Distinct neural encoding of context-dependent error signals and context-dependent changes in behavior during adaptive learning

Chang-Hao Kao[1], Sangil Lee[1], Joshua I. Gold[2], Joseph W. Kable[1]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA

Corresponding authors:

Chang-Hao Kao: chakao@sas.upenn.edu

Joseph W. Kable: kable@psych.upenn.edu

Abstract

Learning effectively from errors requires using them in a context-dependent manner, for example adjusting to errors that result from unpredicted environmental changes but ignoring errors that result from environmental stochasticity. Where and how the brain represents errors in a context-dependent manner and uses them to guide behavior are not well understood. We imaged the brains of human participants performing a predictive-inference task with two conditions that had different sources of errors. Their performance was sensitive to this difference, including more choice switches after fundamental changes versus stochastic fluctuations in reward contingencies. Using multi-voxel pattern classification, we identified context-dependent representations of error magnitude and past errors in posterior parietal cortex. These representations were distinct from representations of the resulting context-dependent behavioral adjustments in dorsomedial frontal, anterior cingulate, and orbitofrontal cortex. The results provide new insights into human brain that represent and use errors in a context-dependent manner to support adaptive behavior.

Introduction

Errors often drive adaptive adjustments in beliefs that inform behaviors that maximize positive outcomes and minimize negative ones (Sutton & Barto, 1998). A major challenge to error-driven learning in uncertain and dynamic environments is that errors can arise from different sources that have different implications for learning. For example, a bad experience at a restaurant that recently hired a new chef might lead you to update your belief about the quality of the restaurant, whereas a similar experience at a well-known restaurant with a chef that has long been your favorite might be written off as a one-time bad night. That is, the same errors should be interpreted differently in different contexts. In general, errors that represent fundamental changes in the environment or that occur during periods of uncertainty should probably lead you to update your beliefs and change your behavior, whereas those that result from environmental stochasticity are likely better ignored (d'Acremont & Bossaerts, 2016; Li, Nassar, Kable, & Gold, 2019; Nassar, Bruckner, & Frank, 2019; O'Reilly et al., 2013).

Neural representations of key features of these kinds of dynamic, error-driven learning processes have been identified in several brain regions. For example, several studies focused on variables derived from normative models that describe the degree to which individuals should dynamically adjust their beliefs in response to error feedback under different task conditions, including the probability that a fundamental change in the environment just occurred (change-point probability, or CPP, which is a form of surprise) and the reducible uncertainty associated with estimates of environmental features (relative uncertainty, or RU). Correlates of these variables have been identified in dorsomedial frontal (DMFC) and dorsolateral prefrontal (DLPFC) cortex and medial and lateral posterior parietal cortex (PPC) (Behrens, Woolrich, Walton, & Rushworth, 2007; McGuire, Nassar, Gold, & Kable, 2014; Nassar, McGuire, Ritz, & Kable, 2019). These and other studies also suggest specific roles for these different brain regions in error-driving learning, including representations of surprise induced by either state changes or outliers (irrelevant to state changes) in the PPC that suggest a role in error monitoring (Nassar, Bruckner, et al., 2019; O'Reilly et al., 2013), and representations of variables more closely related to belief and behavior updating in the prefrontal cortex (PFC) (McGuire et al., 2014; O'Reilly et al., 2013). However, these previous studies, which typically used continuous rather than discrete feedback, were not designed to identify neural signals related to a key aspect of

flexible learning in uncertain and dynamic environments: responding to the same exact errors differently in different contexts.

To identify such context-dependent neural responses to errors, we adapted a paradigm from our previous single-unit recording study (Li et al., 2019). In this paradigm, we generated two different dynamic environments by varying the amount of noise and the frequency that change-points occur (i.e., hazard rate; Behrens et al., 2007; Glaze, Kable, & Gold, 2015; Nassar et al., 2012; Nassar, Wilson, Heasly, & Gold, 2010). In the unstable environment, noise was absent and the hazard rate was high, and thus errors unambiguously signaled a change in state. In the high-noise environment, noise was high and the hazard rate was low, and thus small errors were ambiguous and could indicate either a change in state or noise. Thus, effective learning requires treating errors in the two conditions differently, including adjusting immediately to errors in the unstable environment but using the size of errors and recent error history as cues to aid interpretation of ambiguous errors in the high-noise condition.

In our previous study, we found many single neurons in the anterior cingulate cortex (ACC) or posterior cingulate cortex (PCC) that responded to errors or the current context, but we found little evidence that single neurons in these regions combined this information in a context-dependent manner to discriminate the source of errors or drive behavior. In the current study, we used whole-brain fMRI and multi-voxel pattern classification to identify context-dependent neural responses to errors and activity predictive of context-dependent behavioral updating in the human brain. The results show context-dependent encoding of error magnitude and past errors in PPC and encoding of behavioral shifts in a large array of frontal regions including ACC, DMFC, DLPFC and orbitofrontal cortex (OFC), which provide new insights into the distinct roles these brain regions play in representing and using, respectively, errors in a context-dependent manner to guide adaptive behavior.

