

Subgoal- and Goal-related Reward Prediction Errors in Medial Prefrontal Cortex

José J. F. Ribas-Fernandes^{1*}, Danesh Shahnazian¹,
Clay B. Holroyd¹, and Matthew M. Botvinick^{2,3*}

Abstract

■ A longstanding view of the organization of human and animal behavior holds that behavior is hierarchically organized—in other words, directed toward achieving superordinate goals through the achievement of subordinate goals or subgoals. However, most research in neuroscience has focused on tasks without hierarchical structure. In past work, we have shown that negative reward prediction error (RPE) signals in medial prefrontal cortex (mPFC) can be linked not only to superordinate goals but also to subgoals. This suggests that mPFC

tracks impediments in the progression toward subgoals. Using fMRI of human participants engaged in a hierarchical navigation task, here we found that mPFC also processes positive prediction errors at the level of subgoals, indicating that this brain region is sensitive to advances in subgoal completion. However, when subgoal RPEs were elicited alongside with goal-related RPEs, mPFC responses reflected only the goal-related RPEs. These findings suggest that information from different levels of hierarchy is processed selectively, depending on the task context. ■

INTRODUCTION

Learning and behavioral adaptation depend on reward prediction error (RPE) signals, which indicate when expectations about future reward are violated. The neural correlates of RPEs have been observed in ventral striatum (VS) and medial prefrontal cortex (mPFC) across a variety of experimental paradigms (Hyman, Holroyd, & Seamans, 2017; Dolan & Dayan, 2013; Roesch, Esber, Li, Daw, & Schoenbaum, 2012; Niv, 2009). Yet, although RPEs reflect surprise concerning progress toward obtaining rewards, human behavior often involves achieving subgoals to obtain those rewards (or goals; Logan & Crump, 2011; Botvinick, Niv, & Barto, 2009; Lashley, 1951). The hierarchical reinforcement learning (HRL) theory suggests that progress toward subgoals should also elicit RPEs, irrespective of goal-related progress.

Evidence for such subgoal-related RPEs was provided by our past work (Ribas-Fernandes et al., 2011, henceforth RF2011), where we used a multistep navigation task to manipulate independently the distances to subgoals and to goals. In that experiment, we observed that negative subgoal-related RPEs—which indicate unexpected failure to progress to the subgoal—were associated with

an increased BOLD response in mPFC and anterior insula. Further analyses confirmed that these effects did not result from perceptual- or motor-related task confounds. Since RF2011, the HRL framework has been adopted in numerous theoretical and experimental studies (Chiang & Wallis, 2018; Umemoto, HajiHosseini, Yates, & Holroyd, 2017; Balaguer, Spiers, Hassabis, & Summerfield, 2016; Zarr & Brown, 2016; Holroyd & McClure, 2015; Badre & Frank, 2012).

Although RF2011 introduced novel evidence for neural mechanisms of HRL, it had three important limitations. First, the study only examined negative subgoal-related RPEs and not positive subgoal-related RPEs. A positive RPE consists of situations where the outcome is better than expected, whereas a negative RPE consists of situations where the outcome is worse than expected. In the context of subgoal-related RPEs, outcomes refer to progress toward the subgoal. Second, the task did not elicit goal-related RPEs, which makes RF2011 less comparable to more standard studies of goal-related RPEs. Third, RF2011 elicited RPEs with changes in effort expenditure rather than changes in monetary incentives or other standard task elements, again confounding comparisons with the prior literature.

We report here two experiments that address these limitations directly. In Experiment I, we explored the neural correlates of positive subgoal-related RPEs. In Experiment II, we explored the interplay between subgoal-related RPEs and goal-related RPEs by eliciting RPEs related to both effort expenditure and monetary

¹University of Victoria, ²DeepMind, London, ³University College London

*This research was conducted while José J. F. Ribas-Fernandes and Matthew M. Botvinick were at Department of Psychology, Princeton University.

incentives. The inclusion of monetary-incentive RPEs allows us to compare the neural areas involved with processing of effort-related subgoal RPEs with those of money-related goal RPEs. As we shall detail, our experiments yielded surprising results that enrich the original HRL-based interpretation.

EXPERIMENT 1: AN fMRI EXAMINATION OF POSITIVE SUBGOAL-RELATED RPEs

As we have reviewed, RF2011 provided evidence for subgoal-related RPEs in mPFC in an effortful spatial navigation task that incorporated an explicit subgoal–goal structure. Subgoal-related RPEs were elicited by unexpectedly making the subgoal harder to attain by increasing the distance to the subgoal. Here, we adopt the same task, modifying it so that the subgoal is sometimes unexpectedly easier (rather than harder) to attain, thus eliciting—by hypothesis—a positive subgoal-related RPE.

Methods

Participants

Participants were recruited from the Princeton University community, and all gave their informed consent. Thirty

individuals were recruited (ages 18–25 years, $M = 20.5$ years; 11 men, all were right-handed). All participants received monetary compensation at a departmental standard rate. To further encourage performance, participants also received a small monetary bonus based on task performance.

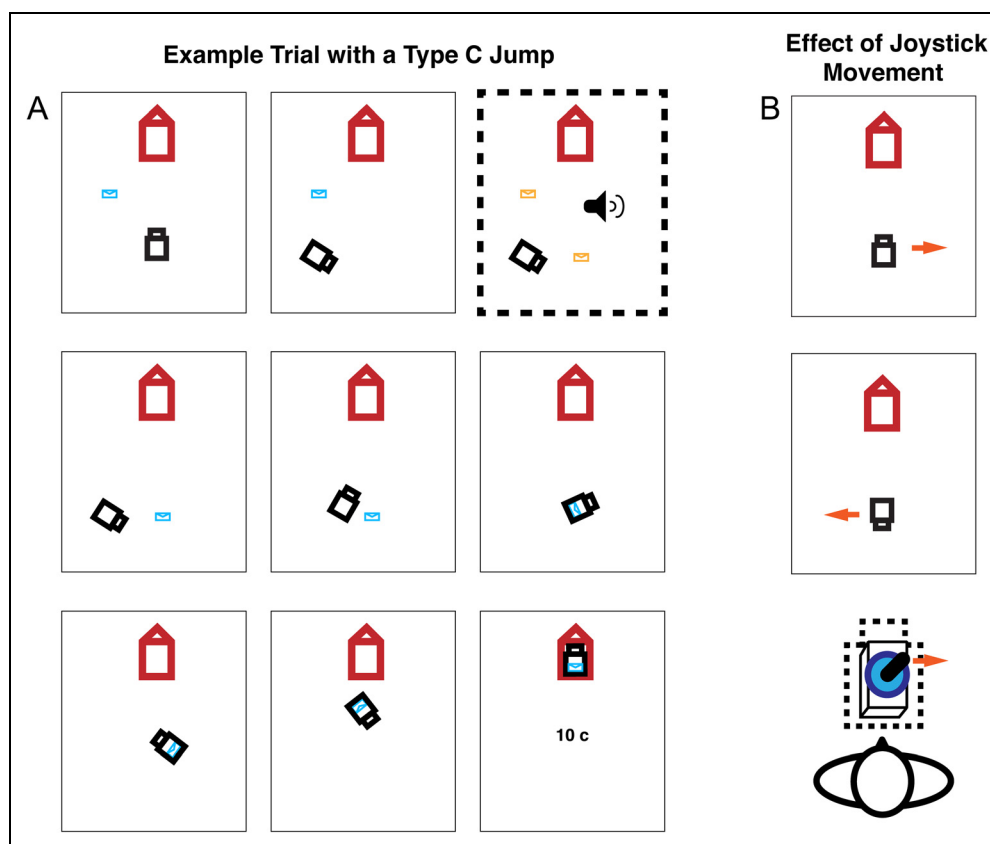
Task and Procedure

Task rationale. We used a hierarchical multistep spatial paradigm similar to RF2011 (see Figure 1A). In what follows we will describe the published paradigm, highlighting any experimental details that depart from the original task design.

