigned prediction errors are known to increase arousal and deploy the LC-norepinephrine system (Nassar et al. 2012), which co-releases dopamine signals that induce hippocampal plasticity (Kempadoo et al. 2016; Takeuchi et al. 2016) and enhance episodic memory (Clewett et al. 2018). Our results therefore suggest that LC activity is modulated more by surprising negative events in depressive individuals and by surprising positive events in healthy individuals. Given projections between the LC and regions within the salience network, such as the anterior cingulate cortex and amygdala, previous work lends support to this hypothesis: depressive individuals show greater striatal-cingulate functional connectivity (Admon et al. 2015) and more amygdala-modulated memory for negative versus positive events, whereas the opposite pattern is true for healthy controls (Leal et al. 2014; Young et al. 2016; Young et al. 2017).

Interestingly, in another line of work, the lateral habenula, which is associated with the processing of negative prediction errors (Matsumoto and Hikosaka 2007), has been strongly implicated in modulating symptoms of depression (Yang et al. 2018). This link suggests that greater activity of the lateral habenula in depressive individuals may support the mnemonic bias towards negative prediction error events. Future neuroimaging work should characterize how unsigned prediction errors differentially modulate memory for negative versus positive prediction error events in depression.

An alternative explanation is that an attentional bias for negative events (Gotlib and Joormann 2010; Joormann and Quinn 2014) leads depressive participants to spend more time looking at scenes associated with strong negative prediction

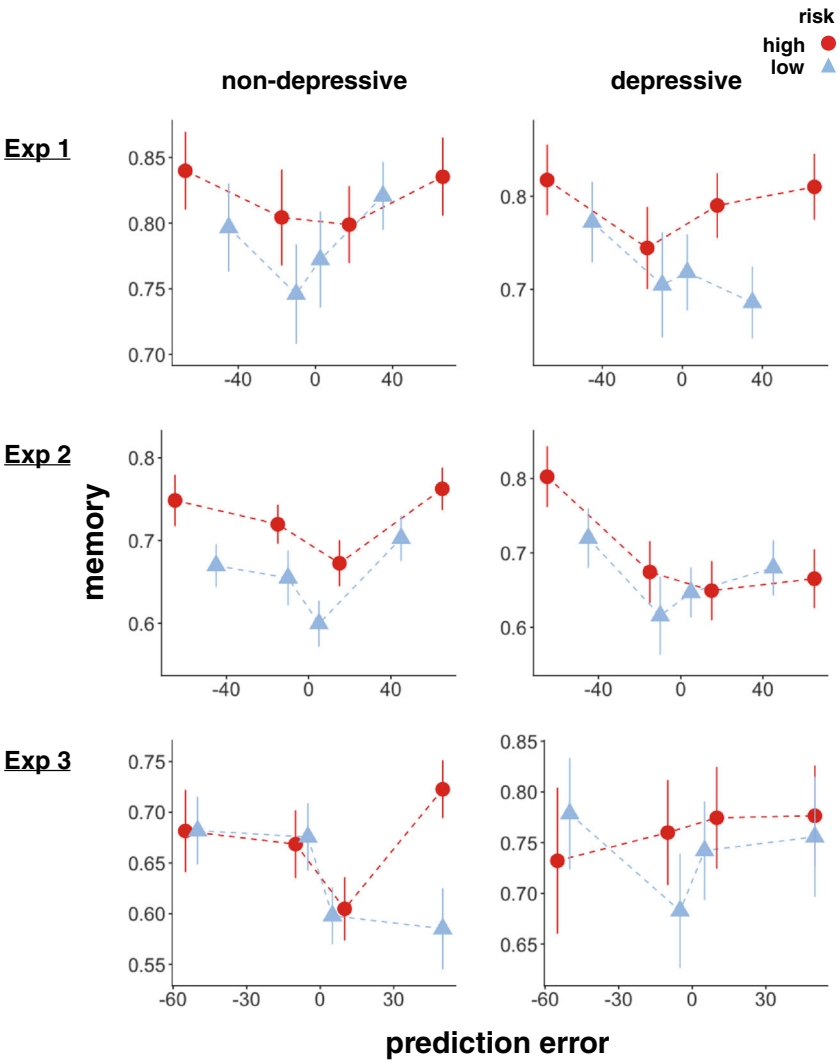

Fig. 5 Memory. Binned item recognition memory as a function of prediction error in high- and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. Item memory was binned by the quartile values of prediction errors within each room to illustrate the effects of prediction errors on memory; each dot represents the average value within that quartile. Note that no statistics were run on the binned values, and they are plotted only to illustrate the mixed-effects regression modeling. High-magnitude prediction errors increased item recognition memory across all experiments and groups. There were no overall differences in memory between depression groups. However, there was a three-way interaction between the unsigned prediction error, the valence of the prediction error, and depression group, such that depressive participants are more likely to remember high-magnitude, negative prediction error items, whereas non-depressive participants are more likely to remember high-magnitude, positive prediction error items. Error bars represent SEM