Results

Sixteen human participants performed a predictive-inference task (Figure 1A) while fMRI was used to measure their blood-oxygenation-level-dependent (BOLD) brain activity. The task required them to predict the location of a single rewarded target from a circular array of ten targets. The location of the rewarded target was sampled from a distribution based on the location of the current best target and the noise level in the current condition. In addition, the

location of the best target could change according to a particular, fixed hazard rate (H). Two conditions with different noise levels and hazard rates were conducted in separate runs. In the high-noise condition (Figure 1B–C), the rewarded target would appear in one of the five locations relative to the location of the current best target, and the hazard rate was low ($H = 0.02$). In the unstable condition (Figure 1D–E), the rewarded target always appeared at the location of the best target, and the hazard rate was high ($H = 0.35$). On each trial, participants made a prediction by looking at a particular target, and then were given explicit, visual feedback about their chosen target and the rewarded target. Effective performance required them to use this feedback in a flexible and context-dependent manner, including typically ignoring small errors in the high-noise condition but responding to small errors in the unstable condition by updating their beliefs about the best-target location.

Behavior

Nearly all of the participants' choice patterns were consistent with a flexible, context-dependent learning process (closed symbols in Figure 2). On average, they learned the location of the best target after a change in its location more quickly and reliably in the unstable versus high-noise condition (Figure 2A). This flexible learning process had two key signatures. First, target switches (i.e., predicting a different target than on the previous trial) tended to follow errors of any magnitude in the unstable condition but only errors of high magnitude (i.e., when the chosen target was 3, 4, or 5 targets away from the rewarded target) in the high-noise condition (sign test for H_0: equal probability of switching for the two conditions; error magnitude of 1: median = -0.35, interquartile range (IQR) = [-0.62, -0.25], $p<0.001$; error magnitude of 2: median = -0.30, IQR = [-0.70, -0.11], $p<0.001$; Figure 2B–C). Second, target switches depended on error history only for low-magnitude errors (i.e., when the chosen target was 1 or 2 targets away from the rewarded target) in the high-noise condition but not otherwise (sign test for H_0: switching was unaffected when recent history contained fewer errors; error magnitude of 1: median = -0.29, IQR = [-0.42, -0.10], $p=0.004$; error magnitude of 2: median = -0.25, IQR = [-0.38, -0.14], $p<0.001$; Figure 2D–F).

We accounted for these behavioral patterns with a reduced Bayesian model that is similar to ones we have used previously to model belief updating in a dynamic environment (open symbols in Figure 2; Tables 1 and 2). According to this model, the decision-maker's trial-by-trial

choices are governed by ongoing estimates of the probability that the best target changed (change-point probability, or CPP) and reducible uncertainty about the best target's location (relative uncertainty, or RU). Both quantities are influenced by the two free parameters in the model, subjective hazard rate and noise level, which were fitted separately in each condition for each participant. As expected, the fitted hazard rates were higher in the unstable condition than in the high-noise condition, although both tended to be higher than the objective values, as we have observed previously (Nassar et al., 2010). The fitted noise estimates were not reliably different between the high-noise versus unstable condition (Table 2).

In the reduced Bayesian model, both CPP and RU contribute to processing errors in a context-dependent manner. CPP increases as the current error magnitude increases and achieves high values more quickly in the unstable condition because of the higher hazard rate (Figure 3A). These dynamics lead to a greater probability of switching targets after smaller errors in the unstable condition. RU increases on the next trial after the participant makes an error and does so more in the high-noise condition because of the lower hazard rate (Figure 3B). These dynamics lead to a greater probability of target switches when the last trial was an error, which is most prominent for small errors in the high noise condition. Thus, CPP and RU each account for one of the two key signatures of context-dependent learning that we identified in participants' behavior, with CPP driving a context-dependent influence of error magnitude and RU driving a context-dependent influence of error history on target switches (Figure 3C).

We also tested several alternative models but they did not provide as parsimonious descriptions of the data (Figure 2 – figure supplement 1, and Tables 1 and 2). Notably, an alternative model that assumed a condition-specific fixed learning rate also assumed errors were treated differently for the two conditions but did not include trial-by-trial adjustments of learning rates used by the reduced Bayesian model. Thus, although this model performed better than the reduced Bayesian model in the unstable condition, it cannot capture participants' behaviors in the high-noise condition, where dynamically integrating both current and past errors is required for adapting trial-by-trial behavior. Other hybrid models performed worse than the reduced Bayesian model under both conditions.

Neural representation of CPP and RU

To compare our current data directly to our previously identified neural representations of CPP and RU (McGuire et al., 2014), the two key quantities in the reduced Bayesian model, we conducted univariate analyses of our imaging data using those behaviorally derived variables as regressors. This comparison also allowed us to better isolate representations of these variables from those related to visual and motor processing demands that differed considerably for the two tasks (the other task included a more complex visual scene and used hand, not eye, movements). Similar to our previous findings, we found activity that was positively correlated with the levels of CPP and RU across DLPFC and PPC (Figure 3D). We identified these joint neural representations of CPP and RU in the high-noise condition, because both CPP and RU varied across trials in this condition, in contrast to the unstable condition in which RU did not vary. The regions of DLPFC and PPC that were responsive to both CPP and RU were a subset of those identified as showing this conjunction in our previous study (Figure 3E, Figure 3 – figure supplement 1).

Because CPP and RU both contribute to responding to errors in a context-dependent manner, we considered the brain regions that responded to both variables as good candidates for encoding errors in a context-dependent manner that is linked to subsequent behavioral shifts. In the following analyses, we aimed to directly identify context-dependent neural representations of error magnitude and error history, as well as activity that predicts subsequent shifts in behavior, in these and other brain regions.

Context-dependent neural representation of errors

We used multi-voxel pattern analysis (MVPA) to identify error-related neural signals that were similar and different for the two task conditions. Given the two key signatures of flexible learning that we identified in behavior, we were especially interested in identifying neural representations of error magnitude and past errors that were stronger in the high-noise than the unstable condition.