In this navigation task, human participants were required to simulate picking up an envelope and delivering it to a house, using a joystick to guide a truck presented on a computer display. Each joystick movement displaced the truck by a fixed distance on the display. We assumed that participants represent the task hierarchically, meaning that they construe delivery to the house as the top-level or “task-level” goal and acquisition of the envelope as a subgoal or “subtask-level” goal.

Importantly, the task was effortful. This was accomplished by randomly changing the orientation of the truck on the display following each joystick movement

Figure 1. Hierarchical delivery task. (A) In this task, participants had to move a truck, using a joystick, to pick up an envelope and deliver it to a house. Each joystick movement displaced the truck by 50 pixels (note that the distance between start point and envelope was 395 pixels). However, after each movement, the orientation of the truck would change randomly. In two thirds of the trials, the envelope would jump to a new location before the truck had reached it, signaled by a beep and a forced pause for 900 msec (see panel bordered by the dashed line). In the remaining third of trials, only the beep and the pause would happen. After delivering the envelope to the house, participants would be rewarded with 10 cents. (B) To ensure that each step would be cognitively effortful, the effect of joystick movements was contingent on the orientation of the truck relative to the screen. For example, if the truck were facing downward, as illustrated in the bottom panel, a rightward movement would displace the truck to the left of the screen.



(Figure 1A). Thus, to move the truck in a desired direction, participants were required to adjust the angle of the joystick to compensate for the random deviations (Figure 1B). Given that each movement required a challenging sensorimotor adjustment, we expected that participants would prefer to travel shorter distances when delivering the envelope. Indeed, in an independent behavioral assay, in RF2011, where participants chose between two envelope delivery trajectories that differed in their overall distance to the goal, they overwhelmingly preferred the shorter route.

The task design allows for changing the distance to the subgoal (i.e., from the starting location to the envelope location) without changing the distance to the goal (i.e., from the start location to the house via the envelope location; Figure 2). Geometrically, all points on an ellipse with foci on the truck and the house have the same overall distance from the start to the house via the envelope, though different distances from the start to the envelope.

We hypothesized that unexpected changes to the distance to the subgoal by displacing the envelope in a random subset of trials, termed “jumps,” would elicit subgoal-related RPEs. To be specific, an unexpected increase in distance to the envelope should elicit a negative subgoal-related prediction error (see Farther Subgoal condition, which was explored in RF2011, but not in the current experiment). In contrast, an unexpected decrease in distance to the envelope should elicit a positive subgoal-related prediction error (see Closer Subgoal condition in Figure 2, which was used in the current experiment, but not in RF2011). Note that the Mirror Jump condition preserves the distance to subgoal and goal, thus not eliciting any RPEs, while still eliciting perceptual and motor changes on the part of the subject; this condition was featured in both the current and published experiments.

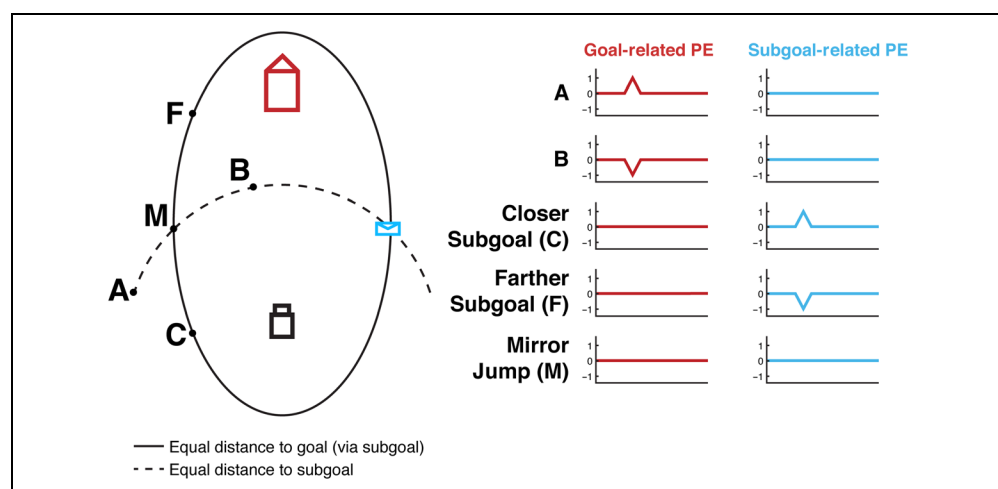
Procedure. The computerized task was coded using MATLAB (The MathWorks) and the MATLAB Psychophysics toolbox, version 3 (Brainard, 1997). An MR-compatible joystick was used for the scanning part of the task (MagConcept), whereas a regular joystick was used for trials outside the scanner (Logitech International).

On each trial, the starting positions of the icons (truck, envelope, house) were vertices in a triangle with fixed distances and angles. The actual positions were random rotations or reflections of the following triangle: truck, 0, 200; envelope, 151, -165; and house, 0, -200 (x, y coordinates in pixels, referenced to the center of a 1024×768 pixels screen). Therefore, the distance between the start point and the envelope was 395 pixels, and the distance between the envelope and the house was 365 pixels, totaling 760 pixels. Note that, as mentioned above, actual positions could vary due to the random rotations and reflections, but the distances and angles between icons were preserved.

Each joystick movement displaced the truck by 50 pixels. The direction of the displacement was a function of the truck's angle with the screen's vertical axis and the angle of the hand movement, inputted through the joystick, relative to center front of the joystick (Figure 1B). After each displacement, the angle between the truck and the screen's vertical axis was changed randomly. Therefore, participants had to adjust the angle of their hand movements on each step to move the truck in the intended direction.

On every trial, after the first, second, or third joystick movement, a brief tone occurred, and the envelope flashed for 900 msec, during which joystick movements were ignored (Figure 1A). On one third of these events (selected at random), the envelope remained in its original location (No Jump condition). On the remaining trials, at the onset of the tone, the envelope jumped to

Figure 2. Different types of RPEs induced by jumps of the envelope to different screen locations. Left view is task display and underlying geometry of the delivery task. Jumps to points on the solid black line, including Closer Subgoal, Farther Subgoal, and Mirror Jump, preserve the overall distance to the goal (start-to-envelope summed with envelope-to-house). Therefore, points on the solid black line only differ in their distance to the subgoal. Right view shows RPE signals generated in each category of jump event. In Experiment I, the envelope would jump to location Closer Subgoal in a third of the trials (triggering a positive subgoal-related RPE) and to location Mirror Jump in a third of the trials and remain in the same place for a third of the trials (No Jump condition). In Experiment II, the envelope would jump to locations that would trigger both a goal-related RPE and a subgoal-related RPE (locations not shown) on two thirds of the trials and remain in the same place in a third of the trials. For illustration, locations where only a goal-related RPE is elicited are shown (A and B).



a new location (jump conditions—described below). Hereafter, we refer to the interruption in the task, common across when the subgoal jumped and did not, No Jump, as “pause event.” In half of the trials in the jump conditions, the distance between the envelope’s new position and the truck position was unchanged by the jump (Mirror Jump in Figure 2). On the remaining third of the trials, Closer Subgoal condition (Closer Subgoal jump) would happen; here, the destination of the envelope was chosen such that the distance between truck and envelope always decreased to 120 pixels whereas the overall path length to the goal (house) was left unchanged. Participants were told that the envelope sometimes stayed in the same place and that sometimes it jumped, with no mention of the distinction between Mirror Jump condition and Closer Subgoal condition.

After the pause event, participants resumed navigating toward the location of the subgoal. When the truck passed within 30 pixels of the envelope, the envelope moved to the truck and remained there for the subsequent moves (Figure 1A). When the truck with the envelope passed within 30 pixels of the house, the image of the truck and the envelope appeared in the house. This image was displayed for 200 msec.

