Functional neural interactions during adaptive reward learning: An functional magnetic resonance imaging study

Ting Wang | Xi Wu | Jiefeng Jiang | Chang Liu | Ming Zhu

1Department of Computer Science, Chengdu University of Information Technology, Chengdu, China
2Department of Psychology, Stanford University, Stanford, California
3Department of Information Technology and Engineering, Chengdu University, Chengdu, China
4Department of Science and Technology, Chengdu University of Information Technology, Chengdu, China

Abstract

A key feature of learning in humans is flexibility in adjusting the weight of new information to update predictions. This flexibility can be computationally captured by changing the learning rate in a reinforcement-learning model. Key components of reinforcement learning—such as prediction error ($\delta$), learning rate ($\alpha$), and reward feedback ($r$)—have been mapped to various brain areas. However, questions regarding the functional integration patterns in the human brain under the modulation of learning factors, and their interactions, remain unanswered. To investigate these phenomena, we first applied a reinforcement-learning model with an adaptive learning rate and functional magnetic resonance imaging to simulate the individual's reward-learning behavior. Psychophysiological interaction (PPI) analysis was then used to examine the functional interactions of the whole brain under the experimental condition of reward ($r$), the integration between reward and learning rate ($\alpha \times r$), and the integration between the prediction error and learning rate ($\alpha \times \delta$) in a reward-learning task. The behavior statistical analyses indicated that the model estimates of $\alpha$ and $\delta$ captured the participants' learning behavioral patterns of getting high reward, by changing $\alpha$ and $\delta$ for different difficulties and after getting different reward feedback. The PPI analysis results showed that motor-related regions (including the supplementary motor area, precentral gyrus, and thalamus) contributed to cognitive control processing regions (including the middle temporal gyrus, anterior and middle cingulate gyrus, and inferior frontal gyrus) by $\alpha \times r$. Finally, $\alpha \times \delta$ modulated the interaction between subregions of the striatum.

Keywords

adaptive reward learning, functional magnetic resonance imaging, learning rate, prediction error, psychophysiological interaction analysis

Abbreviations: AAL, anatomical labeling; ACC, anterior cingulate cortex; BA, Brodmann area; BOLD, blood-oxygenation-level-dependent; EPI, echo planar imaging; fMRI, functional magnetic resonance imaging; FOV, field of view; GLM, general linear model; IFG, inferior frontal gyrus; MCC, middle cingulate cortex; mFG, medial frontal gyrus; MFG, middle frontal gyrus; MNI, Montreal Neurological Institute; mPFC, medial prefrontal cortex; mSFG, medial superior frontal gyrus; MTG, middle temporal gyrus; PPI, psychophysiological interaction; ROI, region of interest; SMA, supplementary motor area; SPM, statistical parametric mapping; STG, superior temporal gyrus; TR, time repetition; UESTC, University of Electronic Science and Technology of China.
1 | INTRODUCTION

Humans can learn in dynamic environments and determine their future behaviors accordingly. Adaptive learning behavior is characterized by a process of flexible belief adjustment or updating. Researchers have used model-based reinforcement learning to investigate this flexibility in adaptive learning during a reward-learning task.\(^1\)\(^2\) Human decision making can be framed as a computational model of reinforcement learning. In the model, the adaptive reward-learning process involves collecting information from outcomes (reward feedback), calculating the difference between the prediction of the outcomes and the actual environmental outcomes (prediction error, \(\delta\)), and deciding how to update predictions about the future according to errors in prediction (learning strategy or learning rate). Computational models of reinforcement learning infer the free model parameters by fitting the model to the behavior. In one trial of a reward-learning task, one can directly observe the state of reward feedback, actual outcomes, and the participant's prediction. Prediction in the next trial is then updated using the prediction error in the current trial, weighted by a learning rate, which controls the influence of the prediction error on the updating of prediction. The prediction error and prediction together provide trial-wise estimates of the participant's learning rate.

At the behavioral level, recent research has demonstrated that the reinforcement-learning model of reward learning captures the patterns of behavioral performance during a reward-learning task.\(^3\)-\(^5\) In reward-learning tasks, the reward feedback affects the subsequent strategy of participants performing a gambling task.\(^6\),\(^7\) In our previous study, we used a reinforcement-learning model with a flexible learning rate to account for human reward-learning behavior. The results showed that the learning rate estimates were higher after high reward than low reward.\(^8\) In addition, humans seem to use an asymmetric learning rate to determine positive and negative prediction errors.\(^7\),\(^9\),\(^10\) In the reinforcement-learning model,\(^11\) the updating of prediction at time \(i+1\) (denoted as \(\Delta p_i+1\)) is the prediction error (\(\delta_i\)) multiplied by the learning rate (\(a_i\)) at time \(i\); thus, \(\Delta p_i+1 = a_i \times \delta_i\).