We found robust, context-dependent representations of the magnitude of the error on the current trial in PPC. Consistent with the context-dependent behavioral effects, this representation of error magnitude was stronger in the high-noise than the unstable condition (Figure 4 and Table 3). Specifically, we could classify correct versus error feedback on the current trial across almost the entire cortex, in both the unstable and noisy conditions. However, for error trials, we

could classify error magnitude (in three bins: 1, 2, 3+ targets away from the rewarded target) only for the high-noise condition and most strongly in the lateral and medial parietal cortex and in the occipital pole. In a parallel set of analyses, we found that univariate activity in PPC also varied in a context-dependent way, responding more strongly to error magnitude in the high-noise than the unstable condition (Figure 4 – figure supplement 1).

We also found robust, context-dependent representations of past errors in PPC. These representations also were stronger in the high-noise than the unstable condition, particularly on trials for which past errors had the strongest influence on behavior. Specifically, we could classify correct versus error on the previous trial in PPC for both task conditions (Figure 5). This classification of past errors depended on the outcome of the current trial. We separated trials according to whether the current feedback was correct or an error, or whether the error magnitude provided ambiguous (error magnitudes of 1 or 2) or unambiguous (error magnitudes of 0 or 3+) feedback in the high-noise condition (Figure 5). We found reliable classifications of past errors in the lateral and medial parietal cortex in both conditions for correct trials and unambiguous feedback. Moreover, these representations depended on the current context, and, consistent with behavioral effects of error history, were stronger for error trials and ambiguous feedback in the high-noise than in the unstable condition (Table 3). These context-dependent signals for past errors were not clearly present in univariate activity (Figure 5 – figure supplement 1). An additional conjunction analysis across MVPA results showed that PPC uniquely encoded context-dependent error signals for both error magnitude of the current trials and past errors when the current trial provided ambiguous feedback (Table 3).

Neural prediction of subsequent changes in behavior

Although PPC responds to errors in a context-dependent manner that could be used for determining behavioral updates, we did not find that activity in this region was predictive of the participants' future behavior. Instead, we found such predictive activity more anteriorly throughout the frontal lobe. Specifically, we investigated whether multi-voxel neural patterns could predict participants' target switches on the subsequent trial. We focused on the trials with small error magnitudes (1 or 2) in the high-noise condition, because these were the only trial types that participants consistently exhibited an intermediate probability of switching (20–80%, Figure 2). We found that activity patterns in widespread regions in OFC, ACC, DMFC, and

DLPFC could predict subsequent stay/switch decisions (Figure 6, Table 4). We did not find any regions where univariate activity reliably predicted participants' subsequent behavior (Figure 6 – figure supplement 1).

Discussion

We identified context-dependent neural representations of errors in humans performing a dynamic learning task. The task required participants to learn in two different dynamic environments. In the unstable condition (high hazard rate and low noise), errors unambiguously indicated a change in the state of the environment, and participants reliably updated their behavior in response to errors. In contrast, in the high-noise condition (low hazard rate and high noise), small errors were ambiguous, and participants used both the current error magnitude and recent error history to distinguish between those errors that likely signal change-points and those likely arising from environmental noise. Using MVPA, we showed complementary roles of PPC and prefrontal regions (including OFC, ACC, DMFC and DLPFC) in the outcome-monitoring and action-selection processes underlying these flexible, context-dependent behavioral responses to errors. Neural patterns in PPC encoded the magnitude of errors and past errors, more strongly in the high-noise than the unstable condition. These context-dependent neural responses to errors in PPC were not reliably linked to subsequent changes in behavior. In contrast, neural patterns in prefrontal regions could predict subsequent changes in behavior (whether participants switch their choice on the next trial or not) in response to ambiguous errors in the high-noise condition.

Context-dependent behavior adaptation

Consistent with previous studies of ours and others (d'Acremont & Bossaerts, 2016; McGuire et al., 2014; Nassar, Bruckner, et al., 2019; Nassar et al., 2012; Nassar et al., 2010; O'Reilly et al., 2013; Purcell & Kiani, 2016), human participants adapted their response to errors differently in different contexts. In the unstable condition, participants almost always switched their choice after errors and quickly learned the new state after change-points. In contrast, in the high-noise condition, participants ignored many errors and only slowly learned the new state after change-points. In this condition, participants had to distinguish true change-points from environmental noise, and they used error magnitude and recent error history as a cue for whether the state had recently changed or not. These flexible and context-dependent responses to errors

could be accounted for by a reduced Bayesian model (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). In this model, beliefs and behavior are dynamically updated according to two key quantities, CPP and RU.

Neural representation of change-point probability and relative uncertainty

Replicating our previous work (McGuire et al., 2014), we identified neural activity correlated with both CPP and RU in PPC and DLPFC. This replication shows the robustness of these neural representations of CPP and RU across experimental designs that differ dramatically in their visual stimuli and motor demands, yet share the need to learn in dynamic environments with similar statistics. In addition, given that CPP and RU account for the context-dependent behavioral responses to error magnitude and recent error history, respectively, the regions responding to both CPP and RU are strong candidates for neural representations of errors and subsequent behavioral updates that are context dependent.