At the completion of each trial (which required, on average, 17.16 steps or joystick movements), participants were rewarded with 10 cents (U.S. dollars). This was indicated by a screen displaying “10 c” for 500 msec. Immediately following this, a fixation cross appeared for 2500 msec, followed by the onset of the next trial, signaled by the appearance of a new spatial arrangement of icons.

Given that the task requires complex sensory–motor coordination, participants practiced the task before functional data acquisition. Practice consisted of 15 min outside the scanner, followed by an 8-min session inside the scanner during structural scan acquisition.

Inside the scanner, for the actual task, participants performed 90 trials, in six runs of 15 trials each, separated by a self-paced rest interval. Each run was approximately 6.8 min, depending on participants’ speed (range = 4.7–10.7 min). Functional data were acquired during these 90 trials.

Behavior Analysis

For each participant, we extracted the mean RT of each condition (Closer Subgoal condition, Mirror Jump condition, and No Jump). We then performed two-tailed paired *t* tests of the mean of the jumps against the mean of No Jump and Closer Subgoal conditions against Mirror Jump condition. We applied a similar analysis to assess the effect of jumps on movement’s accuracy as elaborated below.

Image Acquisition

Data were acquired with a 3-T Siemens Allegra head-only MRI scanner, with a circularly polarized head volume coil.

High-resolution (1 mm³ voxels) T1-weighted structural images were acquired with an MP-RAGE pulse sequence at the beginning of the scanning session. Functional data were acquired using an EPI pulse sequence (3 × 3 × 3 mm voxels, 34 contiguous slices, interleaved acquisition, repetition time = 2000 msec, echo time = 30 msec, flip angle = 90°, field of view = 192 mm, aligned with the anterior commissure–posterior commissure plane). The first five volumes of each run were ignored.

Data Analysis

Data were analyzed using AFNI software (Cox, 1996). The T1-weighted anatomical images were aligned to the functional data. Functional data were corrected for interleaved acquisition using Fourier interpolation. Head motion parameters were estimated and corrected allowing six-parameter rigid body transformations, referenced to the initial image of the first functional run. Data were spatially smoothed with a 6-mm FWHM Gaussian kernel. Each voxels’ signal was converted to percent change.

General Linear Model Analysis

For each participant, we created a design matrix modeling events of interest and nuisance variables. At the time of an event of interest, we defined an impulse and convolved it with a hemodynamic response. The following regressors were included in the model: (a) an indicator variable marking the occurrence of all pause events, (b) an indicator variable marking the occurrence of Mirror Jump condition and Closer Subgoal condition, (c) an indicator variable marking the occurrence of Closer Subgoal condition (note that the events Mirror Jump condition and Closer Subgoal condition are combined together as one regressor to control for the effect of the displacement), (d) a parametric regressor indicating the change in distance to subgoal induced by each jump (only relevant for Closer Subgoal condition, and zeros otherwise), mean-centered, (e and f) indicator variables marking subgoal and goal attainment, and (g) an indicator variable marking all periods of task performance, from the initial presentation of the icons to the end of the trial. Also included were head motion parameters, first- to third-order polynomial regressors to regress out scanner drift effects, and global signal, estimated as the mean for each volume. Parameter estimates from the general linear model were normalized to Talairach space (Talairach & Tournoux, 1988), using SPM5 (www.fil.ion.ucl.ac.uk/spm/).

Group Analysis

For each regressor and for each voxel, we tested the sample of 30 participant-specific coefficients against zero in a two-tailed *t* test. We defined a threshold of *p* = .01 and applied correction for multiple comparison based on

cluster size using Monte Carlo simulations based on estimates of spatial autocorrelation function as implemented in AFNI's 3dClustSim (Version 2018). We report results at a corrected $p < .01$.

ROI Analysis

Given the strong prior for BOLD correlates of RPEs in nucleus accumbens (NAcc) and amygdala (Lee, Seo, & Jung, 2012; Niv, 2009), we defined these anatomical areas as ROIs. NAcc was defined based on anatomical boundaries on a high-resolution T1-weighted image for each participant. Amygdala was defined at the group level using the Talairach atlas in AFNI. Mean coefficients were extracted from these regions for each participant. Reported coefficients for all ROIs are from general linear model analyses without subtraction of global signal. The sample of 30 participant-specific coefficients from these regions was tested against zero in a two-tailed t test, with a threshold of $p < .05$.

Results

Behavior

Each trial took 17.16 steps, on average, across participants ($SEM = 0.60$ steps). There was no difference in the number of steps between conditions (one-way ANOVA, $F(2, 29) = 0.08, p = .92$). Mean RT for each joystick movement was 1090 msec ($SEM = 60$ msec). On average, at 3.96 steps ($SEM = 0.11$ steps), the program interrupted the execution of the task by introducing a pause of 900 msec (i.e., a pause event, which encompasses the Closer Subgoal, Mirror Jump, and No

Jump conditions). In two thirds of the trials, the envelope jumped to a new location at the onset of the pause (jump condition), and in the remaining third, it remained in the same place (No Jump condition). After the pause event was completed, participants took, on average, 610 msec to produce a new joystick movement ($SEM = 70$ msec; note that the enforced delay of 900 msec for all conditions is not included in this measurement). Participants were significantly slower to react to a pause in the jump condition ($M = 690$ msec, $SEM = 70$ msec) than in the No Jump condition ($M = 460$ msec, $SEM = 60$ msec) as revealed by a two-tailed paired t test, $t(29) = 7.96, p < .01$. However, there was no significant difference between Closer Jump ($M = 700$ msec, $SEM = 80$ msec) versus Mirror Jump ($M = 660$ msec, $SEM = 80$ msec; $t(29) = 1.53, p = .14$, two-tailed paired t test).

Whole-brain Analysis

We regressed BOLD activity onto two regressors of interest, a categorical regressor indicating a positive subgoal-related RPE (elicited by Closer Subgoal condition, see Figure 2) and a parametric regressor for the magnitude of subgoal-related RPE (measured as mean-centered decrease in truck-subgoal distance). In the same model, we included three task-specific control regressors (common effect of jump: Closer Jump + Mirror Jump, mean-centered displacement distance, and common effect of pause event: Closer Jump + Mirror Jump + No Jump), along with standard control regressors (see Methods). As in RF2011, we found an effect of subgoal-related RPEs on the BOLD activity in the mPFC. However, the sign of the effect was notable: Whereas in the previous

Table 1. Closer Subgoal Condition. Experiment I. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	t Statistic	Coordinates (x, y, z)
R. lingual G.	462	-.36	-5.74	+0, +71, +1
L. postcentral G.	360	-.25	-3.41	+33, +29, +67
R. superior frontal G.	127	.30	3.26	-33, -46, +31
L. medial frontal G. (see Figure 3)	119	.14	4.62	+0, -7, +43
L. lentiform nucleus	91	-.11	-4.41	+24, -1, -2
R. insula	90	.12	3.03	-45, -13, +1
L. medial frontal G.	65	-.11	-4.26	+3, +20, +49
R. lentiform nucleus	60	.09	-3.80	-24, -7, +4
L. middle frontal G.	57	.17	3.37	+36, -28, +43
R. superior temporal G.	53	-.21	-2.83	-51, -19, -20
L. superior temporal G.	53	-.11	-4.20	+54, +20, -2

Primary threshold $p < .01$, cluster-corrected to $p < .01$, $df 29$. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

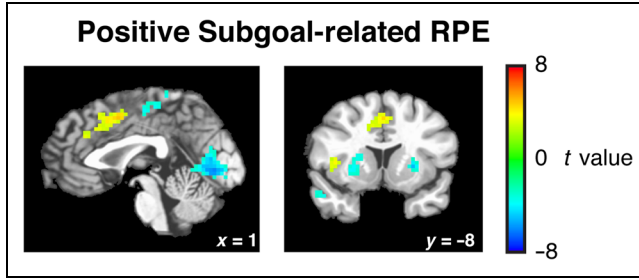


Figure 3. Whole-brain results of Experiment I, positive subgoal-related RPE (Closer Subgoal). Jumps that featured a decrease in distance to the envelope, without changing the overall distance to the house, were associated with an increase in BOLD activity in mPFC and anterior insula. This effect is independent from spatial reorientation, as suggested by the absence of activity in these areas to Mirror Jump, a condition with the highest angle of displacement and no changes in distance to the envelope.

study we observed an increase in BOLD activity in mPFC and right anterior insula to negative subgoal-related RPEs, here we found this activity to positive subgoal-related RPEs (cluster-corrected, $p < .01$; Table 1 and Figure 3). Results for control regressors are in Tables 2–5.

