Encoding Motivation Prediction Errors in the Human Dopaminergic Reward System

Yinmei Ni
Beijing Normal University
Sidong Wang
Beijing Normal University
Jie Su
Beijing Normal University
Jian Li
Peking University https://orcid.org/0000-0002-3941-2622
Xiaohong Wan (xhwan@bnu.edu.cn)
Beijing Normal University https://orcid.org/0000-0002-6472-1435

Article

Keywords: Motivation, Dopamine, Reinforcement learning, Prediction error, Salience, Model-based control, Model-free learning, Ventral striatum, Primary motor cortex, fMRI

Posted Date: August 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-51287/v1

Abstract

The dopaminergic reward system encoding the reward PE signals is vital for reinforcement learning (RL). Although this reward PE hypothesis has been extensively validated, it remains considerable debates on the alternative account of motivation. In the current study, we diverted the participants' motivation from the conditioned stimulus (CS)-associated valences to the CS-elicited actions in a variant Pavlovian conditioning task under appetitive and aversive conditions. We found that the regions in the dopaminergic reward system did not encode such bidirectional reward PE signals, but the PE magnitudes, namely, the motivation PE signals. These neural signals without indicating the directions of learning could not be directly used for model-free RL, but probably for model-based control. Specifically, the ventral striatum during the feedback phase might encode the need of adjusting the learning policy, while the putative substantia nigra pars compacta (SNc) in the midbrain and the putamen during the prediction phase might sustain the intended actions. Meanwhile, the primary motor cortex encoded the salience PE signals for model-free RL. Therefore, our findings demonstrate that the human dopaminergic reward system could encode the motivation PE signals to substantialize model-based control, rather than model-free learning, suggesting that its involvement in RL should be motivation-dependent.

Introduction

Humans and animals are motivated to make precise predictions in uncertain environments[1]. The predictions are adaptively updated by a recursive process, which can be well characterized by the normative reinforcement learning (RL) theory[2,3]. The principle of such a model-free RL algorithm is that the temporal difference between the sequential predictions is proportional to the prediction error (PE), the discrepancy between the experienced and expected outcome (*i.e.*, delta-rule). The directions of learning concur with the signs of the PEs. This bidirectional regulation by the PEs allows the predictions to progressively converge to the actual outcomes.

A great number of neurophysiological and neuroimaging studies in animals and humans have demonstrated that such a neurocomputational algorithm could be implemented in the dopaminergic reward system (Fig. 1a), particularly in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) of the midbrain[4–11]. For instance, in the Pavlovian conditioning tasks, the dopaminergic neural activity encodes the predicted reward when the conditioned stimulus (CS) is presented. Critically, the dopaminergic neural activity also represents the bidirectional reward PE signals (Fig. 1b), immediately after the reward is delivered or omitted[4–8], while the measures by functional magnetic resonance imaging (fMRI) could also detect such phasic neural activities[9–12]. These bidirectional PE signals are then used to retroactively update the predicted reward associated with the CS[13]. Recently, it has been substantially demonstrated that an artificial dopamine PE signal generated by optogenetic stimulations are sufficient to cause a change of the CS-associated value by RL[14–16]. On the other hand, the dopaminergic reward system could also encode the punishers and the aversive PEs[7,9,17–19]. Thus, the dopaminergic reward system might broadly represent saliences and salience PEs (Fig. 1b), regardless of the signs of valences.

Nonetheless, the salience PE signals are bidirectional too, thereby, could be still directly used to update the predictions as suggested by the RL theory.

On the contrary, the dopaminergic reward system has been argued to be alternatively associated with motivation[20], the drive of controlling the intended actions (Fig. 1a). In most of the instrumental conditioning tasks (Fig. 1c right), the motivation in control of an action and the expected reward of the action, however, are highly converged. For example, when an action is predicted to be associated with a higher reward, participants are more motivated to pursue such an action[21,22]. This makes it difficult to clearly distinguish the two alternative hypotheses. In contrast, the performance of an action is not necessary for the outcome or unconditioned stimulus (US) delivery in the traditional Pavlovian setting (Fig. 1c left). However, an extra CS-elicited action could be operationally dissociated from the CS-associated valence in a variant Pavlovian setting (Fig. 1c middle), differing strikingly from the instrumental conditioning. Thus, this new Pavlovian setting provides us an opportunity to test whether the dopaminergic reward system could still normatively encode the bidirectional PE signals that are necessary for updating the CS-associated valences, or instead encode the motivational signals to control the CS-elicited actions (Fig. 1a), while the two factors were orthogonal in this Pavlovian setting.

In the current study, we directly addressed this critical issue using fMRI with this new paradigm in a Pavlovian conditioning task under both appetitive and aversive conditions, where the participants explicitly reported the predicted valences associated with the CS by bimanually adjusting the cursor position prior to the outcome delivery. We found that the neural activities in the human dopaminergic reward system, including the ventral striatum (VS), the putamen, and the putative SNc, all selectively encoded the unsigned PEs (UPEs), rather than the reward PEs or the salience PEs [both are signed PEs (SPEs)]. These neural signals without indicating the directions of learning cannot be directly used for RL. Instead, they might serve as phasic motivational signals in control of RL in association with the CS-elicited actions.

Results

Task paradigm and behavioral results

During a variant Pavlovian conditioning task (Fig. 2a), two different cues, one associated with rewards (the gain condition) and another associated with punishers (the loss condition), were randomly interleaved (Fig. 2b). The specific task sequence with random alternation of the two conditions made it possible to temporally separate the prediction phase from the feedback phase. About two-thirds of all the trials with the same CS were interleaved and thus non-contiguous, while the left contiguous trials (about one-third) had a long ITI (3–7 s). Hence, it became possible to examine the neural activities during the feedback phase of the current trial and those during the prediction phase of the subsequent trial with the same CS (see below). The trial-by-trial outcomes associated with each CS were stochastically varied following beta distributions with different means across blocks (see Methods). The critical change of the

l Pavlovian conditioning tasks was that when the CS was

presented the participants needed to explicitly report their prediction about the CS-associated valence via a combination of several button presses to move the cursor position after the CS presentation in each trial, other than only passively viewing the CS presentation. The right button presses increased and the left button presses decreased the prediction magnitude. The default position of the cursor was always at the prediction reported in the previous trial with the same cue or at the central position for the first trial of each run. Thereby, the participants moved the cursor in regarding with the prediction change between the consecutive trials with the same CS. The participants also immediately reported their confidence in their predictions. Differing strikingly from the instrumental conditioning tasks, the actions of reporting the predictions in the current Pavlovian setting did not affect the potential rewards or punishment associated with the CSs, one of which was randomly chosen from all the trials after the experiment was completed. To generally invigorate the participants to engage in reporting the predictions, an additional fixed bonus would be given for their good performance.

By virtue of the reported predictions, we could precisely measure the trial-by-trial PEs and learning effects (*e.g.*, learning rates). Due to the volatility of the environments, the participants continuously kept learning from the outcomes across all the trials in each run (Fig. 3a). The prediction update in the subsequent trial was linearly proportional to the PE, the discrepancy between the actual outcome and the reported prediction in the current trial (Fig. 3b; but see below). That is, the participants (n = 33) updated their predictions largely following the RW delta-rule[2,3]. The learning effects were similar between the gain and loss conditions (Fig. 3c). Nonetheless, significantly positive biases of their predictions and larger confidence ratings in the gain condition, relative to the loss condition, consistently showed that the participants displayed optimism bias[23], suggesting that they should treat the CS-associated numerical values as valences (Supplementary Fig. 1b-d).

The fMRI activities in the dopaminergic reward system was associated with the UPEs, not the SPEs

According to the prior work in neurophysiology[4–8] and neuroimaging[9–11,17–19], the neural activities in the striatum (*e.g.*, VS) and the midbrain regions should encode the SPEs during the feedback phase and the prediction values during the prediction phase underpinning RL. However, none of these predictions was observed in any region of the dopaminergic reward system. No voxels in the dopaminergic reward system showed their fMRI activities had significant correlation with the SPEs or the prediction values, during either phase of the gain or loss condition ($z < 1.96$, $P > 0.05$, uncorrected).

On the contrary, we found that robust fMRI activities in several regions of the dopaminergic reward system were significantly correlated with the UPEs: positively in the putamen and the putative SNc of the nigrostriatal system during the prediction phase (Fig. 4a), and negatively in the VS during the feedback phase (Fig. 5a) [$z > 3.1$, $P < 0.05$ after family-wise error (FWE) correction]. Notably, the UPEs and the SPEs were by nature uncorrelated (Pearson's $r \approx 0.01$, $P = 0.48$). Significant correlations with the UPEs, rather than the SPEs, were caused by asymmetrical distributions (Supplementary

Fig. 1a), nor by asymmetrical fMRI responses (Supplementary Fig. 2) between the positive and negative PEs, for either the gain or loss condition.

The discrepancy between the associations with the SPEs and those with the UPEs was essentially originated from the case of the negative PEs. Despite that it has been often difficult to detect the positive neural correlates with the negative PEs in the regions of the dopaminergic reward system using fMRI, such as the VS and the VTA/SNc regions[9,10,17,18], we here found robust negative neural correlates with the negative PEs in the putamen and the putative SNc regions, and positive neural correlates with the negative PEs in the VS (Supplementary Fig. 2).