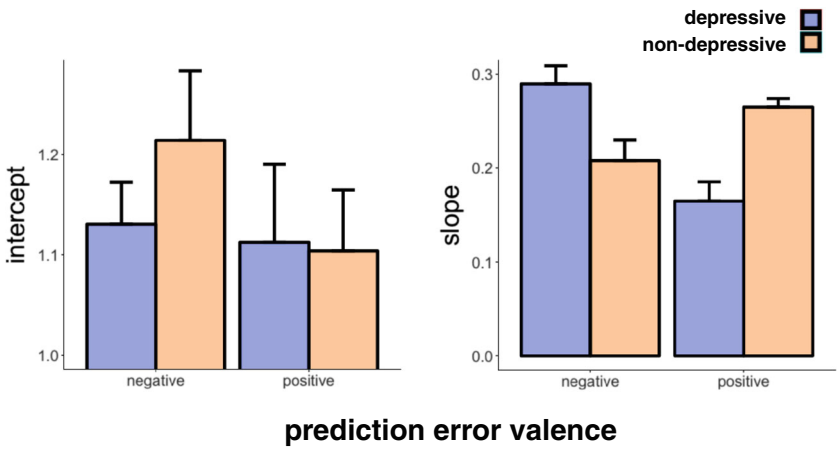

Fig. 6 Intercept and slope values for the unsigned prediction error effect on memory. Mixed-effects logistic regression models were run separately for positive and negative prediction error outcomes in the depressive and non-depressive groups. Bar plots represent average intercept value (left) and slope value (right) as a function of the valence of the prediction error and depression group. There were no differences in the intercept value, but we found an interaction in the slope of this effect (representing the degree to which unsigned prediction errors are improving item memory), such that unsigned prediction errors are boosting memory more so for negative events in depressive individuals and for positive events in non-depressive individuals. Error bars represent SEM

errors (and thereby encoding them in memory). Future studies could test this by using eye tracking as a measure of attention.

Our study has several limitations. First, our depressive group was not a clinical sample, and our findings need to be tested specifically in patients suffering from MDD. Moreover, given the heterogeneity of symptoms in MDD, future studies should take additional measures to allow testing for the modulation of the interaction between learning and memory by the severity of symptoms such as anhedonia, rumination, and anxiety. In particular, anhedonia has been shown to impair reward learning performance regardless of depression severity (Admon and Pizzagalli 2015; Chase et al. 2010; Gradin et al. 2011) and can similarly desensitize individuals to negative outcomes, whereas anxiety increases sensitivity to negative outcomes (Mueller et al. 2015). Individual measures of depressive symptoms along with co-morbid symptoms of anxiety could provide a better picture of which aspects of the disorder are giving rise to the biases in memory. We additionally did not collect medication information so could not test for the potential effects of neuroactive substances on performance.

Nevertheless, it is notable that 26% of our Amazon Mechanical Turk (mTurk) sample scored moderately to severely depressed. This is consistent with a recent finding that depression is two to three times higher in mTurk workers (under 50 years old) than matched national samples (Walters et al. 2018). This further suggests that researchers can characterize or, alternatively, need to control for the effects of depression in their mTurk experiments. In conclusion, our findings, in a heterogenous, online, non-clinical population, are fully in line with previous literature, suggesting that our task, and the interactions we found between learning and memory, may prove useful in clinical settings as well.

Funding information This work was supported by grant W911NF-14-1-0101 from the Army Research Office (Y.N.), the Ellison Foundation (Y.N.), grant R01MH098861 from the National Institute for Mental Health (Y.N.), and the National Science Foundation's Graduate Research Fellowship Program (N.R.).

Compliance with ethical standards

Participants completed informed consent online and were required to correctly answer questions checking for their understanding of the task before proceeding; procedures were approved by Princeton University's Institutional Review Board.

Conflict of interest The authors declare that they have no conflict of interest.