ROI Analysis

To investigate whether areas known to process goal-related RPEs were responsive to subgoal-related RPEs in our experiment, we anatomically delineated two ROIs, NAcc, and amygdala.

As in RF2011, we found no significant change in the BOLD response associated with Closer Subgoal condition in an anatomically defined ROI around bilateral NAcc (mean of regression coefficient, $M = -1.83 \times 10^{-3}$, $p = .94$). Qualitatively similar results were obtained when the same analysis was performed on separate ROIs encompassing either left or right NAcc. However, in contrast with RF2011, we did not observe any subgoal-related RPE in the amygdalar region (Closer Subgoal condition, $M = -0.04$, $p = .06$; parametric decrease in subgoal

distance, $M = -3.58 \times 10^{-4}$, $p = .30$, similar null results for bilateral and unilateral analyses).

Interim Discussion

Although RF2011 found that the BOLD activity in the mPFC is elevated by negative subgoal-related RPEs, that study did not directly test for the neural correlates of positive subgoal-related RPEs. We showed here that activity in mPFC is also correlated with positive subgoal-related RPEs. Taken with the previous results, these findings suggest that the mPFC BOLD signal response to subgoal-related RPEs is unsigned, that is, sensitive to the magnitude of the RPE, but not its valence (Roesch et al., 2012; Hayden, Heilbronner, Pearson, & Platt, 2011). We discuss the implications of this observation in the general discussion. Note that we did not find any effect for the parametric regressor of subgoal RPE. However, it is hard to interpret this finding because our parametric regressor had a restricted range. A stronger test of unsigned prediction errors would entail eliciting changes in magnitude of the RPE, as well as presence/absence of an RPE. However, given our task design, where the subgoal jumped to a fixed subgoal location, the restricted range in the magnitude of positive subgoal-related RPEs limits our experimental power to detect an effect of magnitude.

EXPERIMENT II: AN EXAMINATION OF GOAL AND SUBGOAL-RELATED RPEs

RF2011 and present Experiment I provided evidence for subgoal-related RPEs. However, one limitation of both studies is that they did not elicit goal-related RPEs and thus cannot assess the direct relations between the neural correlates of subgoal-related and goal-related RPEs. Therefore, the present experiment included subgoal displacements that simultaneously changed the distances to both the goal and the subgoal, thereby eliciting both

Table 2. Decrease in Distance to Subgoal. Experiment I. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	t Statistic	Coordinates (x, y, z)
L. precuneus	1105	.01	5.59	+0, +77, +49
L. medial frontal G.	118	−.00	−3.26	+0, −61, +19
L. middle temporal G.	113	−.00	−4.39	+66, +5, −11
L. angular G.	76	−.01	−3.36	+45, +74, +34
L. medial frontal G.	72	−.00	−2.89	+0, −43, +16
L. middle temporal G.	68	−.01	−3.11	+69, +44, +1
R. superior temporal G.	58	−.00	−3.86	−42, −1, −11

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 29. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

Table 3. Pause Event. Experiment I. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. superior temporal G.	1346	.37	4.91	−69, +41, +10
L. precuneus	1151	−.37	−4.04	+12, +80, +46
L. precuneus	1017	.34	5.57	0, +59, +43
R. middle frontal G.	882	−.32	−3.39	−36, −43, +34
L. middle frontal G.	835	−.22	−4.50	24, −19, +55
L. superior temporal G.	798	.35	6.35	+66, +44, +13
L. middle frontal G.	786	−.34	−6.24	+45, −28, +37
R. cuneus	772	.44	2.90	−3, +95, +13
R. inferior parietal lobule	703	−.43	−4.91	−51, +44, +52
L. inferior frontal G.	463	−.34	−4.58	+54, −16, −2
L. postcentral G.	366	.29	4.47	+39, +38, +64
R. medial frontal G.	352	.28	3.02	−3, −61, +4
L. lentiform nucleus	137	.22	6.43	+21, −4, −2
R. lentiform nucleus	128	.18	4.95	−21, −4, −2
R. middle occipital G.	124	−.33	−3.21	33, +95, +4
L. inferior temporal G.	118	−.21	−3.27	+66, +53, −11
R. middle frontal G.	71	−.12	−3.93	−24, +5, +58
R. caudate nucleus	63	−.16	−6.06	−12, +8, +19
L. caudate nucleus	63	−.16	−4.29	+12, +8, +19

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 29. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

subgoal- and goal-related RPEs within a single event. We dissociated these two types of RPEs by systematically manipulating the displacements over the course of the experiment.

Our prediction—which turned out to be incorrect—was that the activity of mPFC would be modulated by

both subgoal- and goal-related RPEs. This prediction was informed by prior empirical research showing simultaneous subgoal- and goal-related activity in an anatomically overlapping region (Diuk, Tsai, Wallis, Botvinick, & Niv, 2013). A second, subordinate objective of this experiment was to examine the relation between RPEs

Table 4. Mirror Jump and Closer Subgoal Conditions. Experiment I. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. precuneus	839	.42	3.30	−3, +68, +52
R. lingual G.	556	.25	6.09	0, +74, +1
R. insula	188	−.11	−5.42	−39, +32, +19
L. fusiform G.	69	.13	3.47	+24, +56, −8
L. superior frontal G.	64	.18	3.51	+30, +8, +64
R. middle frontal G.	58	−.10	−3.90	−39, −31, +16

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 29. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

Table 5. Displacement Distance. Experiment I. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
L. precuneus	228	.01	3.41	+0, +74, +52
L. lingual G.	143	.00	4.38	+3, +74, +4
R. traverse temporal G.	71	−.00	−4.13	−42, +29, +13
R. postcentral G.	59	−.00	−3.31	−60, +20, +40

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 29. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

due to changes in effort/time expenditure versus RPEs associated with monetary feedback. Previous research has found that both types of reinforcers are processed by the VS (Satterthwaite et al., 2012; Botvinick, Huffstetler, & McGuire, 2009; though see Skvortsova, Palminteri, & Pessiglione, 2014, for a dissociation between physical effort RPEs and monetary RPEs).

To be specific, this task paradigm elicits different types of prediction errors by having the subgoal unpredictably jump to different points in space (as illustrated in Figure 2). In Experiment II, similar to Experiment I, two thirds of the trials featured a jump to a new location, whereas in the remaining third the location of the envelope did not change. However, in contrast with Experiment I, all of the jumps were to the locations that should putatively elicit both goal-related RPEs and subgoal-related RPEs. We manipulated the displacement of the jumps so that the magnitude of different types of prediction errors was uncorrelated across the experiment. We opted for a parametric design rather than a categorical design because the categorical manipulation of types and valence of prediction errors would have resulted in a prohibitively high number of conditions: four experimental conditions (Closer Subgoal, Farther Subgoal, and equivalent conditions changing goal-related RPE) and two control conditions (Mirror Jump and No Jump trials), resulting in too few trials per condition.