The putative SNc in the midbrain and the putamen encoded the motivational signals supporting the action execution

We found that a number of voxels that were heterogeneously distributed in the midbrain region showed robust positive correlation with the positive PEs and/or robust negative correlation with the negative PEs during the prediction phase, but not during the feedback phase (Fig. 4b). However, no voxels in the midbrain region were identified as their fMRI activities were positively correlated with the negative PEs during either phase, even with different durations as the event period (see Methods). Those voxels showing significant activations across the four different conditions (the combinations of the positive/negative PEs and the gain/loss contexts) demonstrated partial overlap and convergence around the putative SNc region (Fig. 4c), where the neural responses were similar to those concurrently elicited in the putamen.

Specifically, even in the conventional regions of interest (ROIs) of the dopaminergic reward system, defined by voxels whose activities were significantly correlated with the positive PEs during the feedback phase in the gain condition ($z > 2.6$, $P < 0.005$, uncorrected), namely, encoding the reward PE signals[10,11,19], the fMRI activities were significantly correlated with the UPEs, but neither the reward PEs (Fig. 4c; Supplementary Fig. 2) nor the cue-associated predicted or outcome valences (Supplementary Fig. 3c-d). Critically, the neural regression values with the UPEs in these regions were equivalent between the gain and loss conditions (Fig. 4c; Supplementary Fig. 2), thereby, inconsistent with neural encoding of either the reward PEs or the salience PEs repeatedly observed in most previous studies.

It is conceivable that the selective activities associated with the UPEs in the putamen and the putative SNc during the prediction phase might be responsible for representing a salience signal for attentional learning of the CS-US associations[24,25]. However, the behaviors of the majority of participants consistently complied much better with the RW delta-rule than with the attentional-learning rule (Fig. 3d). Instead, the positive correlations with the UPEs in the putamen and the putative SN of the nigrostriatal system during the prediction phase suggest that these regions might encode a motivational signal for preparing and sustaining the intended actions of reporting the predictions, whereas their fMRI activities were not directly correlated with the frequency of button press in each trial.

The VS encoded the prediction certainty in evaluation of learning

We found that the VS activities were correlated negatively with the UPEs during the feedback phase (Fig. 5a; also in the amygdala, Supplementary Fig. 4). These VS activities were in stark contrast to the conventional observations that the striatum activities measured by fMRI track both the magnitudes and directions of the PEs (Fig. 1b). Instead, the negative correlation with the PE magnitudes suggests that the VS might encode the retrospective certainty about the preceding prediction when the actual outcome was received. A lower PE magnitude indicates a higher degree of prediction certainty[26]. The VS activities were close to zero when the UPEs were small (*i.e.*, high degrees of prediction certainty), but became more negative when the UPEs were larger (*i.e.*, low degrees of prediction certainty; the blue line in Fig. 5b). Thus, the VS seemed to be involved in retrospectively evaluating the learning process, probably encoding the drive to improve the prediction.

Due to the stochastic nature of the environmental changes, the participants actually trial-by-trial adaptively adjusted the learning rates in regarding with the PE magnitudes (Fig. 6a). The greater the UPEs were, the greater the learning rates were. Hence, the underlying RL process was actually deviating from model-free RL. As illustrated in our separate study[27], The brain implemented such a learning process by two separate modules in parallel. The primary motor cortex (PMC) implemented model-free RL process (see below), whereas the anterior cingulate cortex (ACC) implemented an adaptive process to compensate the rigidity of model-free RL in the face of the volatile environment. The fMRI activities in the ACC became greater when the PE magnitudes were larger (the red region in Fig. 6b). Thereby, the fMRI activities in the ACC were also positively correlated with the UPEs.

Intuitively, the negative correlation with the UPEs in the VS might represent the motivation to adjust the learning process that was implemented in the ACC. To test this hypothesis, we made a psychophysiological interaction (PPI) analysis using the VS as the seed to search the voxels across the whole brain whose fMRI activities were modulated by the interaction of the VS activities (the physiological factor) and the PE magnitudes (large *vs.* small; the psychological factor). We found that the fMRI activities in the region largely overlapping with the ACC region mentioned above were significantly modulated (the blue region in Fig. 6b). Specifically, the VS-ACC functional connectivity became more negative when the PE magnitudes became larger (Fig. 6c). Hence, the VS negative activities under the large UPEs seemed to disinhibit the ACC activities, eliciting its involvement in adjusting the learning policy, that is, increasing the learning rate for the subsequent trial with the same CS (Fig. 6a).

In contrast, the VS activities during the prediction phase were not further significantly correlated with the UPEs (the red line in Fig. 5b), but became positively correlated with the reported confidence, consisting with our previous study on the decision-making tasks[28]. Thus, the VS activities during the prediction phase continuously represented the subjective certainty of the prediction in the current trial (Fig. 5c), in particular, prospectively evaluating the prediction in prior to receiving the actual outcome. Consistently

between the two phases, the regression values associated with the UPEs during the feedback phase were highly correlated with those associated with the reported confidence during the prediction phase across all the participants ($r = -0.34$, $P = 0.023$, Fig. 5d), suggesting that the VS should consistently encode the prediction certainty in evaluation of the learning process and action performance.

The PMC contributed to model-free RL

These neural signals in the dopaminergic reward system encoding the UPEs without indicating the directions of learning obviously cannot be directly used to update the predictions as suggested by the normative RL theory[21,22]. Instead, we found that the fMRI activities in the bilateral PMC were prominently correlated with the SPEs during the feedback phase of both the gain and loss conditions. Due to that the opposite hands respectively would be used for the gain and loss conditions, the bilateral PMC activation patterns were reversed (Fig. 7a). Thus, the PMC seemed to encode the salience PEs, irrelevant to the signs of the CS-associated valences, but crucially dependent on the laterality of hands used in the near future. Notably, there was no any motor action during the feedback phase. Importantly, the participants' PMC regression strengths with the SPEs were significantly correlated with their RW learning rates for both the gain and loss conditions ($r = 0.48$, $P = 0.0018$). Hence, the neural signals in the bilateral PMC during the feedback phase were not caused by the current motor actions, could be used to update the predictions following the RW delta-rule and to prepare the motor actions in the subsequent trial[29–31]. In contrast to the ACC activities, The PMC activities were not modulated by the VS activities (Fig. 6c).

We further found that the same activation patterns in the bilateral PMC during the prediction phase in correlation with the SPEs as those during the feedback phase of both the gain and loss conditions (Fig. 7b), when the participants were making motor actions to report their predictions in reference to the prediction changes, which were proportional to the SPEs (Fig. 3a-b). Concurrently, the neural signals in the putamen and the putative SNc in the midbrain might support the intended motor actions that were concurrently executed in the PMC. Notably, both the CS-associated outcome valences during the feedback phase of the current trial and the CS-associated predicted valences during the prediction phase of the subsequent trial were also associated with the fMRI activities in the PMC and the primary visual cortex, other than the dopaminergic reward system (Supplementary Fig. 3a). However, both the regressions with the outcome valences during the feedback phase of the current trial and those with the predicted valences during the prediction phase of the subsequent trial were completely explained away by the SPEs at the current trial, as both the outcome valences at the current trial (mean $r = 0.57$) and the predicted valences at the subsequent trial (mean $r = 0.44$) were highly correlated with the SPEs at the current trial. Only the fMRI activities in the primary visual cortex still remained correlated with the outcome valences at the current trial and the predicted valences at the subsequent trial. These results further suggest that the model-free RL process should primarily occur in the sensorimotor cortical areas of the PMC and the primary visual cortex, rather than the dopaminergic reward system.

It has been well documented that the regions in the dopaminergic reward system play crucial roles in RL[4–12,17–19]. However, there are considerable debates on their exact functional roles. The neural activities in these regions have been argued to alternatively encode valences or motivation. Under instrumental conditioning paradigms, it is difficult to delineate the effects of the two alternative hypotheses due to their high correlation. In the current study, by asking participants to explicitly report their predictions about the CS-associated valences in a Pavlovian conditioning task, we created a new Pavlovian setting to plausibly test the critical issue about whether the neural activities in the dopaminergic reward system should encode the motivation for controlling the CS-elicited actions or the CS-associated valences and SPEs for model-free RL, since the two factors were orthogonal in the current task. Although the participants' behaviors appeared to follow the RW delta-rule, their neural responses measured by neuroimaging did not provide evidence for any region in the dopaminergic reward system encoding the SPEs for model-free RL as predominately observed in the literature[4–12,17–19]. On the contrary, the neuroimaging results revealed that several regions in the dopaminergic reward system including the VS, the putamen, and the putative SNc were robustly associated with the UPEs, negatively in the VS during the feedback phase and positively in the putamen and the putative SNc during the prediction phase (but see ref. 11). A serious consequence following this disparity with the normative findings in the literature is that the neural signals in these regions of the dopaminergic reward system without indicating the directions of learning could not be directly used to update the predictions, thereby, arguing against the automaticity and generality of encoding the reward (or salience) PE signals in the human dopaminergic reward system for model-free RL. In contrast, the bilateral PMC activities encoded the SPEs, probably substantializing model-free RL observed in behaviors. To the best of our knowledge, these findings in the current study provide the first direct evidence supporting that the human dopaminergic reward system could also encode motivation PE signals that are deviating from the dopamine reward or salience PE hypothesis posited by the RL theory, implicating that the human dopaminergic reward system should play much complex roles in RL.