At the neural level, existing model-based reinforcement-learning analyses, combined with functional magnetic resonance imaging (fMRI), have located brain structures that encode these model estimates in humans. For example, change in reward feedback activates the dorsolateral prefrontal cortex\(^12\); learning rate is correlated with fMRI signals in the anterior cingulate cortex (ACC),\(^13\) anterior insula, inferior frontal gyrus (IFG),\(^14\) and medial prefrontal cortex (mPFC).\(^15\) Additionally, the prediction error is encoded in subcortical structures such as the striatum\(^16\),\(^17\) and cortical areas such as the mPFC.\(^16\) See Figure 1 for an illustration of reinforcement learning and related neural response in the brain. Furthermore, a great number of researchers have focused on investigating functional connectivity during reinforcement learning and/or reward processing.\(^3\)-\(^5\) For example, using fMRI, Camara and coworkers\(^3\) investigated functional connectivity patterns while participants performed a gambling task—featuring unexpectedly high monetary gains and losses—and found responses for the gain and loss conditions in the insular cortex, amygdala, and hippocampus that correlated with the activity observed in the seed region of the ventral striatum. Li and coworkers\(^4\) used a similar method and found that reduced responses of nucleus accumbens, ventromedial prefrontal cortex, and hippocampal complex to reward feedback were functionally correlated with activation of the dorsolateral prefrontal cortex.

However, questions regarding the functional integration patterns in the human brain under the modulation of learning factors, and their interactions, remain unanswered. One question is how brain regions integrate to reflect the neural joint effect of learning rate and prediction error on updating prediction. Similarly, another question is how to understand the modulation of reward feedback to learning rate and learning strategy from the perspective of brain region functional integration.

Psychophysiological interaction (PPI) analysis is a useful method to test how the temporal correlation of neural activity between brain regions is modulated by a psychological variable. Therefore, in the present study, in order to illustrate the joint effects of learning rate and prediction as well as reward feedback and learning rate, we performed PPI analysis to investigate functional interactions under the corresponding experimental conditions. Existing studies implicate the lateral/medial prefrontal areas, striatum, insula, amygdala, and parietal regions in specific aspects of decision making.\(^13\)-\(^16\) We hypothesize that integration of learning rate and prediction error, as well as reward feedback, would occur in regions that are associated with each factor, such as the medial frontal gyrus (mFG), IFG, and striatum.

2 | MATERIALS AND METHODS

2.1 | Participants

In the present study, we recruited 33 normal, right-handed college students (of which, 19 were men). All participants provided their informed consent, as required by the Institutional Review Board of the University of Electronic Science and Technology of China (UESTC). Data from eight participants were excluded for the following reasons: two owing to head movement; four for...
synchronization failure (the fMRI scanning and task did not start simultaneously) in at least one run; and two for poor performance where there was no response to more than half of the trials in one run. Therefore, the final sample in the fMRI analysis included 25 participants (15 men; ages 21-25 years, mean age of 22 years).

2.2 Experimental design

Each participant underwent fMRI scanning while performing a decision task, which was projected onto a screen sited within the scanner bore. All participants had normal or corrected-to-normal vision. The participants could view the screen via a mirror mounted on the head coil, and they responded by pressing a button on one of two response boxes (one button box assigned to each hand). The presentation of stimuli, recording of responses, and external scanner trigger timing pulses were controlled using Psychophysics Toolbox Version 3 (http://psychtoolbox.org/).

In each trial, participants were instructed to choose between a red square and a green square to accumulate reward points, which determined the monetary reward they received after the task (Figure 2A). At the beginning of each trial, a fixation cross was presented at the center of the screen, along with the two colored squares (one on each side of the screen), for exponentially stepped intervals (4-5.5 seconds; step size, 0.5 seconds), during which the participants chose one square by pressing the button on the same side as the selected square. Once a response was made, the unselected square was removed from the screen. After this interval, the reward obtained (either one point or five points, corresponding to a low or high reward, respectively) was presented at the center of the screen for one second, followed by a fixation cross shown for exponentially stepped intervals (4-5.5 seconds; step size, 0.5 seconds). In the case of no response (<0.3% of all trials), no points were obtained and the message "0" was shown. The total points obtained were displayed at the bottom of the screen throughout the task.