Methods

Participants

Forty-eight participants were recruited from the Princeton University community, and eight participants were excluded, seven for having incidents of head movements larger than 2.5 mm and one for failure to complete the task inside the scanner (ages 18–27 years, $M = 20$ years, 15 men, 38 were right-handed and 2 were left-handed, joystick was always held in the right hand). All participants received monetary compensation at a departmental standard rate. In addition, participants received

two types of monetary bonuses, one is based on performance and the other is a probabilistic “tip,” as described below.

Materials, Task and Procedure

As before, the overall approach of RF2011 was used to manipulate the subgoal- and goal-related RPEs in an independent manner. The task consisted of three parts: a short behavioral practice outside the scanner, for 12 trials, using a joystick held in the right hand (Logitech International); a 12-trial practice inside the scanner, using an MR-compatible joystick (MagConcept) during structural scan acquisition; and a third phase of 132 trials (six runs of 22 trials) for approximately 60 min (run duration, $M = 11.7$ min, $SEM = 0.3$ min), where functional data were collected. Each run started and ended with a central fixation cross, displayed for 10,000 msec. At the end of each run, participants were given a self-paced break.

On each trial, the house occupied the same vertex as in Experiment I (0, −200). The initial position of the truck and the envelope, which were different than the values used in Experiment I and in RF2011, were determined as follows. The initial position of the truck (−90, 320) was set so that it would be 150 pixels or three optimal steps from a virtual line beyond which a jump would be triggered. This line was parallel to the house–envelope line and would go through the point (0, 200), a point where the envelope is at the same distance to the house and to the truck. This location was utilized for convenience because it allows for equal variance in both positive and negative prediction errors.

As in Experiment I and RF2011, when a jump was triggered, a brief tone was played, the truck and the envelope flashed yellow, and joystick movements were ignored for 900 msec. This pause event happened, on average, after 5.6 steps ($SEM = 0.1$ steps). In one third of the trials (44 trials), the envelope stayed in the same location. In the remaining two thirds of the trials (88 trials), it jumped to a new location (see the following).

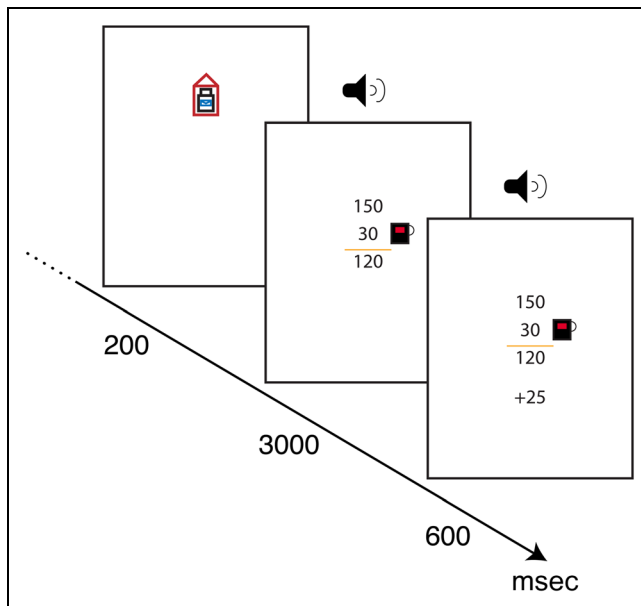


Figure 4. Eliciting monetary prediction errors. In Experiment II, at the end of each trial, participants would receive information about their performance. A delivery yielded 150 points, and any additional step from the shortest distance possible would be deducted from this rate. In the example above, 30 points were deducted for extra steps. In addition, a probabilistic outcome was introduced (+25). Unrelated to their performance, participants would receive a bonus of 25 points, 0 points, or -25 points, with equal probability. Points accrued would be exchanged for U.S. dollars at the end of the experiment.

In this experiment and in contrast to both RF2011 and Experiment I, each jump generated a goal-related RPE and a subgoal-related RPE. We applied a Monte Carlo approach to find a set of 88 jump locations for which the goal-related RPE and subgoal-related RPE were minimally correlated. Our sampling approach also minimized the

correlation between the nuisance variable displacement distance (distance between old and new subgoal locations) and the variables of interest. This procedure was done for each participant independently. In the observed behavioral data, the average correlation between the subgoal-related RPEs and the goal-related RPEs was .31, the correlation between the subgoal-related RPEs and the displacement distances was 0, and the correlation between the goal-related RPEs and the displacement distances was -.37.

As in Experiment I and RF2011, after the jump, the participants were required to navigate toward the new location of the subgoal. When the truck passed within 30 pixels of the envelope, the envelope moved to the truck and remained there for the subsequent moves. When the truck with the envelope passed within 30 pixels of the house, the truck with the envelope appeared within the house. This image was displayed for 200 msec.

To encourage participants to pay attention to the consequences of the change in the amount of effort expenditure incurred by each displacement event, we penalized deviations from the optimal path using a point-based system. Note that this feature of the task was not present in RF2011 nor in Experiment I. Participants could attain a maximum of 150 points per delivery, with a penalty of 0.1 points per excess pixel traveled. At the end of each trial, the number of points obtained was presented after the truck entered the house (for 3000 msec; see Figure 4), accompanied by the sound of cash register. At the end of experiment, the sum of points was converted to U.S. dollars, up to a maximum of \$12.

As mentioned before, a secondary objective unique to this experiment was to compare the neural correlates

Figure 5. Comparison of unsigned responses across Experiments I and II and RF2011. For comparison, only the positive clusters in Experiment I are shown (see Figure 1 for an image with positive and negative clusters).

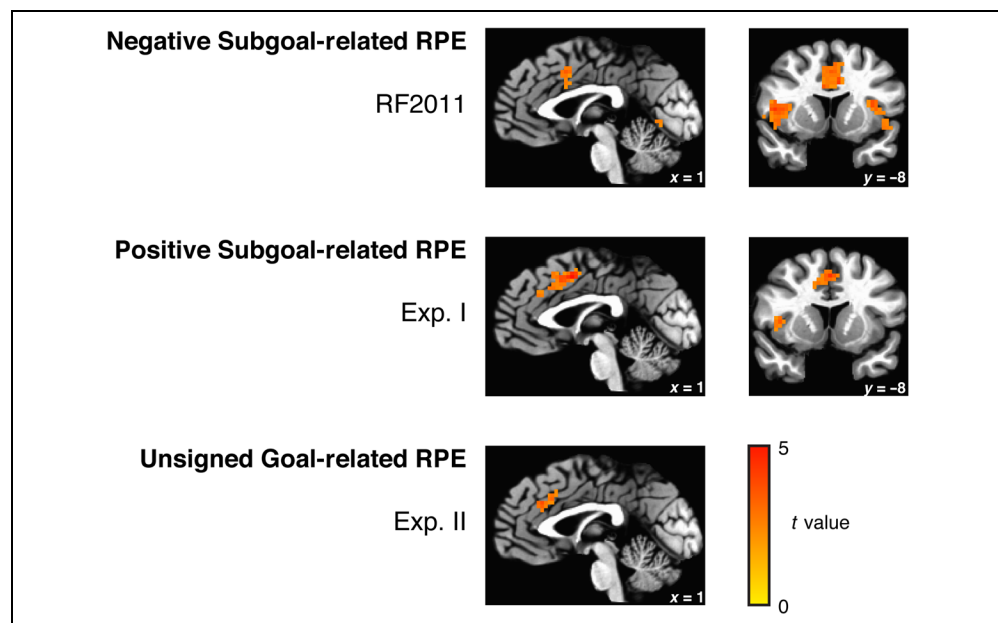


Table 6. Unsigned Distance-driven Goal-related RPE. Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
L. cingulate G.	45	.00	4.20	+1, -28, +29

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

related to effort expenditure versus monetary reward. To do so, after presenting the points feedback on each trial (as described above), a probabilistic monetary reward, in the same point currency, was delivered (Figure 4). Unbeknown to the participants, outcomes were 25 points (framed as a tip), -25 (framed as shortchange), or 0 points, randomly sampled with equal probability. To ensure attentional capture, a sound was delivered simultaneously with presentation of this information (coin sound for 25, a sad trumpet sound for -25, and a brief tone for 0; all sounds 100-msec duration). This probabilistic monetary reward was displayed for 600 msec. The trial was followed by intertrial interval with a fixation cross that remained on screen for 700 msec.