Context-dependent neural representation of errors

Advancing beyond previous work, we identified context-dependent encoding of errors in neural activity in the PPC. Mirroring the context dependence of behavior, the multivariate neural pattern in PPC encoded current error magnitude more strongly in the high-noise condition than in the unstable condition and encoded past errors more strongly on trials that provided ambiguous feedback in the high-noise condition. That is, the multivariate pattern in PPC could distinguish between the same exact error stimuli depending on the context. These same regions of PPC have been shown previously to represent errors, error magnitudes, surprise and salience (Fischer & Ullsperger, 2013; Gläscher, Daw, Dayan, & O'Doherty, 2010; McGuire et al., 2014; Nassar, Bruckner, et al., 2019; Nassar, McGuire, et al., 2019; O'Reilly et al., 2013; Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013). In addition, these regions have been shown to integrate recent outcome or stimulus history in human fMRI studies (FitzGerald, Moran, Friston, & Dolan, 2015; Furl & Averbeck, 2011) and in animal single neuron recording studies (Akrami, Kopec, Diamond, & Brody, 2018; Brody & Hanks, 2016; Hanks et al., 2015; Hayden, Nair, McCoy, & Platt, 2008; Hwang, Dahlen, Mukundan, & Komiyama, 2017). Our results extend on these past findings by demonstrating that the neural encoding of errors in PPC is modulated across different contexts in precisely the manner that could drive adaptive behavior.

These whole-brain fMRI results complement our previous results recording from single neurons in ACC and PCC in the same task (Li et al., 2019). In that study, we identified single neurons in both ACC and PCC that encoded information relevant to interpreting errors, such as the magnitude of the error or the current context. However, we did not find any neurons that combined this information in a manner that could drive adaptive behavioral adjustments. Our whole-brain fMRI results suggest that PPC would be a good place to look for context-dependent error representations in single neurons, including a region of medial parietal cortex slightly dorsal to the PCC area we recorded from previously.

Neural representations of context-dependent behavioral updating

Also advancing beyond previous work, we identified neural activity predictive of context-dependent behavioral updates in DLPFC and across the frontal cortex. In the high-noise condition, small errors provided ambiguous feedback that could reflect either a change in state or environmental noise. Accordingly, after small errors in the high-noise condition, participants exhibited variability across trials in whether they switched from their current choice on the subsequent trial or not. In these ambiguous situations, the multivariate neural pattern across frontal regions, including OFC, ACC, DMFC and DLPFC, predicted whether people switched or stayed on the subsequent trial. That is, the multivariate pattern in frontal regions could distinguish whether people would update their behavior or not in response to the same exact error stimuli. These results suggest a dissociation between PPC regions that monitor error information in a context-dependent manner and frontal regions that may use this information to update beliefs and select subsequent actions.

This ability to decode subsequent choices might arise from different kinds of representations in different areas of frontal cortex. Activity in DMFC reflects the extent of belief updating in dynamic environments (Behrens et al., 2007; Hampton, Bossaerts, & O'Doherty, 2006; McGuire et al., 2014; O'Reilly et al., 2013), and the multivariate pattern in this region can decode subsequent switching versus staying in a reversal learning task (Hampton & O'Doherty, 2007). OFC and DMFC encode the identity of the current latent state in a mental model of the task environment and neural representations in these regions changes as the state changes (Chan, Niv, & Norman, 2016; Hunt et al., 2018; Karlsson, Tervo, & Karpova, 2012; Nassar, McGuire, et al., 2019; Schuck, Cai, Wilson, & Niv, 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014).

Neural activity in frontopolar cortex (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006) and DMFC (Blanchard & Gershman, 2018; Kolling, Behrens, Mars, & Rushworth, 2012; Kolling et al., 2016; Muller, Mars, Behrens, & O'Reilly, 2019) increases during exploratory choices, which occur more frequently during periods of uncertainty about the most beneficial option. In a recent study, we identified distinct representations of latent states, uncertainty, and behavioral policy in distinct areas of frontal cortex during learning in a dynamic environment (Nassar et al., 2019). Our results extend these past findings and demonstrate the role of these frontal regions in adjusting behavior in response to ambiguous errors.

Conclusion

People adapt their behavior in response to errors in a context-dependent manner, distinguishing between errors that indicate change-points in the environment versus noise. Here we used MVPA to identify two distinct kinds of neural signals contributing to these adaptive behavioral adjustments. In PPC, neural patterns encoded error information in a context-dependent manner, depending on error magnitude and past errors only under conditions where these were informative of the source of error. In contrast, activity in frontal cortex could predict subsequent choices that could be based on this information. These findings suggest a broad distinction between outcome monitoring in parietal regions and action selection in frontal regions when learning in dynamic and uncertain environments.

Materials and Methods

Participants

All procedures were approved by University of Pennsylvania Internal Review Board. We analyzed data from sixteen participants (9 female, mean age = 23.5, SD = 4.3, range = 18–33 years) recruited for the current study. One additional participant was excluded from analyses because of large head movements during MRI scanning (>10% of timepoint-to-timepoint displacements were >0.5 mm). All participants provided informed consent before the experiment. Participants received a participation fee of $15, as well as extra incentives based on their performance (mean = $15.09, SD = $2.26, range = $8.5–17.5).

Task

Participants performed a predictive-inference task during MRI scanning. On each trial, participants saw a noisy observation sampled from an unobserved state. The participants' goal was to predict the location of the noisy observation. To perform this task well, however, they should infer the location of the current state.

In this task (Li et al., 2019), there were 10 targets aligned in a circle on the screen (Figure 1A). At the start of each trial, participants had to fixate a central cross for 0.5 seconds to initialize the trial. After the cross disappeared, participants could choose one of 10 targets (red) by looking at it within 1.5 seconds and keeping fixation on the chosen target for 0.3 seconds. Then, an outcome would be shown for 1 second. During the outcome phase, a green dot indicated the chosen target. A purple or cyan target indicated the rewarded target, with color denoting 10 or 20 points of reward value, respectively. At the end of experiment, every 75 points were converted to $0.25 as participants' extra incentives.