References

Admon R, Pizzagalli DA (2015) Dysfunctional reward processing in depression. Curr Opin Psychol 4:114–118. https://doi.org/10.1016/j.copsyc.2014.12.011

Admon R, Nickerson LD, Dillon DG, Holmes AJ, Bogdan R, Kumar P, Dougherty DD, Iosifescu DV, Mischoulon D, Fava M, Pizzagalli DA (2015) Dissociable cortico-striatal connectivity abnormalities in major depression in response to monetary gains and penalties. Psychol Med 45(1):121–131. https://doi.org/10.1017/S0033291714001123

Barto AG (1995) Adaptive critic and the basal ganglia. In: Houk JC, Davis JL, Beiser DG (eds) Models of information processing in the basal ganglia. MIT press, p 215

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, … Grothendieck G (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

Bornstein AM, Khaw MW, Shohamy D, Daw ND (2017) Reminders of past choices bias decisions for reward in humans. Nat Commun 8:15958. https://doi.org/10.1038/ncomms15958

Burt DB, Zembar MJ, Niederehe G (1995) Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. Psychol Bull 117(2):285–305. https://doi.org/10.1037/0033-2909.117.2.285

Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW (2010) Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. Psychol Med 40(3):433–440. https://doi.org/10.1017/S0033291709990468

Clewett D, Huang R, Velasco R, Lee T-H, Mather M (2018) Locus coeruleus activity strengthens prioritized memories under arousal. J Neurosci 2097–17. https://doi.org/10.1523/JNEUROSCI.2097-17.2017

Dillon DG, Pizzagalli DA (2018) Mechanisms of memory disruption in depression. Trends Neurosci 41:137–149. https://doi.org/10.1016/j.tins.2017.12.006

Dillon DG, Dobbins IG, Pizzagalli DA (2014) Weak reward source memory in depression reflects blunted activation of VTA/SN and parahippocampus. Soc Cogn Affect Neurosci 9(10):1576–1583 Retrieved from http://scan.oxfordjournals.org/. Accessed 31 Oct 2018

Duncan KD, Shohamy D (2016) Memory states influence value-based decisions. J Exp Psychol Gen 145(9):3–9

Duszkiewicz AJ, Mcnamara CG, Takeuchi T, Genzel L (2018) Novelty and dopaminergic modulation of memory persistence: a tale of two systems. Trends Neurosci. https://doi.org/10.1016/j.tins.2018.10.002

Gaddy MA, Ingram RE (2014) A meta-analytic review of mood-congruent implicit memory in depressed mood. Clin Psychol Rev 34:402–416. https://doi.org/10.1016/j.cpr.2014.06.001

Gershman SJ, Daw ND (2017) Reinforcement learning and episodic memory in humans and animals: an integrative framework. Annu Rev Psychol 68:101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gotlib IH, Joormann J (2010) Cognition and depression: current status and future directions. SSRN. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, Reid I, Hall J, Steele JD (2011) Expected value and prediction error abnormalities in depression and schizophrenia. Brain 134(6):1751–1764. https://doi.org/10.1093/brain/awr059

Huys QJ, Pizzagalli DA, Bogdan R, Dayan P (2013) Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biol Mood Anxiety Disord 3(1):12. https://doi.org/10.1186/2045-5380-3-12

Joormann J, Quinn ME (2014) Cognitive processes and emotion regulation in depression. Depress Anxiety 31(4):308–315. https://doi.org/10.1002/da.22264

Kempadoo KA, Mosharov EV, Choi SJ, Sulzer D, Kandel ER (2016) Dopamine release from the locus coeruleus to the dorsal

hippocampus promotes spatial learning and memory. Proc Natl Acad Sci 113(51):14835–14840. https://doi.org/10.1073/pnas.1616515114

Knutson B, Bhanji JP, Cooney RE, Atlas LY, Gotlib IH (2008) Neural responses to monetary incentives in major depression. Biol Psychiatry 63(7):686–692. https://doi.org/10.1016/j.biopsych.2007.07.023

Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, Brooks NH, Pizzagalli DA (2018) Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology 43(7):1581–1588. https://doi.org/10.1038/s41386-018-0032-x

Leal SL, Tighe SK, Jones CK, Yassa MA (2014) Pattern separation of emotional information in hippocampal dentate and CA3. Hippocampus 24(9):1146–1155. https://doi.org/10.1002/hipo.22298

Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447(7148):1111–1115. https://doi.org/10.1038/nature05860

Matt GE, Vázquez C, Campbell WK (1992) Mood-congruent recall of affectively toned stimuli: a meta-analytic review. Clin Psychol Rev 12(2):227–255. https://doi.org/10.1016/0272-7358(92)90116-P

Mueller EM, Pechtel P, Cohen AL, Douglas SR, Pizzagalli DA (2015) Potentiated processing of negative feedback in depression is attenuated by anhedonia. Depress Anxiety 32(4):296–305. https://doi.org/10.1002/da.22338