Image Acquisition

Data were acquired with a 3-T Siemens Skyra MRI scanner using a 16-channel head coil. High-resolution (1 mm^3 voxels) T1-weighted structural images were acquired with an MP-RAGE pulse sequence at the beginning of the scanning session.

Functional data were acquired using a high-resolution EPI pulse sequence ($3 \times 3 \times 3\text{ mm}$ voxels, 35 contiguous slices, 3-mm-thick, interleaved acquisition, repetition time = 2000 msec, echo time = 30 msec, flip angle = 90° , field of view = 192 mm, aligned with the anterior commissure–posterior commissure plane). The first five volumes of each run were ignored.

Data Analysis

Data were analyzed using AFNI software (Cox, 1996). The T1-weighted anatomical images were aligned to the functional data. Functional data were corrected

for interleaved acquisition using Fourier interpolation. Head motion parameters were estimated and corrected allowing six-parameter rigid body transformations, referenced to the initial image of the first functional run. A whole-brain mask for each participant was created using the union of a mask for the first and last functional images. Spikes in the data were removed and replaced with an interpolated data point. The data were spatially smoothed with a 6-mm FWHM Gaussian kernel. Each voxel's signal was converted to percent change by normalizing it based on intensity.

General Linear Model Analysis

For each participant, we created a design matrix modeling experimental events and including events of no interest. At the time of an experimental event, we defined an impulse and convolved it with a hemodynamic response. The following regressors were included in the model: (a) an indicator variable marking the occurrence of all auditory tone/envelope events; (b) an indicator variable marking the occurrence of all jump events; (c) a parametric regressor indicating the change in distance to subgoal induced by each jump, mean-centered; (d) a parametric regressor indicating the change in distance to goal induced by each jump, mean-centered; (e and f) indicator variables marking subgoal and goal attainment; (g) a variable marking all periods of task performance, from the initial presentation of the icons to the end of the trial; (h) an indicator variable for delivery of monetary reward (encompassing the positive, 25, negative, -25, and neutral, 0, events); (i) an indicator variable for the positive reward, 25; and (j) an indicator variable for the negative reward, -25. Also included were head motion parameters and first- to third-order polynomial regressors to

Table 7. Unsigned Subgoal-related RPE. Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. superior parietal lobule	278	.00	3.78	-28, +73, +47
R. culmen	59	.00	3.48	-13, +67, -6

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

Table 8. Pause Event. Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. superior frontal G. (cluster extends to anterior medial surface)	9555	.40	6.00	−1, −37, +56
L. inferior parietal lobule	8646	−.46	−5.55	+46, +46, +53
L. superior frontal G.	1642	−.41	−0.60	+37, −40, +32
R. middle frontal G.	686	−.34	−3.84	−34, −43, +35
R. inferior frontal G.	684	−.28	−5.69	−49, −16, 0
L. cuneus	468	.43	5.37	+1, +9, +5
L. inferior temporal G.	63	−.13	−4.19	+55, +58, −3

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

regress out scanner drift effects. A global signal regressor was also included. In additional analyses, instead of indicator variables encompassing signed positive and negative events, we separated regressors for positive and negative events or included them in an unsigned way, with one regressor for the jump RPEs and one regressor for the monetary RPEs. All parametric regressors were mean-centered. The estimates from the general linear model were normalized to Talairach space (Talairach & Tournoux, 1988).

Group Analysis

For each regressor and for each voxel, we tested the sample of 40 participant-specific coefficients against zero in a two-tailed t test. We defined a threshold of $p = .01$ and applied correction for multiple comparison based on cluster size using Monte Carlo simulations based on estimates of spatial autocorrelation function as implemented

in AFNI's 3dClustSim (Version 2018). We report results at a corrected $p < .01$.

ROI Analysis

We followed the same procedure as in Experiment I.

Results

Behavior

A trial lasted, on average, 19.81 steps ($SEM = 0.40$ steps). A linear regression of types of prediction errors on number of steps revealed a significant effect of goal-related RPE ($\beta = -1.52$, $p < .01$), and no effect of subgoal-related RPE ($\beta = 0.18$, $p = .06$). Average RT was 1160 msec ($SEM = 30$ msec). The pause happened, on average, at 5.57 steps ($SEM = 0.08$ steps). Average RT for the first movement after pause events was 1460 msec ($SEM = 30$ msec). A linear regression of RTs of pause

Table 9. Displacement Distance. Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. superior frontal G.	106	.00	2.90	−28, −4, +65
L. inferior parietal lobule	96	.00	4.26	+52, +43, +53
L. superior temporal G.	81	.00	2.86	−49, −10, +2
R. precuneus	71	.00	3.11	−1, +55, +47
L. postcentral G.	63	−.00	−3.33	+28, +31, +65
L. superior frontal G.	59	.00	3.55	+37, −52, +26
R. supramarginal G.	50	−.00	3.49	−61, +46, +29
R. superior temporal G.	49	−.00	−3.42	−52, +28, +14

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

Table 10. All Jumps. Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. precuneus	4438	.43	5.41	-10, +73, +53
L. precentral G.	925	-.14	-3.18	+58, -4, +5
L. middle frontal G.	758	.19	4.07	-19, -10, +62
L. anterior cingulate	717	-.18	-3.43	+1, -16, -6
L. precentral G.	398	-.15	-5.06	+34, +19, +65
R. middle frontal G.	222	.12	3.76	-31, -34, +41
L. inferior frontal G.	198	-.15	-2.71	+43, -25, -12
L. uncus	194	-.11	-3.78	+16, +7, -21
R. transverse temporal G.	165	-.10	-3.59	-64, +13, +11
L. middle frontal G.	139	.13	4.25	+40, -31, +38
L. declive	121	-.09	-2.87	+13, +73, -15
L. caudate	119	.07	5.34	+16, -10, +8
L. parahippocampal G.	98	.07	5.52	+28, +46, -6
L. cingulate G.	89	-.07	-3.47	+1, +25, +41
L. postcentral G.	84	-.13	-3.38	+1, +49, +23
R. middle temporal G.	76	-.12	-2.99	-61, +4, -21
R. inferior frontal G.	70	-.13	-3.21	-40, -22, -12
R. parahippocampal G.	68	-.08	-2.91	-16, +4, -9
R. superior temporal G.	62	-.08	-2.86	-58, -1, +5
R. caudate	58	.04	4.76	-16, -7, +14
R. parahippocampal G.	53	.05	3.63	-31, +34, -9
L. precuneus	48	-.07	-3.08	+1, +70, +23

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

Table 11. +25 and -25, Compared with +25, 0, and -25 Points (Unsigned Monetary Probabilistic Reward Independent of Delivery; see Figure 6). Experiment II. Whole Brain

Area	Size (Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
R. superior temporal G.	149	.05	3.06	-67, +19, +5
L. anterior insula	106	.08	3.43	+34, -4, -9
L. middle occipital G.	96	-.07	-4.08	+49, +76, 0
R. superior temporal G.	60	.11	3.75	-61, +4, +2
R. fusiform G.	56	.10	4.00	-34, +43, -15
R. cuneus	52	-.04	-4.28	-10, +73, +23
R. middle frontal G.	49	.05	4.25	-34, -4, +32