Dopaminergic reward system might be not always necessary for model-free RL

The failures of finding that the dopaminergic reward system encoded the bidirectional PE signals in the current study would be first impressed by the account that the weak neural activities could not be sensitively detected by the fMRI measures with low signal-to-noise ratios, especially for the depressed activities when the actual outcomes were worse than the predictions[10,18] (*i.e.*, the negative PEs). However, in the current study, the VS and the nigrostriatal system including the putamen and the putative SNc were found to robustly decrease and increase their neural activities in response to negative PEs, respectively. Thereby, these neural activities detected by fMRI were indeed sensitive to the PE magnitudes, but not the PE directions. Therefore, our results showed that the neural activities in the dopaminergic reward system were largely irresponsible for model-free RL in the current Pavlovian setting. Although recent studies using optogenetic stimulations have provided clear evidence for the sufficiency of dopamine PE signals

for RL[14–16], to the best of our knowledge, the necessity of dopamine PE signals in RL has not yet been formally tested. On the contrary, the previous lesion and pharmaceutical studies have showed that the dopamine deficits do not affect learning, but rather action performance[20,32–34].

However, the reward PE hypothesis in the dopaminergic reward system has had a deep influence on the field of association learning. The neural correlates of the reward PEs or the salience PEs have been found in a variety of regions, such as the ACC[35], the medial and lateral orbitofrontal cortex[36], the insular[17], the amygdala[19,37], the primary visual cortex[38], the ventral and dorsal striatum[17–19, 39], the cerebellum[40], the periaqueductal gray[41], and currently the PMC, for either appetitive or aversive condition. Most of these neural signals have been simply thought to mirror the dopamine PE signals in the midbrain, even for non-valence association learning, given that the dopaminergic neurons have wide connections with most of the brain regions. However, it is worth noting that most of these inferences have been lacking direct evidence.

On the contrary, our findings in the current study alternatively opt for that the PE signals associated with the CS features might be not always necessarily originated from the dopaminergic reward system, as there were even no such PE signals encoded in these regions. Instead, it seems plausible that association learning on a particular CS-associated feature should recruit the neural locus specific to that associated feature to encode the PE signals for model-free RL. For instance, the PE signals for sensory features might be encoded in the relevant sensory system[38,42] (*e.g.*, the primary visual cortex), while those for CS-associated motor actions might be instead encoded in the motor system (*e.g.*, the cerebellum or the PMC). Only those for learning the CS-associated valences need to encode such PE signals for model-free RL in the dopaminergic reward system. In short, the model-free RL process could be implemented in a wide range of brain regions other than the dopaminergic reward system.

Dopaminergic reward system could encode phasic motivational signals

The extant neuroimaging studies often used the evidence of encoding the PE signals in the VS to demonstrate that the PE signals originated from the midbrain dopamine neurons are involved in RL, as the VS receives direct projection from the midbrain dopamine neurons[18,39]. However, our results showed that the VS activities were temporally and functionally dissociated from the activities in the midbrain (VTA and SNc), consisting with the findings in previously electrophysiological and voltammetrical studies in animals[43,44] and neuroimaging studies in humans[45]. Instead of encoding the reward or salience PE signals for model-free RL, our current findings support the notion that the VS should, in general, represent motivational signals[46,47]. The motivation PE signals encoded in the VS during the feedback phase specifically modulated the adjustment of learning policy that was implemented in the ACC, but did not affect the model-free RL process implemented in the PMC. This is consistent with the previous findings that the VS is not necessary for stable or model-free RL, or well-trained responses to CS, but considerably

affects learning the stochastic CS-reward associations or model-based RL[33,34]. The dopamine releases in the VS coincide with the progress of approaching the time- and effort-consuming goals[48,49], but independent of the midbrain dopamine spikes[45]. Taken together, our findings are in line with the perspective that the VS might represent the phasic motivational signals to properly adjust model-based learning and control[33,34,45,46,50], but not model-free learning and control. Importantly, the lack of evidence of encoding the SPE signals in the dopaminergic reward system in the current study does not speak to that the neural activities in the dopaminergic reward system only represent the model-free PE signals, but not the model-based PE signals. Indeed, the neural activities in the dopaminergic reward system have been also suggested to play critical roles in model-based RL[50–52].

In contrast with the external motivation of maximizing the rewards or minimizing the punishment in the RL framework, the motivational signals encoded in the VS to drive the behavioral control in the current study could be intrinsic. The intrinsic motivation, such as curiosity or information seeking, is encoded in the dopaminergic reward system[53–55]. In the current case, making a precise prediction was irrelevant to the CS-associated valence that the participants would potentially obtain in each trial. Instead, the motivation could be intrinsically minimizing the PE magnitudes at the trial-by-trial basis. Although the goal of minimizing the PE magnitudes could be the accrual of the fixed amount of bonus, this overall motivation would be instead represented by continuously tonic dopaminergic neural activities[56], but should not coincide with the trial-by-trial phasic signals in the VS. Instead, the internal drive of uncertainty reduction[1], might be the underlying motivation of minimizing the PE magnitudes at the trial-by-trial basis. This is evidenced by the observation that the VS also represented the reported confidence[28].

On the other hand, the putamen and the putative SNc in the nigrostriatal system also encoded the motivation PE signals, but not the predicted valences (Supplementary Fig. 3), while the participants made motor actions to report the predictions during the prediction phase. Thereby, these PE signals should be also not relevant to RL, but rather probably associated with the motor execution. It has been well known that volitional movements need indispensable supports from the nigrostriatal system[57]. These neural activities in encoding the action variable of the movement distance of the cursor position, that is, the UPEs, are critical to prepare or sustain the movements that are executed in the PMC[31]. In this perspective, the current findings are consistent with the previous findings that the neural activities in the dopaminergic reward system are also associated with non-valence CS features, such as novelty[58], uncertainty[59], and salience[7], because these CS-associated features are crucial to prepare the coming movements. As the consequence, these features could swiftly reorient the participants' attentions towards the CS[60].

Motivation-dependent control of RL in the dopaminergic reward system

How can the current findings be reconciled with the extant evidence of encoding reward or salience PE

One plausible converging mechanism could be that the

signals encoded in the dopaminergic reward system are subject to the underlying motivation. The current paradigm switched the participants' intentions from the CS-associated valences to the CS-elicited actions[21,22]. Accordingly, the neural PE signals were alternatively encoded in the PMC, rather than in the dopaminergic reward system. Thereby, the PE signals encoded in the dopaminergic reward system, even for model-free RL, should be not automatically computed, consisting with the recent proposal that the neural computation for RL in dopaminergic reward system is generally goal-directed[51,52]. In other words, it should be motivation-dependent.

Despite that the dopaminergic reward system in the current Pavlovian setting was not directly involved in model-free RL, the two subsystems of the VS and the nigrostriatal system had their differential functional roles in regulation of model-based RL. Specifically, the putamen and putative SNc of the nigrostriatal system might sustain action performance (actor), whereas the VS might instead evaluate the current learning policy (critic). In the normative RL tasks, the reward PE signals in the VS and the putamen of the dorsal striatum (DS) have been also proposed to work as the critic in encoding the teaching signals of the PEs and the actor in encoding the updated predictions, respectively[39,61]. Hence, the two subsystems of the human dopaminergic reward system coordinate together to form a general control system with the actor-critic architecture in control of RL.

Lastly, although the fMRI signals in the dopaminergic reward system could be influenced by dopamine[12], the inverse inference from the fMRI activities to the dopaminergic neural activities is logically problematic. Hence, we should remain cautions to interpret these fMRI activities encoding the phasic motivational signals in the regions of the dopaminergic reward system to be dopamine-depedent. Future neurophysiological and optogenetic studies on animals using the similar paradigm are deserved to carefully investigate this outstanding issue.

In conclusion, using a new Pavlovian setting where the participants explicitly reported their predicted valences associated with the CS in both the gain and loss conditions, we found that the neural activities in the regions of the dopaminergic reward system were predominately correlated with the UPEs, rather than the SPEs. These neural signals without indicating the directions of learning could not be directly used for model-free RL. Instead, these neural activities might represent the phasic motivational signals in control of model-based RL, whereas the PMC might implement the model-free RL process. These findings provide new insight on the neural mechanism of the dopaminergic reward system involving in RL.

Methods

Participants.

Thirty-three right-handed participants (18–27 years old, 22 females) participated in the fMRI experiment with a variant Pavlovian conditioning task. Informed consent was obtained from each individual participant, in accordance with a protocol approved by the Beijing Normal University Research Ethics

Experimental paradigm.

In a variant Pavlovian conditioning task, two different cues were stochastically associated with either rewards (gain) or punishers (loss). The two conditions were alternately and randomly intermixed (Fig. 2). The associated gain and loss values (unit: Chinese Yuan) were randomly drawn from a beta distribution, with the mean $\in$ [± 22, ± 26, ±30, ± 34, ±38], using the same standard deviation of 3.6 within a block of 4–8 trials. The number of trials within each block was randomly drawn from a uniform distribution, between 4 and 8 trials. Hence, the outcomes were noisy and volatile[62–64]. The sequence was randomly generated for each participant. Although the neighboring blocks of the same cue type always had different mean values, the cue-associated values appeared to be continuously varied, within [± 10, ± 50], and the change points between the neighboring blocks were not apparent, due to the large standard deviations within each block. Each participant was required to explicitly report the prediction value associated with each presented cue by scrolling the bar position to the target position, in combination with several button presses, where the right or left button presses would increase or decrease the magnitudes, respectively. Specifically, pressing the left or right button with the corresponding index finger corresponded to adding or subtracting 1 from the current position, respectively; the buttons for the middle finger resulted in steps of 5, the ring finger resulted in steps of 10, and the little finger button was used to submit the prediction value. The participants were not provided with any information regarding the environment associated with the task and were merely instructed to learn from the outcomes.