The decision task with a flexible learning rate has been described in detail in our previous study. Briefly, this task consisted of eight runs of 40 trials each. Participants were instructed to obtain as many points as possible. Additionally, they were informed of the following: (a) in each trial, one color was more likely to lead to a higher reward than the other color; (b) the higher reward color was reset at the beginning of each run and could change during the course of a run. Unbeknownst to the participants, for any trial, the sum of the probability of either color being the higher reward color was 1. This

FIGURE 1 Diagram of the reinforcement-learning model estimates and related neural response in the brain. The first line represents the fMRI signal of one subject at one run. The second line represents the modulation of reward feedback to fMRI signal and related neural response. The stripe spacing in the figure represents each trial. There are a total of 40 trials in one run. In the same way, the third line and fourth line represents the modulation of prediction error and learning rate to fMRI signal, respectively [Color figure can be viewed at wileyonlinelibrary.com]
constraint was used to maintain the chance of obtaining a high reward at 50%. To create a wide range of probability that either color was the more rewarding one, the two following conditions were created (four runs for each condition; the order of runs was balanced among participants): (a) in a “difficult” run, the underlying probability of obtaining a high reward by selecting the red square changed every four trials, either in the order of 0.2, 0.4, 0.6, 0.8, 0.6, 0.4, 0.2, 0.4, 0.6, 0.8, or its reverse. (b) In an “easy” run, the probability of obtaining a high reward by selecting the red square remained at 0.2 or 0.8 (two runs for each probability) throughout the run. This design ensured that the mean probability of obtaining a high reward by selecting either color was constantly 0.5 (i.e., the two colors were equally likely to be the high rewarding one) across this task. Before beginning, all participants read the instructions for the task and performed a practice run to ensure that they understood the task. After the scanning session, the participants received monetary compensation, which consisted of a fixed amount for participation in the study plus an additional award based on the score of a randomly selected run.

To simulate how participants learn which color leads to a better chance of receiving a high reward, we adapted the flexible control model, which has been shown to capture the flexible learning of control demands in a changing environment. This model had the following five variables (Figure 2B): (a) the (volatility-driven) flexible learning rate ($\alpha$), which quantifies the model’s (or participant’s) belief concerning the rate of color change corresponding to obtaining a higher reward; (b) the probability of the red square leading to a high reward ($p$); (c) the observed color selection ($s$; either 0 or 1, corresponding to green or red, respectively); (d) the observed reward ($r$; either 0 or 1, corresponding to a low or high reward, respectively), and (e) the observed outcome ($o$), which is determined by $s$ and $r$ and encodes whether the selection resulted in the expected outcome. The trial-by-trial color selection ($s$) and obtained reward ($r$) were submitted to the model (see below) to generate a trial-by-trial estimate of $\alpha$ and $p$. The choices of equations and processing steps have been validated in our recent study.8 For each trial, we could directly observe the prediction error, denoted $\delta$, which is defined as the distance between the previous prediction and outcome. We could also observe the update, defined as the subsequent shift in the participant’s prediction. These two variables together provide trial-wise estimates of the participant’s learning rate. We implemented these procedures using MATLAB R2013a. The scripts are available on request.

2.3 | Image acquisition

Experiments were performed using a GE MR750 3.0 T scanner (General Electric, Fairfield, Connecticut) at UESTC. The anatomical images were scanned using a T1-weighted axial sequence parallel to the anterior commissure-posterior commissure line. Each anatomical scan had 156 axial slices with the following parameters: spatial resolution = 1 mm $\times$ 1 mm $\times$ 1 mm; field of view (FOV) = 256 mm $\times$ 256 mm; and time repetition (TR) = 8.124 ms. The functional images were scanned using a T2*-weighted single-shot gradient echo planar imaging (EPI) sequence with a TR of 2 seconds. Each functional volume contained 43 axial slices with the following parameters: spatial resolution = 3.75 mm $\times$ 3.75 mm $\times$ 3.3 mm; FOV = 240 mm $\times$ 240 mm; echo time = 28 ms; and flip angle = 90°.
Each fMRI run lasted for 416 seconds (208 TRs). During image acquisition, head movement was monitored in real time using software that integrated data from the MR scanner system. When head movement exceeded 2.5 mm or 2.5° within a run, scanning for that run was restarted using a new trial sequence.

### 2.4 | Image preprocessing

The data were preprocessed using SPM8 (www.fil.ion.ucl.ac.uk/spm/software/spm8). The first four time points of each run were discarded. The remaining functional images for each participant were realigned using rigid-body transformation to correct for head movements to the mean functional image using seventh-degree B-spline interpolation, and were then unwarp, and coregistered to the anatomical T1-weighted MRI image using a normalized mutual information function. Next, structural images were segmented and normalized into a common stereotactic space [Montreal Neurological Institute 152 T1-template]. Subsequently, the normalization parameters were applied to the functional images, and these were resampled to 3 × 3 × 3 mm$^3$ isotropic voxel seventh-degree B-spline interpolation. Finally, the resampled functional volumes were smoothed using a Gaussian kernel (FWHM = 8 mm).