Participants performed this task in two dynamic conditions separated into two different runs: a high-noise condition and an unstable condition. In the high-noise condition, the rewarded target could be one of five targets, given the underlying state (Figure 1B). The rewarded target probabilities for the relative locations ([-2, -1, 0, 1, 2]) of the current state were [0.05, 0.15, 0.6, 0.15, 0.05]. Thus, the location of the current state was most likely rewarded, but nearby targets could also be rewarded. Occasionally, the state would change its location with a hazard rate of 0.02 (Figure 1C). When a change-point happens, the new state would be selected among the ten targets based on a uniform distribution. In the unstable condition, there was no noise (Figure 1D). That is, the location of the state would be always rewarded. However, the state was unstable, as the hazard rate in this condition was 0.35 (Figure 1E). There were 300 trials in each run.

Behavior analysis

We investigated how participants' used error feedback flexibly across different contexts. Before the behavioral analysis, we removed two different kinds of trials. First, we removed trials in which participants did not make a choice within the time limit (Unstable: median number of trials = 10.5, range = 1–83; High-noise: median = 10, range = 2–88). Second, we also removed trials in which the location of the chosen target was not on the shortest distance between the previously chosen and previously rewarded targets (Unstable: median = 3, range = 0–24; High-noise: median = 17, range = 5–37). These trials implied that participants might have lost track of

the most recently rewarded target and cannot be captured by any of the belief updating models we tested.

First, we investigated how fast participants learned the location of the current state. For each condition and participant, we binned trials from trial 0 to trial 20 after change-points. Then, we calculated the probability of choosing the location of the current state for each bin.

Second, we examined how different magnitudes of errors lead to shifts in behavior. For each condition and participant, we binned trials based on the current error magnitude (from 0 to 5). Then, for each bin, we calculated the probability that participants switch their choice to another target on the subsequent trial. We hypothesized participants would have a lower probability of switching after small error magnitudes (1 or 2) in the high-noise condition than in the unstable condition since such errors could be due to environment noise in the high-noise condition but would signal a state change in the unstable condition.

Third, we further investigated how error history influenced participants' behavioral shifts. Similarly, we binned trials based on the current error magnitude and the error history of the last three trials. Here, we used four bins of error magnitudes (0, 1, 2, 3+). Based on the outcome of correct or error on the last three trials, there were 8 types of error history. For each error magnitude, we calculated the probability of switching for each type of error history. We hypothesized that participants in the high-noise condition would tend to switch their choice after small errors more if they had made more errors recently. To test this hypothesis, we ordered the 8 types of error history based on the number of recent errors and calculated the slope of probability of switching against the order of error history. A negative slope means that participants tend to switch as they receive more recent errors.

Behavior modeling

We fit several different computational models to participants' choices to evaluate which ones could best account for their behavior in the task.

Reduced Bayesian (RB) model

Previous studies have shown that a reduced Bayesian model, which approximates the full Bayesian ideal observer, could account well for participants' behavior in dynamic environments

similar to the current task (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). In this model, belief is updated by a delta rule:

$$\delta_t = x_t - B_t \quad (1)$$

$$B_{t+1} = B_t + \alpha_t \delta_t \quad (2)$$

where B_t is the current belief and x_t is the current observation. The new belief (B_{t+1}) is formed by updating the old belief according to the prediction error ($x_t - B_t$) and a learning rate (α_t). The learning rate controls how much a participant revises their belief based on the prediction error. In this model, the learning rate is adjusted on a trial-by-trial basis according to:

$$\alpha_t = \Omega_t + (1 - \Omega_t)\tau_t \quad (3)$$

where Ω_t is the change-point probability and τ_t is the relative uncertainty. That is, α_t is high as either Ω_t or τ_t is high. The change-point probability is the relative likelihood that the new observation represents a change-point as opposed to a sample from the currently inferred state (Nassar et al., 2010):

$$\Omega_t = \frac{U(x_t|1,10)H}{U(x_t|1,10)H + f_p(x_t|\gamma_t, B_t)(1-H)} \quad (4)$$

where H is the hazard rate, $U(x_t|1,10)$ is the probability of outcome derived from a uniform distribution, and $f_p(x_t|\gamma_t, B_t)$ is the probability of outcome derived from the current predictive distribution. That is, $U(x_t|1,10)$ reflects the probability of outcome when a change-point has occurred while $f_p(x_t|\gamma_t, B_t)$ reflects the probability of outcome when the state has not changed. The predictive distribution is an integration of the state distribution and the noise distribution:

$$f_p(X|\gamma_t, B_t) = C \times P(X|B_t)^{\gamma_t} \times P(X|B_t) \quad (5)$$

where X is a random variable determining the locations of target, $P(X|B_t)$ is the noise distribution in the current condition, $P(X|B_t)^{\gamma_t}$ is the state distribution, γ_t is the expected run

length after the change-point, and C is a normalizing constant to make the sum of probabilities in the predictive distribution equal one. Thus, the uncertainty of this predictive distribution comes from two sources: the uncertainty of the state distribution (σ_s^2) and the uncertainty of the noise distribution (σ_N^2). The uncertainty of the state distribution would decrease as the expected run length increases.