Task sequence.

Each trial started with a 1-s presentation of a fractal image, as the valence-associated cue (*i.e.*, CS). After the cue presentation, the participants reported the valence (prediction) associated with the cue within 3 s. The initial position of the cursor always began at the prediction value reported for the previous trial with the same cue, or at the central position (± 30) for the first trial of each run. Immediately after reporting the prediction value, the participants reported their confidence rating, using a scale from 1 (indicating completely uncertain) to 8 (indicating completely certain), regarding the prediction precision, within 2 s. After a uniformly random jitter, lasting between 3 and 5 s, the actual associated value (outcome) was presented as feedback, for 1 s. The inter-trial interval (ITI) was uniformly random, lasting from 4 to 6 s, causing the prediction and feedback phases to be temporally separated by a 3–7 s gap. Each run consisted of 30 gain trials and 30 loss trials, and a total of eight runs were performed.

The outcomes of one gain trial and one loss trial were independently and randomly chosen to be added to each participant's basic payment (100 Chinese Yuan, approximately 15 US dollars). In addition, each participant was instructed that another bonus equal to 40 Chinese Yuan would be rewarded for good performance in predicting the cue-associated values. In fact, all participants received this bonus. Prior to the fMRI experiment, each participant practiced two runs of the task outside of the scanner.

Behavioral analysis.

We used a simple model-free RL model to characterize the underlying learning process associated with the prediction update (*i.e.*, the RW model[2,3]). Each participant's prediction (p) was assumed to update through a trial-by-trial recursive process, as follows:

$$\mathrm{p_n} = \mathrm{p_{n-1}} + \alpha^{*}\mathrm{pe_n} \quad (1)$$

where $\mathrm{pe_n} = \mathrm{o_n} - \mathrm{p_{n-1}}$, denoting the prediction error; $\mathrm{o_n}$ denotes the actual outcome, α denotes the constant learning rate, and $\mathrm{p_1} = \pm 30$. The updating process is driven by the prediction error.

Alternatively, the participants might progressively gain cue-outcome associations, as described by the Pearce-Hall (PH) model[24], accounting for the associability or attention with the cue, as follows,

$$\alpha_{\mathrm{i}} = (1 - \gamma)\,\alpha_{\mathrm{i-1}} + \gamma\beta\left|\mathrm{pe_i}\right| \quad (2)$$

where α_{i} denotes the associability strength at the trial i, γ denotes the decay constant of the learning rate and β is a scaling coefficient.

We fitted the trial-by-trial predictions with the outcomes, and calculated Bayesian information criterion (BIC) by transforming the minimum of residual sum of square into log-likelihood for each model in each individual participant. We used nonlinear optimization algorithms, implemented in MATLAB (Matlab2012b, Mathworks Inc., Natick, Massachusetts), to separately estimate the parameters of the gain and loss conditions, for each participant.

Further, to illustrate that the learning rates changed with the PEs, we separately divided the trials of the positive and negative PEs (the PEs that equaled to zero were omitted) equally into six bins across the gain and loss conditions for each participant. The mean learning rate in each bin was calculated by a linear model to fit the regression value between the prediction changes and the PEs (Fig. 6a).

fMRI parameters.

All fMRI experiments were conducted using a 3-T Siemens Trio MRI system, with a 12-channel head coil (Siemens, Germany). Functional images were acquired with a single-shot gradient-echo T_2^* echo-planar imaging (EPI) sequence, with a volume repetition time of 2 s, an echo time of 30 ms, a slice thickness of 3.0 mm, and an in-plane resolution of 3.0 × 3.0 mm^2 (field of view: 19.2 × 19.2 cm^2; flip angle: 90 degrees). Thirty-eight axial slices were taken, with an interleaved acquisition, parallel to the anterior commissure-posterior commissure line.

fMRI analyses.

The fMRI analyses were conducted using FMRIB's Software Library[65] (FSL). To correct for rigid head motion, all EPI images were realigned to the first volume of the first scan. Data sets in which the translation motions were larger than 2.0 mm or the rotation motions were larger than 1.0 degree were discarded. No data were discarded from this experiment. Brain matter was separated from non-brain matter by using a mesh deformation approach, which was used to transform the EPI images into individual high-resolution structural images and then into Montreal Neurological Institute (MNI) space, using affine registration with 12 degrees of freedom, and resampling the data with a resolution of $2 \times 2 \times 2$ mm^3. Spatial smoothing, with a 4-mm Gaussian kernel (full width at half-maximum), and high-pass temporal filtering, with a cutoff of 0.005 Hz, were applied to all fMRI data.

For the first-level analyses, two events were applied to each trial. The first event represented the prediction phase, time-locked to the onset of the cue presentation, with the sum of the cue presentation duration (1 s) and the response time (RT) of the prediction report representing the event duration. The second event represented the feedback phase, time-locked to the onset of the outcome presentation, with the presentation duration (1 s) as the event duration. Six general linear model (GLM) analyses with parametric regression were separately applied to the feedback phase of the current trial and to the prediction phase of the subsequent trial with the same cue as follows. (1) We first used the signed PEs (SPEs) to regress with the trial-by-trial fMRI activities, during both the feedback and prediction phases, separately for the gain and loss conditions, to specify the neural encoding of the SPEs in the dopaminergic reward system. (2) To further test the neural encoding of the SPEs in the dopaminergic reward system, we repeated the same process but separated the positive PEs from the negative PEs in both the gain and loss conditions (Supplementary Fig. 2). (3) Because the SPEs and the unsigned PEs (UPEs) were uncorrelated (mean $r = 0.01$, $P = 0.48$), we simultaneously regressed the SPEs and the UPEs with the trial-by-trial fMRI activities in both the feedback and prediction phases of the gain and loss conditions (Figs. 4 & 5). (4) We used the reported confidence as a parameter to regress the trial-by-trial fMRI activities during the prediction phase of the gain and loss conditions (Fig. 5). (5) To examine the neural correlates of the CS-associated valences, we regressed the outcomes and the predictions with the trial-by-trial fMRI activities during the feedback and prediction phases of the gain and loss conditions, respectively (Supplementary Fig. 3). (6) As both the outcomes (mean $r = 0.57$) at the current trial and the predictions at the subsequent trial (mean $r = 0.44$) were highly correlated with the SPEs at the current trial, we then regressed the outcomes and the predictions after orthogalization with the SPEs with the trial-by-trial fMRI activities during the feedback and prediction phases of the gain and loss conditions, respectively (Supplementary Fig. 3). We added the currently irrelevant SPEs or UPEs associated with the alternative CS as the confounding variables during both the feedback and prediction phases in each trial. All the regressors were convolved with the canonical hemodynamic response function, using two-gamma kennels. Further, we also used a delta function at the onsets of both phases to look for the possibly sharp phasic neural responses to the SPEs. We obtained very similar results as used the GLMs described above.

For the group-level analyses, we used FMRIB's local analysis of mixed-effects (FLAME), which model both the 'fixed effects' of within-participant variance and the 'random effects' of between-participant variance, using Gaussian random field theory. Statistical parametric maps were generated by the threshold, with $z > 3.1$, $P < 0.05$ after family-wise error (FWE) correction for multiple comparisons, unless mentioned otherwise.

Regions-of-interest (ROI) definition.

We focused our analyses on the three ROIs of the dopaminergic reward system: the VS, putamen, and putative SNc. These ROIs were defined by the voxels within the anatomically defined regions that reached a significance level at $z > 2.6$ ($P < 0.005$) for the parametric regression of the positive PEs with fMRI activities during the gain condition in the voxel-wise whole-brain analysis. Therefore, the defined ROIs of the VS and putative SNc should agree with the conventional regions of the dopaminergic reward system thought to be responsive to the reward PEs. We then assessed the regression values of fMRI activities with the negative PEs, the SPEs, the UPEs, the cue-associated values and the reported confidence in these ROIs. The anatomical regions of the VS and putamen were defined by the Harvard Subcortical Structures Atlas (including probabilities > 0.5), and the anatomical region of the putative SN was defined by a mask around the ventral tegmental area/SNc[45] (MNI coordinates: x: − 8 to + 6, y: − 26 to − 14, z: − 20 to − 12). The ROI of the amygdala was extracted using the same way as the VS. The ROI of the PMC was defined using the same approach, an anatomically defined PMC area that reached a significance level at $z > 2.6$ ($P < 0.005$) for the parametric regression of the SPEs with fMRI activities in both the gain and loss conditions during both the feedback and prediction phases in the voxel-wise whole-brain analysis.

ROI analyses.

The mean beta values of the GLMs were averaged from the voxels of the ROIs. Further, we also used a trial-based GLM to obtain the trial-by-trial values of the response activities during the prediction and feedback phases. Different from the normal GLM analyses, which use two common regressors across all the trials as described above, here, each trial had independent regressors[66,67]. We then divided all the trials for each participant equally, into ten bins, according to the normalized SPEs. The mean response beta value in each bin was calculated (Fig. 5b, Supplementary Fig. 4b).