### 2.5 | Psychophysiological interaction analysis

Conventional functional connectivity analysis provides information about how spatially separated brain areas are temporally correlated with each other. PPI analysis further tests how the temporal covariance of neural activity between brain regions is modulated by a psychological variable. That is, PPI analysis identifies how the influence of A (source region 1) on B (source region 2) is altered by the experimental context or task, based on regression models.

PPI analyses identify voxels in which activity is more related to activity in a seed region of interest (seed ROI) in a given psychological context, such as during a particular behavioral task. In other words, a PPI analysis aims to identify regions of which their activity depends on an interaction between psychological factors (the task) and physiological factors (the time course of a region of interest). A task-specific increase in the relationship between brain regions (a PPI effect) is suggestive of a task-specific increase in the exchange of information. In the present study, we used PPI analysis to identify functional connectivity patterns of reward learning under the psychological context of $r$, $\alpha \times r$, and $\alpha \times \delta$. We referred to the psychological context as the participant receiving reward feedback (ie, $r$), the interaction of learning rate and reward feedback (ie, $\alpha \times r$) reflecting the modulation of reward feedback to learning rate, and the interaction of learning rate and prediction error (ie, $\alpha \times \delta$) reflecting the updating of learning. The PPI models were estimated in the following steps.

A general liner model (GLM)-based analysis was first performed on the preprocessed fMRI data at each voxel, for each individual. This GLM consisted of the following five regressors, time-locked to the onset of the reward feedback in each trial: the stick function, the reward feedback ($r$), the signed reward prediction error ($\delta$ or $r - p$), the updating in learning ($\alpha \times \delta$), and the interaction between reward ($r$) and $\alpha$ (ie, $\alpha \times r$), which accounted for the behavioral pattern when participants increased their learning rate after receiving a high reward. All regressors were concatenated across the eight runs of this experiment.

We first divided the brain into 90 regions of interest (ROIs) by using the automated anatomical labeling (AAL) template regions. We then extracted individual average time series of blood-oxygenation-level-dependent (BOLD) activity within each ROI from the resulting GLM corresponding to the model estimate of interest. The resulting time courses were convolved with canonical hemodynamic function. A new GLM was then applied with the following regressors: interaction between BOLD activity in the ROI and an indicator function for the model estimate of interest; and the original BOLD time series in the ROI. These regressors were also convolved with a canonical hemodynamic function such that the observed BOLD signal would be a linear combination of these regressors. The individual beta values from the PPI analysis were used in a second level analysis of group effect and tested against the estimated model parameters.

### 2.6 | Statistical analyses

A correction for multiple comparisons was applied using false discovery rate (FDR) across all voxels showing the main effect of repetition in each experiment (at a voxel-wise $P$-value threshold of <.05). For the behavioral data, we conducted repeated measures ANOVAs and two-tailed $t$ tests (significance was set at $P < .05$), using IBM SPSS (IBM Corp., Armonk, New York) to analyze the behavioral and model data.

### 3 | RESULTS

We performed PPI analysis to investigate the functional interactions of the whole brain—under the experimental
condition of reward, the integration between reward and learning rate, and the integration between the prediction error and learning rate—in the 25 participants as they performed a decision task. In each trial, the participants were instructed to choose between a red square and a green square to accumulate reward points, which determined the monetary reward they received after the experiments were complete (see Figure 2A). Importantly, at each moment, one color had a better chance of leading to a high reward than the other color (the sum of the two probabilities was always 1). To maximize reward, the participants must learn which color was more rewarding. To create a wide range of belief regarding the more rewarding color, the task contained two conditions. In the “easy” condition, the more rewarding color and its probability of yielding high reward remained constant at 80% throughout a run. Conversely, in the “difficult” condition, the more rewarding color and its probability of obtaining high reward varied across time (from 20% to 80%, step size = 20%). Across the whole task, each color had a 50% chance of being the more rewarding color.