The expected run length reflects the expected number of trials that a state remains stable, and thus is updated on each trial based on the change-point probability (Nassar et al., 2010):

$$\gamma_{t+1} = (\gamma_t + 1)(1 - \Omega_t) + \Omega_t \quad (6)$$

where the expected run length is a weighted average conditional on the change-point probability. If no change-point occurs (i.e., change-point probability is low), the expected run length would increase, leading the uncertainty of the state distribution to decrease. That is, as more observations from the current state are received, participants are more certain about the location of the current state. However, if the change-point probability is high, which signals a likely change in the state, the expected run length would be reset to 1. Thus, the uncertainty of the state distribution becomes large. Participants are more uncertain about the current state after a change-point.

The other factor influencing the learning rate is the relative uncertainty, which is the uncertainty regarding the current state relative to the irreducible uncertainty or noise (McGuire et al., 2014; Nassar et al., 2012):

$$\tau_{t+1} = \frac{\Omega_t\sigma_N^2 + (1-\Omega_t)\sigma_s^2 + \Omega_t(1-\Omega_t)[\delta_t(1-\tau_t)]^2}{\Omega_t\sigma_N^2 + (1-\Omega_t)\sigma_s^2 + \Omega_t(1-\Omega_t)[\delta_t(1-\tau_t)]^2 + \sigma_N^2} \quad (7)$$

The three terms in the numerator contribute to the uncertainty about the current state. The first term reflects the uncertainty conditional on the change-point distribution; the second term reflects the uncertainty conditional on the non-change-point distribution; and the third term reflects the uncertainty due to the difference between the two distributions. The denominator shows the total variance which is the summation of the uncertainty about the current state and the

noise. As more precise observations are received in a given state, this relative uncertainty would decrease.

During model fitting, the noise distribution was approximated by the von Mises distribution, which is a circular Gaussian distribution:

$$P(x_t|B_t, K) = \frac{e^{K\cos(x_t - B_t)}}{\sum_{i=1}^{10} e^{K\cos(x_i - B_t)}} \quad (8)$$

where B_t is the location of the current belief, x_i is the location of target, and K controls the uncertainty of this distribution. When K is 0, this is a uniform distribution. As K increases, the uncertainty decreases. The denominator is used as a normalization term to make sure the sum of all the probabilities equals one. Thus, there are two free parameters in this model: hazard rate (H) and noise level (K). The range of hazard rate is between 0 and 1 and the noise level is greater than or equal to zero.

Fixed learning rate (fixedLR) model

We also consider an alternative model in which participants used a fixed learning rate in each of the two dynamic conditions. That is, the learning rate is the same over all trials in a condition. This model has one free parameter, the fixed learning rate (α_{fixed}), for each condition (Eq. 2). The fixed learning rate is between 0 and 1.

Hybrid of RB model and fixedLR model

Furthermore, we consider a hybrid model, in which the learning rate on each trial is a mixture of the learning rates from the RB model and the fixedLR model:

$$\alpha_t = w\alpha_{RB} + (1 - w)\alpha_{fixed} \quad (9)$$

where α_{RB} is the learning rate from the RB model and is varied trial by trial according to Ω_t and τ_t, α_{fixed} is the learning rate from the fixedLR model and w reflects the weight to integrate these two learning rates. In this model, there are four free parameters: hazard rate, noise level, fixed learning rate and weight. The weight is between 0 and 1.

Hybrid of RB model and P_{stay}

Finally, we consider a hybrid model, which combines the RB model with a fixed tendency to stay on the current target regardless of the current observation. Such a fixed tendency to stay was observed in monkeys in our previous study (Li et al., 2019). Here the belief is updated by:

$$B_{t+1} = B_t + [(1 - P_{stay}) \times \alpha_t (X_t - B_t) + P_{stay} \times 0] \quad (10)$$

where P_{stay} is the probability that participants stay on the current target. This model has three free parameters: hazard rate, noise level and the probability of stay. The probability of stay is between 0 and 1.

Model fitting and comparison

Each model was fitted within each participant and within each condition separately. Optimal parameters were estimated by minimizing the mean of the squared error (MSE) between a participant's prediction and the model prediction.

$$MSE = \frac{\sum_{t=1}^{n} (B_t - \widehat{B_t})^2}{n} \quad (11)$$

where *t* is the trial, *n* is the total number of included trials, B_t is a participant's prediction on trial *t,* and $\widehat{B_t}$ is the model prediction on trial *t*.

Since each model used a different number of parameters and each participant had a different number of included trials, we used Bayesian Information Criterion (BIC) to compare the performance of different models:

$$BIC = n \ln(MSE) + k \ln(n) \quad (12)$$

where n is the number of included trials and k is the number of free parameters in a model. A model with lower BIC performs better.

MRI Data Acquisition and Preprocessing

We acquired MRI data on a 3T Siemens Prisma with a 64-channel head coil. Before the task, we acquired a T1-weighted MPRAGE structural image (0.9375 X 0.9375 X 1 mm voxels, 192 X 256 matrix, 160 axial slices, TI = 1,100 ms, TR = 1,810 ms, TE = 3.45 ms, flip angle = 9°). During each run of the task, we acquired functional data using a multiband gradient echo-planar imaging (EPI) sequence (1.9592 X 1.9592 X 2 mm voxels, 98 X 98 matrix, 72 axial slices tilted 30° from the AC-PC plane, TR = 1,500 ms, TE = 30 ms, flip angle = 45°, multiband factor = 4). The scanning time (mean = 24.14 minutes, SD = 1.47, range = 21.85-30.00) for each run was dependent on the participants' pace. After the task, fieldmap images (TR = 1,270 ms, TE = 5 ms and 7.46 ms, flip angle = 60°) were acquired.