Psycho-physiological interaction (PPI) analyses.

To calculate the voxel-wise functional connectivity between the VS region (the seed ROI) and the voxels across the whole brain that changed with the PE magnitudes (*i.e.*, UPEs), we performed another voxel-wise GLM analysis, in which the time course of the VS region (physiological factor), the median-split UPEs (large: 1; small: − 1; psychological factor), and their interaction were put into the feedback phase as

three parametric modulation regressors (Fig. 6b-c). Statistical parametric maps were generated by the threshold, with $z > 2.6$, $P < 0.05$ after FWE correction for multiple comparisons.

Declarations

Author Contributions

Y. Ni conducted the experiments; Y. Ni, S. Wang, J. Su and X. Wan analyzed the data; Y. Ni, S. Wang, J. Su, J. Li and X. Wan designed the experiments; J. Li and X. Wan wrote the manuscript and supervised the project.

Acknowledgements

This research was funded by the Key Program for International S&T Cooperation Projects of China (MOST, 2016YFE0129100, X.W.) and the National Natural Science Foundation of China (No. 31471068, X.W.), and was partially supported by the Fundamental Research Funds for the Central Universities (2017EYT33, X.W.) and the Thousand Young Talents Program of China (X.W.).

Declaration of Interests

The authors declare no competing interests.

References

1. Friston K. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 2010; *11*: 127–138.
2. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory* (Eds Black, A. H. & Prokasy, W. F.) New York: Appleton Century Crofts, 1972; pp. 64–99.
3. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. (The MIT Press, Cambridge, 1998).
4. Schultz W, Dayan P, Montague RR. A neural substrate of prediction and reward. Science 1997; 275: 1593–1599.
5. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 2005; *47*: 129–141.
6. Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat. Neurosci. 2007; *10*: 1615–1624.
7. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 2009; *459*: 837–841.
8. Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N. Arithmetic and local circuitry underlying 5; *525*: 243–246.

9. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron 2003; *38*, 329–337.
10. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 2008; *319*: 1264–1267.
11. Diederen KMJ, Spencer T, Vestergaard MD, Fletcher PC, Schultz W. Adaptive prediction error coding in the human midbrain and striatum facilitates behavioral adaptation and learning efficiency. Neuron 2016; *90*: 1127–1138.
12. Ferenczi EA, Zalocusky KA, Liston C, Grosenick L, Warden MR, Amatya D, Katovich K, Mehta H, Patenaude B, Ramakrishnan C, Kalanithi P, Etkin A, Knutson B, Glover GH, Deisseroth K. Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. Science 2016; 351: aac9698.
13. Reynolds JN, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. Nature 2001; *413*: 67–70.
14. Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth, K. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 2009; *324*: 1080–1084.
15. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci. 2013; *16*: 966–973.
16. Chang CY, Esber GR, Marrero-Garcia Y, Yau HJ, Bonci A, Schoenbaum G. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat. Neurosci. 2016; *19*: 111–116.
17. Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. Temporal difference models describe higher-order learning in humans. Nature. 2004; *429*: 664–667.
18. Delgado MR, Li J, Schiller D, Phelps EA. The role of the striatum in aversive learning and aversive prediction errors. Philos Trans R Soc Lond B Biol Sci. 2008; *363*: 3787–3800.
19. Metereau E, Dreher JC. Cerebral correlates of salient prediction error for different rewards and punishments. Cereb. Cortex 2013; *23*: 477–487.
20. Berridge KC. From prediction error to incentive salience: mesolimbic computation of reward motivation. Eur. J. Neurosci. 2012; *35*: 1124–1143.
21. Boureau Y-L, Dayan P. Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 2010; 1–24.
22. Guitart-Masip M, Duzel E, Dolan R, Dayan P. Action versus valence in decision making. Trends Cogn. Sci. 2014; *18*: 194–202.
23. Sharot T, Guitart-Masip M, Korn CW, Chowdhury R, Dolan RJ. How dopamine enhances an optimism bias in humans. Curr. Biol. 2012; *22*: 1477–1481.
24. Pearce JM, Hall GA. Model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev. 1980; *87*: 532–552.

25. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND. Differential roles of human striatum and amygdala in associative learning. Nat. Neurosci. 2011; *10*: 1250–1252.
26. Pouget A, Drugowitsch J, Kepecs A. Confidence and certainty: distinct Probabilistic quantities for different goals. Nat. Neurosci. 2015; *19*: 366–374.
27. Ni Y, Wang S, Su J, Li J, Wan X. Distributed Neural Implementation of Bayesian Learning in the human brain (forthcoming).
28. Qiu L, Ni Y, Su J, Zhang X, Bai Y, Li X, Wan X. The neural system of metacognition accompanying decision-making in the prefrontal cortex. PLoS Biol. 2018; *16*: e2004037.
29. Kutas M, Donchin E. Studies of squeezing: handedness, responding hand, response force, and asymmetry of readiness potential. Science 1974; *186*: 545–548.
30. Tanji J, Evarts EV. Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. J. Neurophysiol. 1976; *39*: 1062–1068.
31. Vyas S., O'Shea DJ, Ryu SI, Shenoy KV. Causal role of motor preparation during error-driven learning. Neuron 2020; *106*: 1–11.
32. Smittenaar P, Chase HW, Aarts E, Nusselein B, Bloem BR, Cools R. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection-learning or perforamcne? Eur. J. Neurosci. 2012; *35*: 1144–1151.
33. Saunders B, Robinson TE. The role of dopamine in the accumbens core in the expression of Pavlovian-conditioned responses. Eur. J. Neurosci 2012; *36*: 2521–2532.
34. Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB. Amygdala and ventral striatum make distinct contributions to reinforcement learning. Neuron 2016; *92*: 505–517.
35. Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. Nature Neurosci. 2007; *10*: 647–656.
36. Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Koltzenburg M, Frackowiak RS, Friston KJ, Dolan RJ. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat. Neurosci. 2005; *8*: 1234–1240.
37. Paton J, Belova M, Morrison S, Salzman C. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 2006; *439*: 865–870.
38. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 1999; *2*: 79–87.
39. O'Doherty J, et al. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 2004; *304*: 452–454.
40. Sendhilnathan N, Ipata AE, Goldberg ME. Neural correlates of reinformcement learning in mid-lateral cerebellum. Neuron (In press, 2020).
41. Roy M, Shohamy D, Daw D, Jepma M, Wimmer GE, Wager TD. Representation of aversive prediction errors in the human periaqueductal gray. Nat. Neurosci. 2014; *17*: 1607–1612.

42. Li Z, Yan A, Guo K, Li W. Fear-related signals in the primary visual cortex. Curr. Biol. 2019; *29*: 4078–4083.
43. Floresco SB, Yang CR, Phillips AG, Blaha CD. Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. Eur. J. Neurosci. 1998; *10*: 1241–1251.
44. Mohebi, A., et al. Dissociable dopamine dynamics for learning and motivation. Nature 2019; *570*: 65–70.
45. Klein-Flugge MC, Hunt LT, Bach DR, Dolan RJ, Behrens TEJ. Dissociable reward and timing signals in human midbrain and ventral striatum. Neuron 2011; *72*: 654–664.
46. Salamone JD, Correa M. The mysterious motivational functions of mesolimbic dopamine. Neuron 2012; *76*: 470–485.
47. Floresco SB. The nucleus accumbens: an interface between cognition, emotion, and action. Annu. Rev. Psychol. 2015; *66*: 25–52.
48. Howe MW, Tierney PL, Sandberg SG, Phillips PE, Graybiel AM. Prolonged dopamine signaling in striatum signals proximity and value of distant rewards. Nature 2013; *500*: 575–579.
49. Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD. Mesolimbic dopamine signals the value of work. Nat. Neurosci. 2016; *19*: 117–126.
50. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan R. Model-based influences on humans' choices and striatal prediction errors. Neuron 2011; *69*: 1204–1215.
51. Dayan P, Berridge KC. Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. Cogn. Affect Behav. Neurosci. 2014; *14*: 473–492.
52. Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y. Model-based predictions for dopamine. Current Opinion in Neurobiology 2018; *49*: 1–7.
53. Bromberg-Martin ES, Hikosaka O. Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron 2009; *63*: 119–126.
54. Gruber MJ, Gelman BD, Ranganath C. States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit. Neuron 2014; *84*: 486–496.
55. Klucharev V, Hytonen K, Rijpkema M, Smidts A, Fernandez G. Reinforcement learning signal predicts social conformity. Neuron 2008; *61*: 140–151.
56. Niv Y, Daw N, Dayan P. How fast to work: response vigor, motivation and tonic dopamine. Adv. Neural Inf. Process. Syst. 2006; *18*: 1019.
57. Packard MG, Knowlton BJ. Learning and memory functions of the basal ganglia. Annu. Rev. Neurosci. 2002; *25*: 563–593.
58. Schultz W. Predictive reward signal of dopamine neurons. J. Neurophysiol., 1998; *80*: 1–27.
59. Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. Science 2003; *299*: 1898–1902.

60. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 2010; *68*: 815–834.
61. Joel D, Niv Y, Ruppin E. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. 2002; *15*: 535–547.
62. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J. Neurosci. 2010; *30*: 12366–12378.
63. Nassar MR, et al. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat. Neurosci. 2012; *15*:1040–1046.
64. McGuire JT, Nassar MR, Gold JI, Kable JW. Functionally dissociable influences on learning rate in a dynamic environment. Neuron 2014; *84*: 870–881.
65. Smith SM, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; *23*: S208–219.
66. Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 2012; *59*: 2636–2643.
67. Rissman J, Gazzaley A, D'Esposito M. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 2004; *23*: 752–763.