3.1 | Behavioral results

To test whether participants followed the task instructions for learning to choose the color with the higher reward, we conducted one-sample t tests. Participants chose the higher rewarding color more frequently than chance level (ie, 50%) under two conditions of difficulty (easy condition: 69.23 ± 1.1%, t (24) = 18.11, P < .001, one-sample t test; difficult condition: 52.38 ± 1.0%, t (24) = 2.48, P < .021, one-sample t test; Figure 3A). These outcomes indicated that participants followed the task instructions to learn the more rewarding color. Additionally, participants chose the higher reward color more frequently under the easy condition than under the difficult one (t (24) = 10.35, P < .001, paired t test; Figure 3A); this suggests that manipulation of the difficulty of learning to choose the high reward color was successful.

To further probe how participants adjusted their choice of color based on the reward obtained, we tested the frequency of participants repeating their previous color choice. Because the higher rewarding color was more likely to remain the same as that in the previous trial, it could be expected that participants would repeat their choices more than at chance level. As expected, the overall frequency of repeated choice was significantly higher than chance (74.19 ± 9.5%, t (24) = 12.73, P < .001, one-sample t test). To further test the difference in the frequency of choice repetition among the experimental conditions, we conducted a repeated measures 2 (obtained reward: high/low) × 2 (difficulty: easy/difficult) ANOVA. This ANOVA revealed a significant main effect of the reward obtained [F(1, 24) = 116.15, P < .001; Figure 3A], driven by a higher frequency of repeating the previous color choice after receiving a high reward (55.26 ± 6.8%) than after receiving a low reward (18.93 ± 6.1%); this suggests that a high reward reinforced the participants’ belief that the previously selected color was the higher rewarding one. The main effect of difficulty was also statistically significant [F(1, 24) = 440.45, P < .001], driven by a higher likelihood of repeating the previous color choice under an easy condition (80.83 ± 9.1%) than a difficult condition (67.55 ± 10.8%). This difference indicated the fact that the higher reward color changed more frequently in the difficult condition than in the easy one. The reward type × difficulty interaction was also significant [F(1, 24) = 106.68, P < .001].

3.2 | Model-based behavioral analysis

The flexible learning model generated trial-by-trial estimates of \( \alpha \) and \( \delta \), which drove the learning of the more rewarding color (Figure 4A). To test whether the learning
rate ($\alpha$) and prediction error ($\delta$) reflected behavioral patterns, we conducted the following repeated measures ANOVA analysis: 2 (obtained reward: high/low) $\times$ 2 (difficulty: easy/difficult) ANOVA on the absolute values of $\delta$ and $\alpha$, respectively.

The first ANOVA revealed a significant main effect of the reward obtained [$F(1, 24) = 34.79, P < .001$; Figure 4B], whereby high rewards ($0.21 \pm 0.002$) were associated with lower $\delta$ estimates than low rewards ($0.22 \pm 0.003$). The main effect of difficulty was also statistically significant [$F(1, 24) = 231.36, P < .001$; Figure 4B], driven by lower $\delta$ estimates under the easy condition ($0.39 \pm 0.004$) than the difficult condition ($0.47 \pm 0.004$). This difference possibly implied the fact that the higher rewarding color changed more frequently in the difficult condition than in the easy one. Additionally, we found a significant interaction between the reward type and the difficulty [$F(1, 24) = 59.98, P < .001$] in the $\delta$ estimates.

The second ANOVA analysis revealed a significant influence of the reward obtained [$F(1, 24) = 333.81, P < .001$; Figure 4C], whereby high rewards ($0.162 \pm 0.002$) were associated with higher $\alpha$ estimates than low rewards ($0.156 \pm 0.002$). The difference possibly reflected the fact that an increased learning rate followed a high reward. The main effect of difficulty was also statistically significant [$F(1, 24) = 8.50, P < .008$; Figure 4C], driven by higher $\alpha$ estimates under the easy condition ($0.059 \pm 0.001$) than the difficult one ($0.100 \pm 0.002$). Additionally, we found a significant interaction between the reward type and the difficulty [$F(1, 24) = 113.30, P < .001$] in the $\alpha$ estimates.

Taken together, these results suggest that the flexible learning model estimates of $\alpha$ and $\delta$ captured the behavioral patterns and, therefore, provided meaningful learning-related information for the following imaging analyses.

### 3.3 PPI analysis results

In the behavioral results, we showed that the reward feedback (high or low reward) affected the subsequent...
strategy of color selection. These results also revealed that the learning rate and prediction error mediated the learning strategy, demonstrating that the reward feedback mediated the learning rate, and the learning rate further drove the reward prediction error to update the future decision. To study the brain connectivity patterns of the model estimates and their interactions, we conducted PPI analysis under the experimental context of $r$, $\alpha \times r$, and $\alpha \times \delta$. The reward-modulated PPI analysis results showed that the interactions (ie, functional connectivity) between the right orbital part of the middle frontal gyrus (MFG) and the ipsilateral medial frontal gyrus (mFG)—as well as between the left postcentral gyrus and the right superior temporal gyrus (STG)—were altered by reward feedback (see Figure 5A, Figure 6A, and Table 1).