Data were preprocessed using FMRIB's Software Library (FSL) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004). Functional data were motion corrected using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), high-pass filtered with a Gaussian-weighted least square straight line fitting of $\sigma = 50\ s$, undistorted and warped to MNI space. To map the data to MNI space, boundary-based registration was applied to align the functional data to the structural image (Greve & Fischl, 2009) and fieldmap-based geometric undistortion was also applied. In addition, the structural image was normalized to the MNI space (FLIRT). Then, these two transformations were applied to the functional data.

fMRI analysis: univariate activity correlated with CPP and RU

Using similar procedures to our previous study (McGuire et al., 2014), we examined the effects of CPP and RU on univariate activity. Both the current study and the previous study investigate the computational process and neural mechanisms during learning in dynamic environments. The underlying task structures (which involved noisy observations and sudden change-points) are similar between the two studies, but the two studies used very different visual stimuli and motor demands. We specifically focused on the high-noise condition in the current study since it was more similar to the underlying structure, in terms of noisy observations and hazard rate of change-points, to our previous study.

We investigated the factors of CPP, RU, reward values and residual updates. The trial-by-trial CPP and RU were estimated from the RB model with subjective estimates of hazard rate and

noise. The residual update reflects the difference between the participants' update and the predicted update, and is estimated from a behavioral regression model in a similar manner as our previous study:

$$Update_t = \beta_0 + \beta_1\delta_t + \beta_2\delta_t\Omega_t + \beta_3\delta_t(1-\Omega_t)\tau_t + \beta_4\delta_t Reward + \varepsilon \quad (13)$$

where $Update_t$ is the difference between B_{t+1} and B_t, δ_t is the error magnitude, both Ω_t and τ_t were derived from the RB model with subjective estimates of hazard rate and noise, and the reward value indicated whether a correct response earned a large or a small value on that trial.

Then, a general linear model using these four factors was implemented on the neural data. Here we further smoothed the preprocessed fMRI data with a 6 mm FWHM Gaussian kernel. We included several trial-by-trial regressors of interest in the GLM: onsets of outcome, CPP, RU, reward value, and residual update. Six motion parameters were also included as confounds. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel $p<0.005$. Statistical testing was then based on the corrected cluster p value. For the conjunction analyses, we used the same procedure as the previous study (McGuire et al., 2014). We kept regions that passed the corrected threshold and showed the same sign of effects. For these conjunction tests, we only kept regions that have at least 10 contiguous voxels.

Since the number of participants was fewer in this study (n=16) than in the previous study (n=32), we might have lower power to detect effects in the whole-brain analyses. Thus, we also implemented ROI analyses. We selected seven ROIs that showed the conjunction effects of CPP, RU and reward value in the previous study (McGuire et al., 2014) and tested the effects of CPP and RU in these ROIs.

fMRI analysis: multi-voxel pattern analysis (MVPA)

We implemented MVPA to understand the neural representation of error signals and subsequent choices. Our analyses focus on the multi-voxel pattern when participants received an outcome. Before implementing MVPA, we estimated trial-by-trial beta values using the unsmoothed preprocessed fMRI data. We used the general linear model (GLM) to estimate the beta weights for each trial (Mumford, Turner, Ashby, & Poldrack, 2012). In each GLM, the first

regressor is the trial of interest and the second combines the rest of trials in the same condition. These two regressors were then convolved with a gamma hemodynamic response function. In addition, six motion parameters were included as control regressors. We repeated this process (one GLM per trial) to estimate trial-by-trial beta values for all the trials in the two conditions. We then used these beta values as observations for MVPA. A whole-brain searchlight was implemented (Kriegeskorte, Goebel, & Bandettini, 2006). In each searchlight, a sphere with the diameter of 5 voxels (10 mm) was formed, and the pattern of activity across the voxels within the sphere were used to run MVPA.

A support vector machine (SVM) with a linear kernel was used to decode different error signals and choices in our whole-brain searchlight analysis. We implemented SVM through the LIBSVM toolbox (Chang & Lin, 2011). To avoid overfitting, we used 3-fold cross-validation, with one fold used as testing data and the other two as training data. Training data were used to train the classifier and then this classifier was used on testing data to examine the classification accuracy. In linear SVM, a free parameter *c* regularizes the trade-off between decreasing training error and increasing generalization. Thus, during the training of classifier, the training data were further split into 3-folds to select the optimal value of the parameter *c* through cross-validation. We pick the optimal value for *c* from [0.001, 0.01, 0.1, 1, 10, 100, 1000] and this optimal parameter should maximize the cross-validation accuracy. Then, we used the optimal parameter *c* to train the model again based on the entire training data and calculated the classification accuracy on the testing data. We repeated this procedure with each of the three folds held out as testing data and calculated the average of the classification accuracy. To minimized the influence of different number of trials for each category on the classification accuracy, we used balanced accuracy.

We first examined how the multi-voxel neural pattern on the current trial could discriminate correct versus error on the current trial or error magnitudes on the current error trial. For the analysis of correct versus error, the baseline accuracy is 50%. For the analysis of error magnitudes, we split trials into three bins of error magnitude: 1, 2, and 3+. Thus, the baseline accuracy is 33%.