Figures

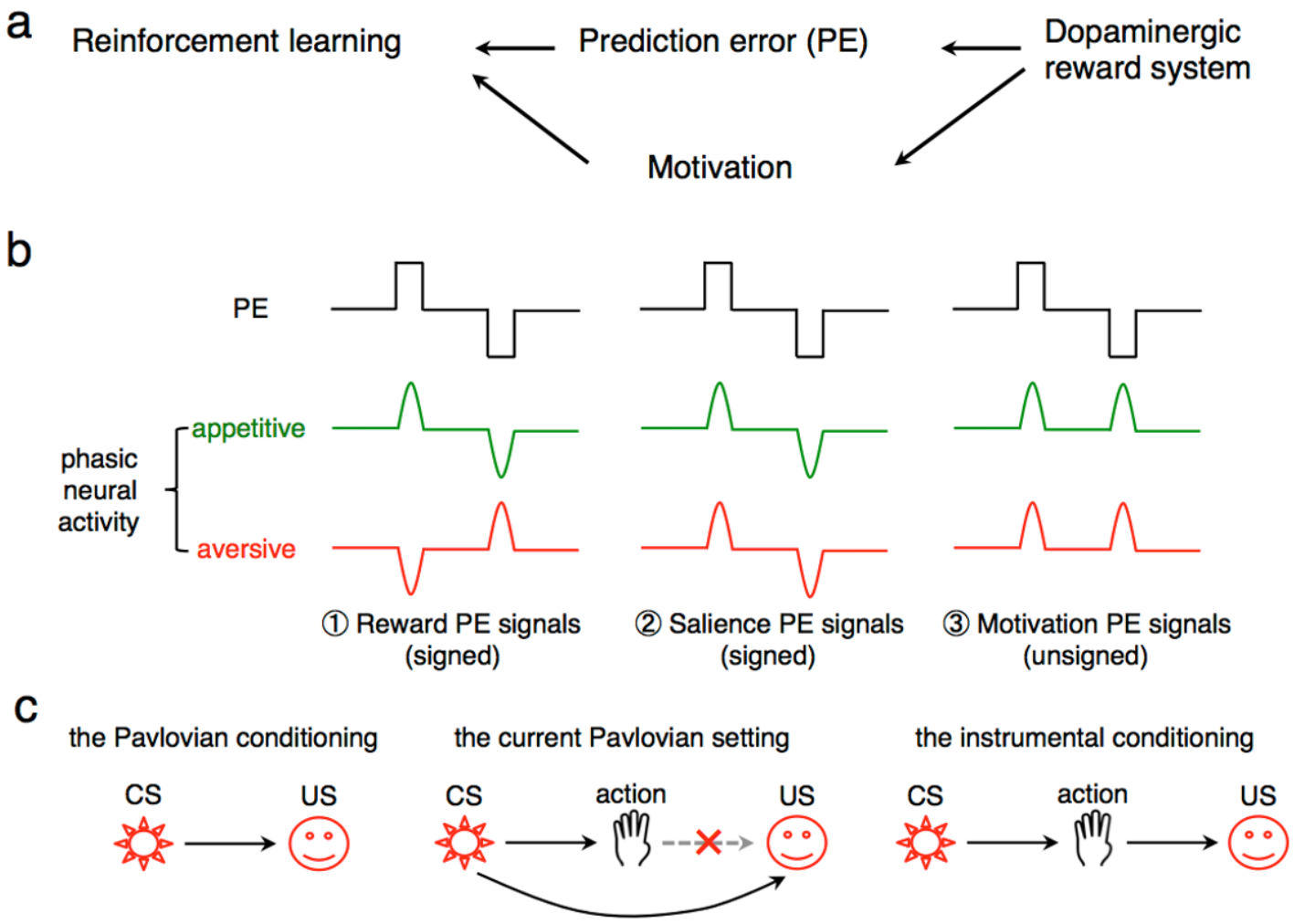

Figure 1

The schematic framework. (a) The two alternative hypotheses about the information encoded by the neural activity in the dopaminergic reward system: (1) the prediction error (PE) signals directly used for model-free reinforcement learning (RL); (2) the motivational signals for controlling the RL process. (b) The phasic neural activity in the dopaminergic reward system (here measured by fMRI) has three possible orthogonal forms in response to the PE, the discrepancy between the experienced and expected valences. (1) The (signed) reward PE signals, which are reverse between the positive PEs and the negative PEs, and are also reverse between the appetitive and aversive conditions; (2) The (signed) salience PE signals, which are reverse between the positive PEs and the negative PEs, but are the same between the appetitive and aversive conditions; (3) The (unsigned) motivation PE signals, which are the same between the positive PEs and the negative PEs, and are also the same between the appetitive and aversive conditions. (c) A new Pavlovian conditioning task used in the current study. In the traditional Pavlovian conditioning task (left), the conditioned stimulus (CS) is directly associated with the unconditioned stimulus (US), there is no need of action performance for the US delivery. In contrast, in the instrumental conditioning task (right), the CS-associated US is conditioned to the CS-elicited actions. In the current Pavlovian setting (middle), The CS-elicited actions are dissociated from the CS-associated US.

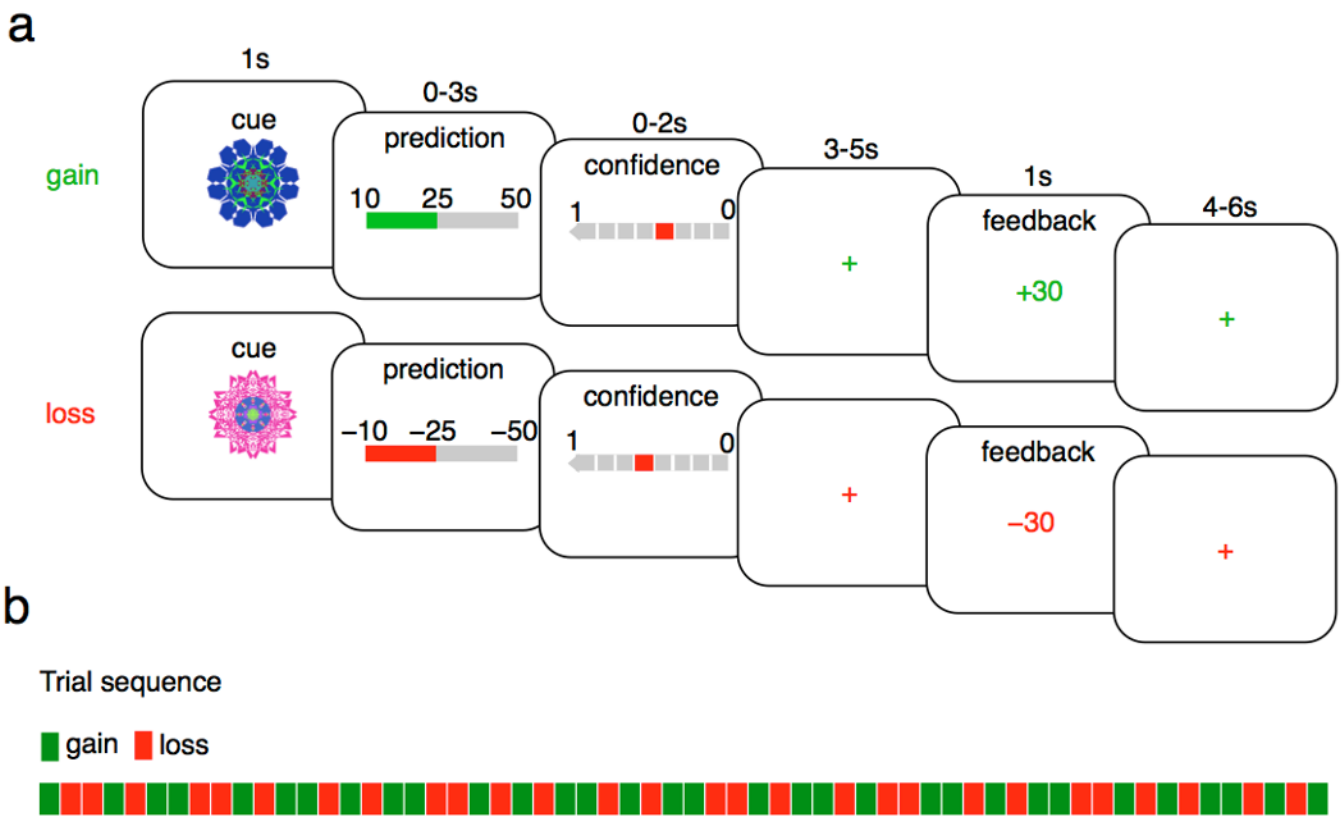

Figure 2

Task paradigm. (a) The cues (fractal images) associated with stochastic rewards (gain) and punishers (loss) were interleaved and randomly presented, the participants reported the predicted valences and reported their confidence, and then the actual outcome was given after a random interval. (b) The gain (green) and loss (red) trials were alternately and randomly intermixed in each run.