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

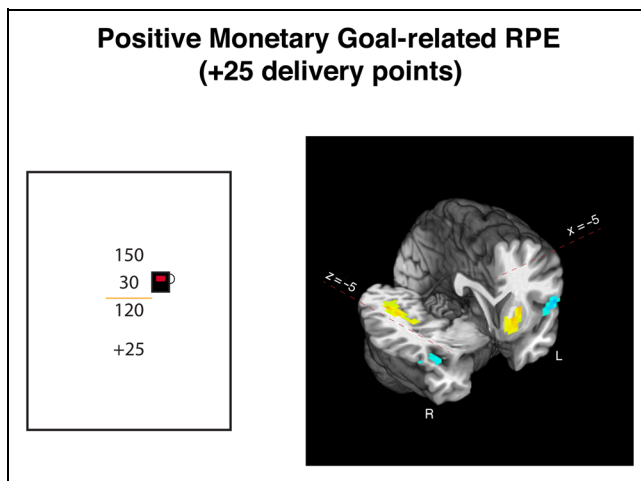


Figure 6. Eliciting positive RPEs with monetary outcomes. In Experiment II, at the end of each trial, a third of trials participants would get +25 delivery points, which would later be converted to U.S. dollars. We observed left ventral putamen increases to tip, compared with outcome (+25, 0, or -25). In addition, we observed bilateral decreases in response in superior temporal gyrus and increases in fusiform gyrus. $p < .01$, cluster-corrected (see Table 12 for coordinates).

events revealed a significant increase in RTs to jumps (mean regression coefficient = 0.07, $SEM = 0.02$; $t(39) = 3.28$, $p < .005$, two-tailed t test). The same regression also revealed that RTs were significantly slower as displacement distance increased (mean regression coefficient = 0.04, $SEM = 0.01$; $t(39) = 4.22$, $p < .001$). No significant effect of subgoal-related RPE or goal-related RPE on RT was observed ($p = .15$ and $p = .78$).

Mean accuracy across all steps was 71.68° ($SEM = 0.21^\circ$), and for the step immediately after the pause event, the accuracy was 35.69° ($SEM = 1.1^\circ$). A linear regression on the accuracy scores on the step succeeding the pause event revealed a significant increase in deviations from the optimal path in the jump condition (mean regression coefficient = 0.08, $SEM = 0.02$; $t(39) = 4.43$, $p < .05$). We also observed that the extent of deviation from optimal path increased with displacement distance (mean regression coefficient = 0.03, $SEM = 0.01$; $t(39) =$

2.11, $p < .05$). Similar to what we observed with RT data, no significant effects of subgoal-related RPEs or goal-related RPEs were observed ($p = .25$ and $p = .11$).

Whole-brain Analysis

We observed an increase in BOLD response in left mPFC to distance-driven unsigned goal-related prediction errors ($M = 1.0 \times 10^{-3}$, $p < .01$, cluster-corrected; Figure 5 and Table 6). Surprisingly, in contrast with the results of the previous experiment and our past study, no response was observed to unsigned subgoal-related RPEs, even at a liberal threshold (Table 7). Regression models with signed regressors yielded results consistent with the models with unsigned responses. Results for control regressors are provided in Tables 8 to 10.

After each delivery, a probabilistic monetary reward was delivered: +25, 0, -25, with equal probability. In contrast with the unsigned distance-driven RPE and contrary to our expectations, we observed no medial prefrontal activity to signed or unsigned monetary RPEs ((+25, -25) compared with (+25, 0, -25)), even at liberal thresholds (see Table 11). Positive monetary RPEs (+25) yielded an increase in BOLD response in left putamen activity, on the border between ventral and dorsal striatum (see Figure 6 and Table 12 for coordinates), relative to the common responses to all the possible monetary outcomes, +25, 0, -25. This was matched by a contralateral striatal cluster at a more liberal threshold. In addition, we observed increased activity in bilateral fusiform gyrus and decreased activity in bilateral superior temporal gyrus.

ROI Analysis

No significant response was observed in anatomically delineated VS to subgoal-related RPEs or distance-driven goal-related RPEs ($p > .05$). Consistent with whole-brain results, we did observe a significant response in the bilateral putamen, on the border with VS, to monetary positive RPEs ($p < .001$).

Table 12. +25 Points (Monetary Probabilistic Reward Independent of Delivery; see Figure 6). Experiment II. Whole Brain

Area	Size (No. of Voxels)	Peak Voxel		
		Parameter Estimate	<i>t</i> Statistic	Coordinates (<i>x</i> , <i>y</i> , <i>z</i>)
L. superior temporal G.	367	-.24	-4.32	+64, +25, +14
R. superior temporal G.	352	-.27	-5.01	-64, +13, +8
R. fusiform G.	201	.09	2.81	-40, +49, -21
L. fusiform G.	94	.08	5.69	+28, +49, -12
L. lentiform nucleus	41	.10	3.65	+16, -1, -6

Primary threshold $p < .01$, cluster-corrected to $p < .01$, df 39. Labels provided by Talairach Daemon. Coordinates in Talairach space and DICOM order. G. = gyrus; R. = right; L. = left.

We tested for unsigned RPEs in anatomically defined amygdalar complex. We found no significant response to distance-driven goal-related RPEs, subgoal-related RPEs, or to monetary goal-related RPEs ($p > .05$).

DISCUSSION

Our previous study, RF2011, provided first evidence for the involvement of the mPFC in the processing of subgoal-related negative RPEs. However, that study left open two important questions: How do the neural correlates of positive RPEs compare with those of negative RPEs? And, how do the neural correlates of subgoal-related RPEs compare with those of goal-related RPEs? We have presented two experiments aimed at addressing these questions. Our experiments yielded surprising results that impact our understanding of the neural mechanisms of hierarchical reward processing.

We examined these two questions using a spatial navigation paradigm that explicitly incorporated goals and subgoals. Importantly, in this paradigm, the level of effort required to attain the subgoal and goal could be independently manipulated. This feature allowed us to selectively examine the neural correlates of positive and negative subgoal-related RPEs, together with the neural correlates of goal-related RPEs. Furthermore, the geometric features of the task allowed us to ensure that our findings were not driven simply by low-level perceptual or motor factors (RF2011).

Concerning the first question, in Experiment I we found an increase in BOLD response in mPFC to positive subgoal-related RPEs. Given that RF2011 found an increase in mPFC response to negative subgoal-related RPEs, the combined results suggest that mPFC may signal unsigned RPEs, rather than signed RPE per se. Unsigned responses in mPFC have been found consistently to RPEs, both in nonhuman electrophysiological and human neuroimaging studies (Hyman et al., 2017; Roesch et al., 2012; Bryden, Johnson, Tobia, Kashtelyan, & Roesch, 2011; Hayden et al., 2011). However, human electrophysiological studies investigating ERPs find signed, rather than unsigned, RPEs originating in the mPFC (Sambrook & Goslin, 2015)—though analysis in the time–frequency component associated with mPFC activity find unsigned responses (Mas-Herrero & Marco-Pallarés, 2014; Cavanagh, Figueroa, Cohen, & Frank, 2012). Our findings of unsigned subgoal-related RPEs are broadly consistent with HRL but suggest a somewhat different interpretation from the one considered in RF2011. Namely, rather than constituting the reinforcement term in RL models, these prediction errors could be used to modulate learning from events based on their saliency, as described in the attention to learning theory (Pearce, Kaye, & Hall, 1982). Our findings suggest that saliency-modulated learning extend to hierarchical settings, in particular, to subtask performance.

It is worth noting that other theories of mPFC function predict an unsigned response to prediction errors. In particular, two theories assign a central role to the finding of unsigned prediction errors in mPFC: the predicted response–outcome model (Alexander & Brown, 2011) and the reward value and prediction model (Silvetti, Seurinck, & Verguts, 2011). Other theories accommodate unsigned prediction errors as a byproduct of the purported function of the mPFC, namely, theories related to HRL (Shahnazian & Holroyd, 2018) and to the expected value of control (Shenhav, Botvinick, & Cohen, 2013).