The key imaging analysis in this study concerned how the flexible learning rate interacts with reward feedback and prediction error. The behavioral analysis revealed an increased learning rate following a high reward. Accordingly, we tested the interaction between the updated learning rate and the reward feedback. PPI analysis results of the integration of the reward and learning rate showed that the altered interactions between the right precentral gyrus and MFG, supplementary motor area (SMA) and middle temporal gyrus (MTG), left medial superior frontal gyrus (mSFG) and Brodmann area 39 (BA39)/STG, right insula and MTG, right ACC and MTG, right middle cingulate cortex (MCC) and BA40, and right thalamus and IFG were modulated by the integration of the reward feedback and learning rate (see Figure 5B, Figure 6B, and Table 2).

PPI analysis results of the integration of prediction error and learning rate showed that the altered interaction between the left caudate and left putamen, as well as between the right mSFG and ipsilateral mFG were modulated by the integration of prediction error and learning rate (see Figure 5C, Figure 6C, and Table 3).

4 | DISCUSSION

In our present study, we applied a reinforcement-learning model with a flexible learning strategy (ie,
learning rate) to simulate human behaviors in a reward-learning task. All participants successfully followed the instructions suggested in the model. In addition, we performed an fMRI-based PPI analysis to examine whole-brain functional interactions with $r$, $\alpha \times r$, and $\alpha \times \delta$. Altered functional connectivity with the mFG was found
in all three PPI analyses, which suggests that the mFG serves as a key hub in cognitive processing. The integration between reward and learning rate engaged the decision control and motor processing pathways, suggesting an integration of decision control and motor information during cognitive decision making. Finally, as expected, integration between the reward and prediction error implicated a pathway related to the striatum, which has been reported as encoding the prediction error in previous studies.8

Our behavioral analysis revealed an increased learning rate and frequency of repeating a color choice after receiving a high reward. These observations are consistent with the existing understanding that reward feedback dominates cognitive decision making and mediates learning rate.22

All PPI analyses of the three model estimates of interest (ie, learning rate, and its integration with reward and prediction error) showed altered functional connectivity with the mFG. This suggests that the mFG may serve as a

### TABLE 1  
Regions showing significant differences of functional connectivity with seed regions in the modulation of reward feedback estimates

| Seed regions | Anatomical region | Hemisphere | Peak MNI co-ordinates $x$ | $y$ | $z$ | $T$ value | Cluster size no. voxels |
|--------------|-------------------|------------|----------------------------|-----|-----|-----------|-------------------------|
| Right middle frontal gyrus, orbital part | BA10/medial frontal gyrus | R | 3 | 59 | 14 | 5.45 | 29 |
| | Medial frontal gyrus/medial superior frontal gyrus | R | 15 | 44 | 23 | 5.13 | 26 |
| Left postcentral gyrus | Superior temporal gyrus | R | 39 | −43 | 14 | 6.60 | 20 |

Note: Peaks are reported for all clusters $\geq$19 voxels at a voxel-wise $P$-value threshold of <.05 (FDR corrected). Abbreviations: BA: Brodmann area; MNI, Montreal Neurological Institute; R, right.

### TABLE 2  
Regions showing significant differences of functional connectivity with seed regions in the modulation of reward feedback $\times$ learning rate estimate

| Seed regions | Anatomical region | Hemisphere | Peak MNI co-ordinates $x$ | $y$ | $z$ | $T$ value | Cluster size no. voxels |
|--------------|-------------------|------------|----------------------------|-----|-----|-----------|-------------------------|
| Right precentral gyrus | Middle frontal gyrus | L | −42 | 2 | 41 | 5.52 | 21 |
| Right supplementary motor area | Middle temporal gyrus | L | −51 | −28 | −4 | 4.96 | 37 |
| Right insula | Medial frontal gyrus | L | −6 | −1 | 65 | 5.55 | 19 |
| Right anterior cingulum | Middle temporal gyrus | L | −42 | −46 | 5 | 4.78 | 26 |
| Right middle cingulum | BA40 | L | −39 | −43 | 35 | 4.30 | 20 |
| Right thalamus | Inferior frontal gyrus | L | −45 | 32 | 5 | 5.15 | 25 |

Note: Peaks are reported for all clusters $\geq$19 voxels at a voxel-wise $P$-value threshold of <.05 (FDR corrected). Abbreviations: BA: Brodmann area; L, left; MNI, Montreal Neurological Institute; R, right.