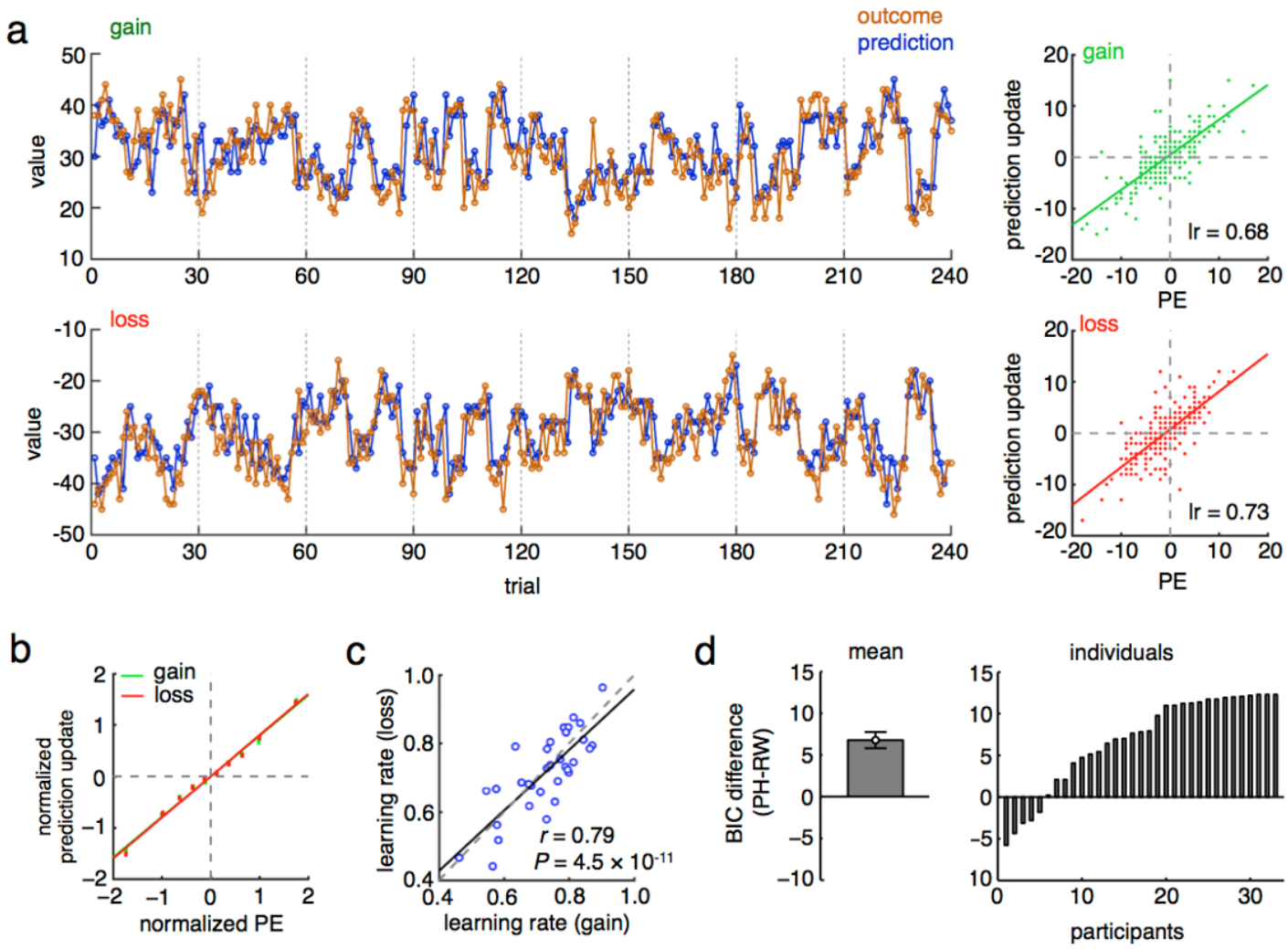

Figure 3

Prediction updates largely followed the RW delta-rule. (a) The predictions at the subsequent trials tracked the outcomes at the current trials associated with the same gain or loss CS in a representative participant (left). The trial-by-trial prediction updates were linearly proportional to the prediction errors (PEs) in both the gain and loss conditions in the same representative participant (right). lr, learning rate. (b) The normalized prediction updates across participants had a linear function of the normalized PEs in both the gain and loss conditions. (c) The learning rates in the gain and loss conditions were consistent (dotted line: y = x; r = 0.79, P = 4.5 × 10-11). (d) The mean (left) and individual (right) Bayesian information criterion (BIC) difference between the PH (attentional-learning rule) and RW (delta-rule) models in fitting with the participants' behavioral data (n = 33).

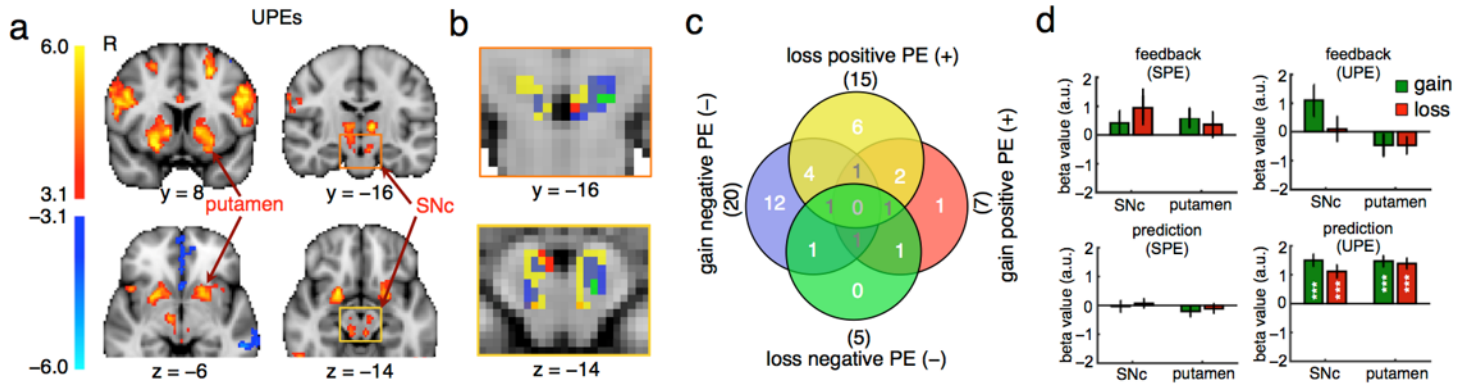

Figure 4

Positive neural correlates of the UPEs in the putamen and the putative SNc during the prediction phase. (a) Activation maps showing significantly positive correlations with the UPEs during the prediction phase, across both gain and loss conditions ($z > 3.1$, $P < 0.05$ after FWE correction). (b) Activation maps in the midbrain region (the rectangles in A), showing significant positive correlations with the positive PEs and negative correlations with the negative PEs, during the prediction phases of the gain and loss conditions, respectively. The same colors represent the same conditions as in (C). (c) The number of voxels (voxel size: 2 × 2 × 2 mm3) in which the fMRI activities were significantly correlated with the SPEs in each condition. (d) The beta values of regression of the fMRI activities with the UPEs and the SPEs in the regions normatively defined by significant correlations with the positive PEs during the prediction phase of the gain condition. ***$P < 0.001$. Error bars indicate s.e.m. across participants.