The second question we addressed was whether goal-related RPEs and subgoal-related RPEs arise in the same location. In Experiment II, we simultaneously manipulated subgoal-related RPEs and goal-related RPEs. In this setting, we expected mPFC to show unsigned subgoal-related RPEs, in keeping with RF2011 and Experiment I, independently of unsigned goal-related RPEs.

Yet, in contrast with Experiment I and RF2011, we only observed the unsigned goal-related RPEs and no subgoal-related RPEs. Although this failure to observe an effect might stem from lack of experimental power (possibly due to the moderate correlation between the two regressors, $r = .31$), these discrepant findings may also result from differential task-related demands placed on the attentional system. Specifically, the data suggest that the RPE signals are generated based on the specific level of task structure that is currently attended, as determined by the specific contingencies of the task.

It could be that these effects were due to different task designs. However, in our view, the most salient difference between the task designs across the two experiments was the simultaneous presentation of subgoal- and goal-related RPEs, which may have drawn attention to the goal. Moreover, these changes do not appear to have introduced any confounding factors. Other than addition of tip and shortchange, which occurs at the end of the trial, all other aspects of the task were kept constant, namely, the way the truck movements were inputted by the joystick and the perceptual properties of a pause event and a jump.

According to a selective hierarchical interpretation of mPFC function, possibly due to attentional constraints, mPFC processed subgoal-related information only in the absence of competing information at the goal level. This interpretation resonates with findings of learning signals in mPFC that are selectively elicited by stimuli that are attended to (Akaishi, Kolling, Brown, & Rushworth, 2016). Notably, in some hierarchical learning algorithms, the agent attends with unequal priority to different levels of hierarchy (e.g., see “recursive optimality” in Dietterich, 2000). In other words, learning simultaneously at all levels of hierarchy is not a necessary condition of a hierarchical agent. Therefore, our results could still be compatible with hierarchical decision-making.

An alternative way to interpret the results of Experiment I is that mPFC may track the distance of the subgoal to the goal rather than the distance to the subgoal itself. Under this interpretation, mPFC evaluates the proximity of the subgoal to the goal. This would still be a hierarchical account, given that prediction errors would be computed with regard to a portion of the task and based on a subgoal. In fact, such an interpretation resonates well with another learning algorithm called skill chaining (Konidaris & Barto, 2009). Both interpretations are plausible given the data, and both implicate the mPFC in HRL.

Overall, it is interesting to compare our findings with the results of Diuk et al. (2013) where VS was sensitive to simultaneous prediction errors at two different levels of task hierarchy. In Diuk et al., the information pertinent to task levels were presented explicitly as two different stimuli, whereas in our study, this information must be inferred by attending to the change in the relative arrangement of stimuli resulting from jump displacement. Taken together, the results of Diuk et al. (2013) and Experiment II suggest that humans can be sensitive to two sources of information simultaneously, provided that these sources are presented separately.

In addition, we addressed whether the same regions that process subgoal-related effort-driven RPEs would process monetary RPEs. We found a dissociation between mPFC and VS: Whereas the former was sensitive to effort-driven RPEs (subgoal-related RPEs and goal-related RPEs, Experiment I), VS was only sensitive to monetary goal-related RPEs (Experiment II). The differential engagement of mPFC and VS is compatible with the HRL theory of mPFC function (Holroyd & Yeung, 2012), which contends that mPFC is highly engaged in tasks that require extended sequences of actions involving effortful behavior. Our spatial navigation task incorporates both features.

In conclusion, our results are consistent with the idea that mPFC processes prediction errors in hierarchical settings. More specifically, we show that (1) mPFC signals subgoal-related RPEs in an unsigned manner, (2) mPFC signals RPEs related to superordinate goals similarly, (3) whether mPFC's BOLD response reflects subgoal- or goal-related RPE is dependent on the specific task manipulation and is presumably determined by attentional factors, and (4) RPE signaling differs between mPFC and VS. We propose that such prediction errors are used to improve behavior at the level of subtasks, which can then be applied to different tasks. Given that ecological tasks are hierarchically structured, mPFC can be instrumental in extending reinforcement learning mechanisms to ecological settings.

Reprint requests should be sent to Matthew Botvinick, DeepMind, London, United Kingdom, and Gatsby Computational Neuroscience Unit, Alexandra House - 17 Queen Square, London WC1N 3AR, or via e-mail: botvinick@google.com.

REFERENCES

- Akaishi, R., Kolling, N., Brown, J. W., & Rushworth, M. (2016). Neural mechanisms of credit assignment in a multicue environment. *Journal of Neuroscience*, *36*, 1096–1112.
- Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344.
- Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*, 527–536.
- Balaguer, J., Spiers, H., Hassabis, D., & Summerfield, C. (2016). Neural mechanisms of hierarchical planning in a virtual subway network. *Neuron*, *90*, 893–903.
- Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, *9*, 16–27.
- Botvinick, M. M., Niv, Y., & Barto, A. G. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*, 262–280.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashtelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *Journal of Neuroscience*, *31*, 18266–18274.
- Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex*, *22*, 2575–2586.
- Chiang, F.-K., & Wallis, J. D. (2018). Neuronal encoding in prefrontal cortex during hierarchical reinforcement learning. *Journal of Cognitive Neuroscience*, *30*, 1197–1208.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, *13*, 227–303.
- Diuk, C., Tsai, K., Wallis, J., Botvinick, M. M., & Niv, Y. (2013). Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *Journal of Neuroscience*, *33*, 5797–5805.
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*, 312–325.
- Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience*, *31*, 4178–4187.
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*, 54–83.
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*, 122–128.
- Hyman, J. M., Holroyd, C. B., & Seamans, J. K. (2017). A novel neural prediction error found in anterior cingulate cortex ensembles. *Neuron*, *95*, 447–456.
- Konidaris, G., & Barto, A. G. (2009). Skill discovery in continuous reinforcement learning domains using skill chaining. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (pp. 1015–1023). Red Hook, NY: Curran Associates.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior; the Hixon symposium* (pp. 1–36). New York: Wiley.

- Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, 35, 287–308.
- Logan, G. D., & Crump, M. J. C. (2011). Hierarchical control of cognitive processes: The case for skilled typewriting. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 54, pp. 1–28). San Diego, CA: Elsevier.
- Mas-Herrero, E., & Marco-Pallarés, J. (2014). Frontal theta oscillatory activity is a common mechanism for the computation of unexpected outcomes and learning rate. *Journal of Cognitive Neuroscience*, 26, 447–458.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53, 139–154.
- Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian learning. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior* (Vol. 3, pp. 241–256). Cambridge, MA: Ballinger Publishing Company.
- Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., et al. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, 71, 370–379.
- Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *European Journal of Neuroscience*, 35, 1190–1200.
- Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, 141, 213–235.
- Satterthwaite, T. D., Ruparel, K., Loughead, J., Elliott, M. A., Gerraty, R. T., Calkins, M. E., et al. (2012). Being right is its own reward: Load and performance related ventral striatum activation to correct responses during a working memory task in youth. *Neuroimage*, 61, 723–729.
- Shahnazian, D., & Holroyd, C. B. (2018). Distributed representations of action sequences in anterior cingulate cortex: A recurrent neural network approach. *Psychonomic Bulletin & Review*, 25, 302–321.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79, 217–240.
- Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience*, 5, 75.
- Skvortsova, V., Palminteri, S., & Pessiglione, M. (2014). Learning to minimize efforts versus maximizing rewards: Computational principles and neural correlates. *Journal of Neuroscience*, 34, 15621–15630.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Umemoto, A., HajiHosseini, A., Yates, M. E., & Holroyd, C. B. (2017). Reward-based contextual learning supported by anterior cingulate cortex. *Cognitive, Affective and Behavioral Neuroscience*, 17, 642–651.
- Zarr, N., & Brown, J. W. (2016). Hierarchical error representation in medial prefrontal cortex. *Neuroimage*, 124, 238–247.