### TABLE 3  
Regions showing significant differences of functional connectivity with seed regions in the modulation of prediction error $\times$ learning rate estimate

| Seed regions | Anatomical region | Hemisphere | Peak MNI co-ordinates $x$ | $y$ | $z$ | $T$ value | Cluster size no. voxels |
|--------------|-------------------|------------|----------------------------|-----|-----|-----------|-------------------------|
| Right medial superior frontal gyrus | Medial frontal gyrus | R | 9 | 47 | −10 | 4.46 | 19 |
| Left caudate | Putamen | L | −21 | 8 | −1 | 4.39 | 22 |

Note: Peaks are reported for all clusters $\geq$19 voxels at a voxel-wise $P$-value threshold of <.05 (FDR corrected). Abbreviations: L, left; MNI, Montreal Neurological Institute; R, right.
key hub that is integrally involved in cognitive decision making. The mFG has been found to be involved in cognitive functions in fMRI analysis, as well as in electroencephalogram studies. Regions correlated with the mFG include the insula, MFG, and mSFG. These regions are mainly located in the frontal cortical networks, which govern executive and cognition functions such as decision making and control monitoring. The insula has been also found to play a role in motor control and cognitive functioning. Engagement of the mFG and these cortical regions during participants’ performance of a reward task suggest that these pathways may encode the process of cognitive function.

The PPI analysis of reward and learning rate interaction revealed an integration of decision control and motor information during cognitive decision making. The results showed a large of altered functional connectivity in cognitive control areas (including the MTG, ACC, MCC, and IFG) and motor-related regions (including the SMA, precentral gyrus, insula, and thalamus). A previous study indicated that decision making is involved in the neural processes of incorporating new information, selecting action, and central execution such as speed-accuracy trade-offs or decision caution. This evidence suggests that the interaction of reward and learning rate engages the decision hierarchy (frontal and temporal cortices) and motor information processes. Additionally, it is noteworthy that most of these regions are connected with the MTG, which is related to language-related cognition such as syntax comprehension and syntactic processing. This region may serve as a central station for reward evaluation and processing because it is functionally connected with not only motor cortices (SMA, precentral gyrus, and insula) but also cognitive control areas (ACC, MCC, and IFG).

Previous fMRI studies have located brain structures, such as the striatum and mFC, encoding prediction error in humans. Interestingly, our PPI analysis of the interaction between prediction error and learning rate revealed functional connectivity between the left caudate and putamen, as well as the right mSFG and ipsilateral mFG. Both the caudate and putamen are parts of the striatum, which are anatomically and functionally interconnected. Our results further confirm the notion that subcortical regions (striatum) and cortical regions (medial prefrontal cortex) are involved in the reward pathway of the brain.

One limitation of this study is that the task used only two levels of reward magnitude, which resulted in a correlation between reward magnitude and reward prediction error. In addition, given that the goal of the present study was to maximize reward, obtaining a low reward can be seen as the outcome of an “incorrect” response, which may further drive learning through prediction error.

5 | CONCLUSION

In conclusion, we applied fMRI-based PPI analysis to investigate the functional connectivity patterns of model estimates and their interaction in a flexible learning model. The findings suggest the following: (a) the mFG may serve as a key hub in the processing of cognitive decision making; (b) the interaction between reward and learning rate integrate decision control and motor information during cognitive decision making; and (c) the interaction between reward and prediction error involves a striatum-related pathway. This study reports the first probe of functional patterns of model estimates and their interaction in a reinforcement-learning model using a flexible learning strategy.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation (grant number 61806029), the Chengdu University of Information Engineering Research Fund (grant number KYZ201719), and the Project of Sichuan Provincial Education Hall (grant number 18ZA0089).

DISCLOSURE OF INTERESTS

None.

AUTHOR CONTRIBUTIONS

T.W. performed the data analyses and wrote the manuscript. X.W. contributed to the conception of the study. J.J. contributed to the interpretation and discussion of the analysis results. C.L. contributed to the data acquisition. M.Z. contributed to the research design and draft revision. All authors reviewed the manuscript.

ETHICS STATEMENT

This study was performed in accordance with the recommendations of the Institutional Review Board of Chengdu University of Information Technology, with written informed consent from all subjects.

DATA AVAILABILITY STATEMENT

The datasets analyzed in the current study are not publicly available because further analysis of the datasets is currently being conducted in our research. However, the data are available from the corresponding author upon reasonable request.