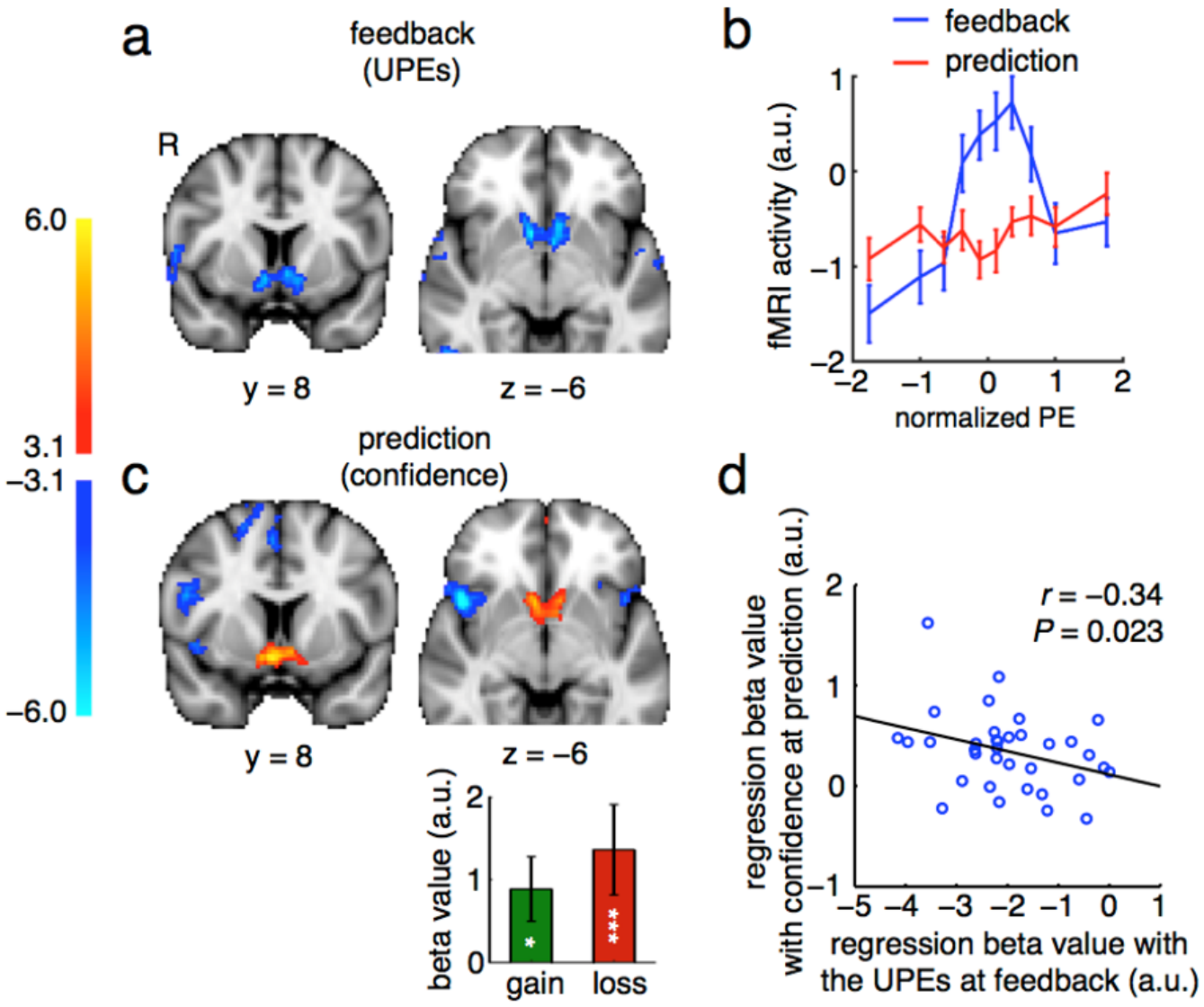

Figure 5

The VS represented the prediction certainty. (a) Activation maps showing significant correlations with the UPEs during the feedback phase, across both the gain and loss conditions ($z > 3.1$, $P < 0.05$ after FWE correction). Significantly negative correlations were identified in the VS. (b) The fMRI activities in the VS as a function of the normalized SPEs. (c) Activation maps showing significant correlations with the reported confidence during the prediction phase of both the gain and loss conditions ($z > 3.1$, $P < 0.05$ after FWE correction). Significantly positive correlations were identified in the VS. *$P < 0.05$, ***$P < 0.001$. Error bars indicate s.e.m. across participants. (d) The regression beta value with the UPEs was negatively correlated with the regression beta value with the reported confidence in the anatomically defined VS across all participants ($r = -0.34$, $P = 0.023$).

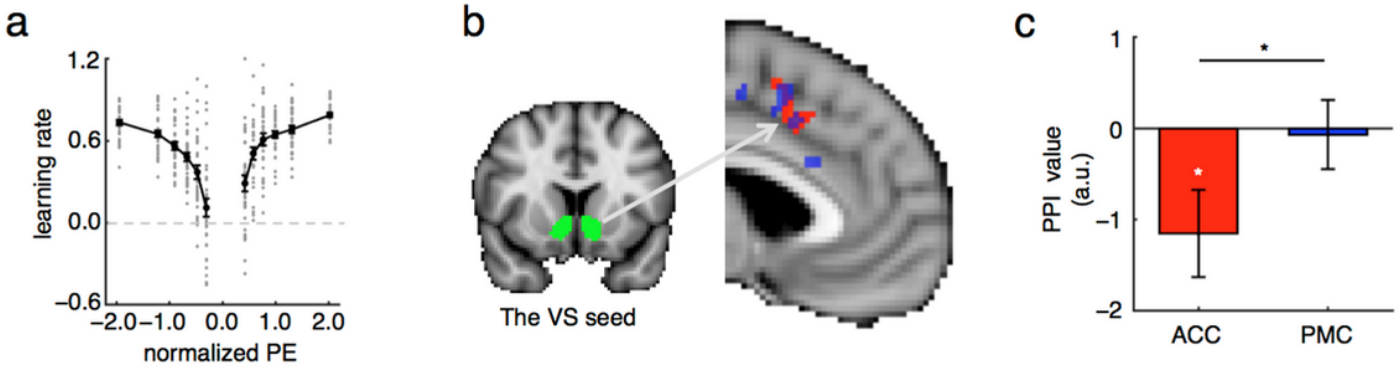

Figure 6

The VS modulated the dACC activities in adjusting learning. (a) The learning rates actually increased with the PE magnitudes. Each gray dot represents the learning rate in each bin for each individual participant. (b) The voxel-wise PPI activation map (blue, $z > 2.3$, $P < 0.01$, uncorrected for illustration), using the fMRI time courses in the VS region as the physiological factor and the median-split UPEs as the psychological factor, overlapping with the anterior cingulate cortex (ACC) region, whose activities were correlated with the UPEs during the feedback phase (red, $z > 3.1$, $P < 0.05$ after FWE correction). (c) The VS-ACC functional connectivity (the PPI values), but not the VS-PMC functional connectivity, became more negative under the greater UPEs. *$P < 0.05$. Error bars indicate s.e.m. across participants.

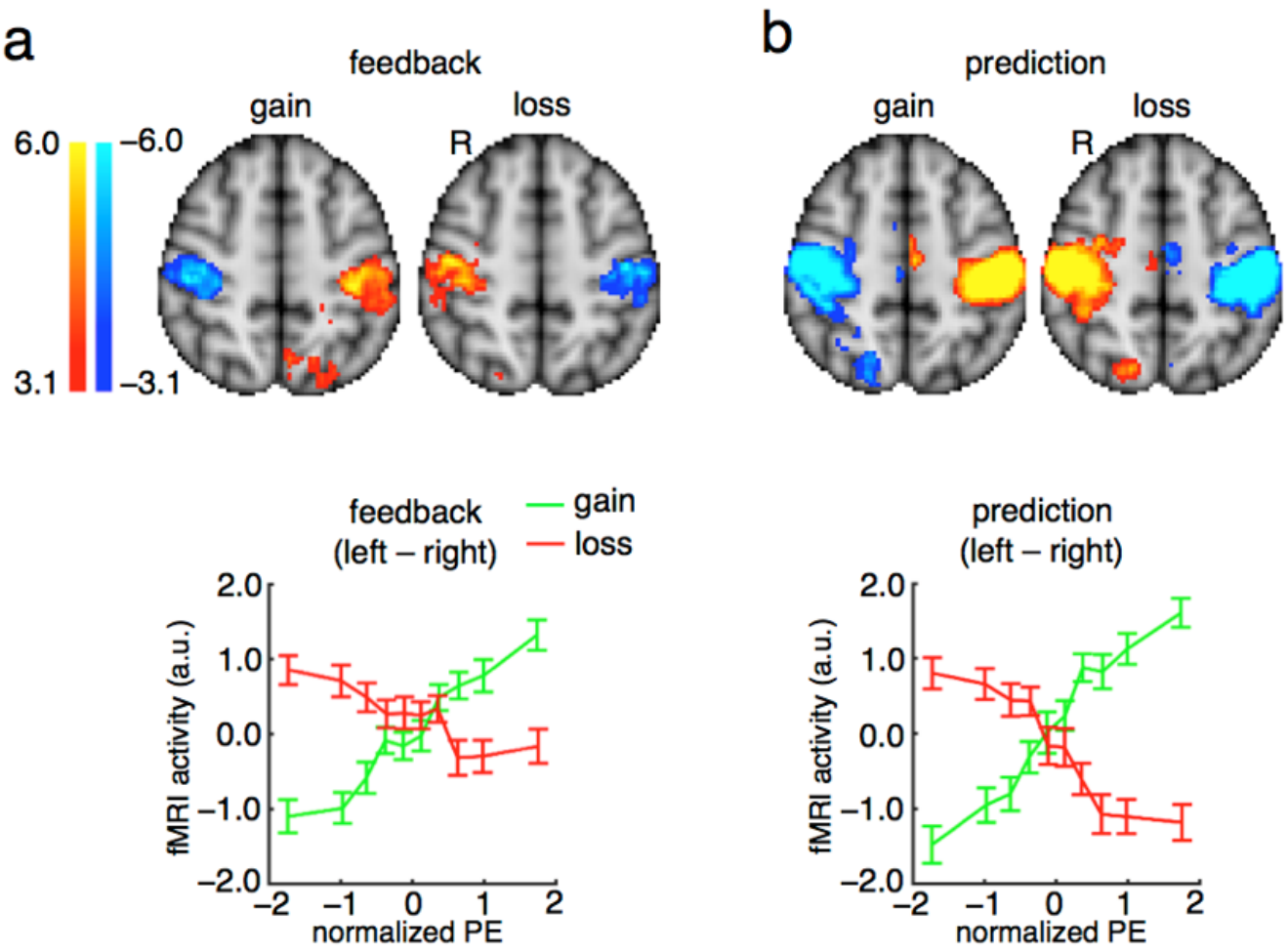

Figure 7

Neural correlates of the SPEs in the PMC. (a) Activation maps showing significant correlations with the SPEs during the feedback phase of the gain and loss conditions ($z > 3.1$, $P < 0.05$ after FWE correction; upper). The fMRI activity as a function of the normalized PEs within individual participants during the feedback phase of the gain and loss conditions (below). (b) Activation maps showing significant correlations with the SPEs during the prediction phase of the gain and loss conditions ($z > 3.1$, $P < 0.05$ after FWE correction; upper). The fMRI activity as a function of the normalized PEs within individual participants during the prediction phase of the gain and loss conditions (below) Error bars indicate s.e.m. across participants.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.docx