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15(7):1040–1046. https://doi.org/10.1038/nn.3130

Pizzagalli DA (2014) Depression, stress, and anhedonia: toward a synthesis and integrated model MDD: major depressive disorder. Annu Rev Clin Psychol 10:393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Pollack I, Norman DA (1964) A non-parametric analysis of recognition experiments. Psychon Sci 1(1–12):125–126. https://doi.org/10.3758/BF03342823

Read Montague P, Dayan P, Sejnowskw TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 76. Retrieved from http://www.jneurosci.org/content/jneuro/16/5/1936.full.pdf. Accessed 31 Oct 2018

Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G (2012) Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. Eur J Neurosci 35(7):1190–1200. https://doi.org/10.1111/j.1460-9568.2011.07986.x

Rouhani N, Norman KA, Niv Y (2018) Dissociable effects of surprising rewards on learning and memory. J Exp Psychol Learn Mem Cogn. https://doi.org/10.1037/xlm0000518

Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH (1996) The inventory of depressive symptomatology (IDS): psychometric properties. Psychol Med 26(03):477–486. https://doi.org/10.1017/S0033291700035558

Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, Lam J, Skandali N, Siegel JZ, Ousdal OT, Prabhu G, Dayan P, Fonagy P, Dolan RJ (2017) Association of neural and emotional impacts of reward prediction errors with major depression. JAMA Psychiatry 74:10–12. https://doi.org/10.1001/jamapsychiatry.2017.1713

Sacchet MD, Levy BJ, Hamilton JP, Maksimovskiy A, Hertel PT, Joormann J, Anderson MC, Wagner AD, Gotlib IH (2017) Cognitive and neural consequences of memory suppression in major depressive disorder. Cogn Affect Behav Neurosci 17(1):77–93. https://doi.org/10.3758/s13415-016-0464-x

Sara SJ (2009) The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 10:211–223. https://doi.org/10.1038/nrn2573

Shohamy D, Adcock RA (2010) Dopamine and adaptive memory. Trends Cogn Sci 14:464–472. https://doi.org/10.1016/j.tics.2010.08.002

Smoski MJ, Felder J, Bizzell J, Green SR, Ernst M, Lynch TR, Dichter GS (2009) fMRI of alterations in reward selection, anticipation, and feedback in major depressive disorder. J Affect Disord 118:69–78. https://doi.org/10.1016/j.jad.2009.01.034

Takeuchi T, Duszkiewicz AJ, Sonneborn A, Spooner PA, Yamasaki M, Watanabe M, Smith CC, Fernández G, Deisseroth K, Greene RW, Morris RGM (2016) Locus coeruleus and dopaminergic consolidation of everyday memory. Nature 537(7620):1–18. https://doi.org/10.1038/nature19325

Ubl B, Kuehner C, Kirsch P, Ruttorf M, Diener C, Flor H (2015) Altered neural reward and loss processing and prediction error signalling in depression. Soc Cogn Affect Neurosci 10(8):1102–1112. https://doi.org/10.1093/scan/nsu158

Walters K, Christakis DA, Wright DR (2018) Are mechanical Turk worker samples representative of health status and health behaviors in the U.S.? PLoS ONE 13(6). https://doi.org/10.1371/journal.pone.0198835

Weber EU, Blais A-R, Betz NE (2002) A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. J Behav Decis Mak 15(4):263–290. https://doi.org/10.1002/bdm.414

Whitton AE, Kakani P, Foti D, Van't Veer A, Haile A, Crowley DJ, Pizzagalli DA (2016) Blunted neural responses to reward in remitted major depression: a high-density event-related potential study. Biol Psychiatry Cogn Neurosci Neuroimaging 1(1):87–95. https://doi.org/10.1016/j.bpsc

Wimmer GE, Shohamy D (2012) Preference by association: how memory mechanisms in the hippocampus bias decisions. Science 338:270–273. https://doi.org/10.1126/science.1223252

Yang Y, Cui Y, Sang K, Dong Y, Ni Z, Ma S, Hu H (2018) Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. Nature 554(7692):317–322. https://doi.org/10.1038/nature25509

Young KD, Siegle GJ, Bodurka J, Drevets WC (2016) Amygdala activity during autobiographical memory recall in depressed and vulnerable individuals: association with symptom severity and autobiographical overgenerality. Am J Psychiatry 173:78–89. https://doi.org/10.1176/appi.ajp.2015.15010119

Young KD, Siegle GJ, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC, Bodurka J (2017) Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. Am J Psychiatry 174:748–755. https://doi.org/10.1176/appi.ajp.2017.16060637