ORCID

Ming Zhu https://orcid.org/0000-0002-3422-2328

REFERENCES

1. Glascher JP, O’Doherty JP. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdiscip Rev Cogn Sci*. 2010;1(4):501-510.
2. Yoshida W, Ishii S. Model-based reinforcement learning: a computational model and an fMRI study. Neurocomputing. 2005;63:253-269.

3. Camara E, Rodriguez-Fornells A, Münte TF. Functional connectivity of reward processing in the brain. Front Hum Neurosci. 2008;2(2):19.

4. Li J, Delgado MR, Phelps EA. How instructed knowledge modulates the neural systems of reward learning. Proc Natl Acad Sci U S A. 2011;108(1):55-60.

5. Van dBW, Cohen MX, Kahnt T, Crone EA. Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cereb Cortex. 2012;22(6):1247-1255.

6. Yeung N, Sanfey AG. Independent coding of reward magnitude and valence in the human brain. J Neurosci. 2004;24(28):6258-6264.

7. Gershman SJ. Do learning rates adapt to the distribution of rewards? Psychon Bull Rev. 2015;22(5):1320-1327.

8. Wu X, Wang T, Liu C, et al. Functions of learning rate in adaptive reward learning. Front Hum Neurosci. 2017;11:592.

9. Niv Y, Edlund JA, Dayan P, O’Doherty JP. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci. 2012;32(2):551-562.

10. Lefebvre G, Lebreron M, Meynial F, Bourgeois S, Palmentier S. Behavioural and neural characterization of optimistic reinforcement learning. Soc Sci Electron Publ. 2017;1(4):0067.

11. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement. In: Black AH, Prokasy WF, eds. Classical Conditioning II: Current Theory & Research; New York: Appleton-Century-Crofts; 1972.

12. Zanolie K, Teng S, Donohue SE, et al. Switching between colors and shapes on the basis of positive and negative feedback: an fMRI and EEG study on feedback-based learning. Cortex. 2008;44(5):537-547.

13. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007;10(9):1214-1221.

14. Jiang JF, Beck J, Heller K, Egner T. An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. Nat Commun. 2015;6:8165.

15. McGuire JT, Nassar MR, Gold JI, Kable JW. Functionally dissociable influences on learning rate in a dynamic environment. Neuron. 2014;84(4):870-881.

16. Eppinger B, Schuck NW, Nystrom LE, Cohen JD. Reduced striatal responses to reward prediction errors in older compared with younger adults. J Neurosci. 2013;33(24):9905-9912.

17. Jochem G, Klein TA, Ullsperger M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J Neurosci. 2011;31(5):1606-1613.

18. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage. 1997;6(3):218-229.

19. O’Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H. Tools of the trade: psychophysiological interactions and functional connectivity. Soc Cogn Affect Neurosci. 2012;7(5):604-609.

20. Tzourio-mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;15(1):273-289.

21. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp. 1996;4(1):58-73.

22. Jiang J, Heller K, Egner T. Bayesian modeling of flexible cognitive control. Neurosci Biobehav Rev. 2014;46(Pt 1):30-43.

23. Hester R, Barre N, Murphy K, Silk TJ, Mattingley JB. Human medial frontal cortex activity predicts learning from errors. Cereb Cortex. 2008;18(8):1933-1940.

24. Silvetti M, Castellar EN, Roger C, Verguts T. Reward expectation and prediction error in human medial frontal cortex: an EEG study. Neuroimage. 2014;84:376-382.

25. Lee DY, Seo HJ. Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. Ann N Y Acad Sci. 2007;1104:108-122.

26. Goodwin S. Neural Correlates of Executive Control in Prefrontal Cortical Networks. [dissertations & theses—Gradworks]; 2012.

27. Critchley HD. Neural mechanisms of autonomic, affective, and cognitive integration. J Comput Neurol. 2005;493(1):154-166.

28. Chen MY, Jimura K, White CN, Maddox WT, Poldrack RA. Multiple brain networks contribute to the acquisition of bias in perceptual decision-making. Front Neurosci Switz. 2015;9:63.

29. Acheson DJ, Hagoort P. Stimulating the brain’s language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. J Cogn Neurosci. 2013;25(10):1664-1677.

30. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci Biobehav Rev. 2013;37(7):1297-1310.

31. Al-Hakim R, Nain D, Levitt J, Shenton M, Tannenbaum A. Semi-automatic parcellation of the corpus striatum. Proc SPIE Int Soc Opt Eng. 2007;6512:651236.

How to cite this article: Wang T, Wu X, Jiang J, Liu C, Zhu M. Functional neural interactions during adaptive reward learning: An functional magnetic resonance imaging study. Int J Imaging Syst Technol. 2020;30:92–103. https://doi.org/10.1002/ima.22387