We next examined how the multi-voxel neural pattern on the current trial could discriminate whether the previous trial was an error or not. We also investigated how the classification of past errors differs conditional on the type of the current trial. We classified trial

609 *t-1* as correct or error separately for four different types of current trials: correct trials, error trials,
610 unambiguous feedback trials and ambiguous feedback trials. Unambiguous feedback trials were
611 trials with error magnitudes of 0 or 3+, while ambiguous feedback trials were trials with error
612 magnitudes of 1 or 2, in which participants would be uncertain about the change of the state in
613 the high-noise condition.
614 Lastly, we examined how the multi-voxel neural pattern on the current trial could classify
615 the choice on the next trial. In this analysis, we focused only on the trials with error magnitudes
616 of 1 or 2 in the high-noise condition, since only under these conditions were participants
617 similarly likely to switch versus stay. For these trials, we examined whether the multi-voxel
618 pattern on the current trial predicted whether the participant stayed or switched on the next trial.
619 The baseline accuracy was 50%.
620 After obtaining the classification accuracy for each participant, we subtracted the baseline
621 accuracy from the classification accuracy. Before conducting a group-level test, we smoothed
622 these individual accuracy maps with a 6 mm FWHM Gaussian kernel. For statistical testing, one-
623 sample cluster-mass permutation was applied with 5,000 iterations. We used uncorrected voxel
624 $p<0.005$ to form a cluster and estimated the corrected cluster p value for each cluster. For the
625 conjunction analyses, we used the same procedure described above.
626

Acknowledgments

This work was supported by grants from National Institute of Mental Health (R01-MH098899 to J.I.G. and J.W.K.) and National Science Foundation (1533623 to J.I.G. and J.W.K.). We thank Yin Li for valuable comments; M. Kathleen Caulfield for fMRI scanning.

Competing interests

The authors declare no competing interests.

635 **References**

636 Akrami, A., Kopec, C. D., Diamond, M. E., & Brody, C. D. (2018). Posterior parietal cortex
637 represents sensory history and mediates its effects on behaviour. *Nature, 554*, 368.
638 doi:10.1038/nature25510
639 Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the
640 value of information in an uncertain world. *Nat Neurosci, 10*(9), 1214-1221.
641 doi:10.1038/nn1954
642 Blanchard, T. C., & Gershman, S. J. (2018). Pure correlates of exploration and exploitation in the
643 human brain. *Cognitive, Affective, & Behavioral Neuroscience, 18*(1), 117-126.
644 doi:10.3758/s13415-017-0556-2
645 Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator.
646 *Current Opinion in Neurobiology, 37*, 149-157. doi:10.1016/j.conb.2016.01.003
647 Chan, S. C. Y., Niv, Y., & Norman, K. A. (2016). A Probability Distribution over Latent Causes,
648 in the Orbitofrontal Cortex. *The Journal of Neuroscience, 36*(30), 7817-7828.
649 doi:10.1523/jneurosci.0659-16.2016
650 Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM*
651 *transactions on intelligent systems and technology, 2*(3), 27.
652 doi:10.1145/1961189.1961199
653 d'Acremont, M., & Bossaerts, P. (2016). Neural Mechanisms Behind Identification of
654 Leptokurtic Noise and Adaptive Behavioral Response. *Cerebral Cortex, 26*(4), 1818-
655 1830. doi:10.1093/cercor/bhw013
656 Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates
657 for exploratory decisions in humans. *Nature, 441*(7095), 876-879.
658 doi:10.1038/nature04766
659 Fischer, Adrian G., & Ullsperger, M. (2013). Real and Fictive Outcomes Are Processed
660 Differently but Converge on a Common Adaptive Mechanism. *Neuron, 79*(6), 1243-
661 1255. doi:10.1016/j.neuron.2013.07.006
662 FitzGerald, T. H. B., Moran, R. J., Friston, K. J., & Dolan, R. J. (2015). Precision and neuronal
663 dynamics in the human posterior parietal cortex during evidence accumulation.
664 *NeuroImage, 107*, 219-228. doi:10.1016/j.neuroimage.2014.12.015
665 Furl, N., & Averbeck, B. B. (2011). Parietal Cortex and Insula Relate to Evidence Seeking
666 Relevant to Reward-Related Decisions. *The Journal of Neuroscience, 31*(48), 17572-
667 17582. doi:10.1523/jneurosci.4236-11.2011
668 Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus Rewards: Dissociable
669 Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement
670 Learning. *Neuron, 66*(4), 585-595. doi:10.1016/j.neuron.2010.04.016
671 Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in
672 unpredictable environments. *eLife, 4*, e08825. doi:10.7554/eLife.08825
673 Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-
674 based registration. *NeuroImage, 48*(1), 63-72. doi:10.1016/j.neuroimage.2009.06.060
675 Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The Role of the Ventromedial
676 Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans.
677 *The Journal of Neuroscience, 26*(32), 8360-8367. doi:10.1523/jneurosci.1010-06.2006
678 Hampton, A. N., & O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related
679 decision making with functional MRI. *Proceedings of the National Academy of Sciences,*
680 *104*(4), 1377-1382. doi:10.1073/pnas.0606297104

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature, 520*, 220. doi:10.1038/nature14066

Hayden, B. Y., Nair, A. C., McCoy, A. N., & Platt, M. L. (2008). Posterior Cingulate Cortex Mediates Outcome-Contingent Allocation of Behavior. *Neuron, 60*(1), 19-25. doi:10.1016/j.neuron.2008.09.012

Hunt, L. T., Malalasekera, W. M. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E. J., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience, 21*(10), 1471-1481. doi:10.1038/s41593-018-0239-5

Hwang, E. J., Dahlen, J. E., Mukundan, M., & Komiyama, T. (2017). History-based action selection bias in posterior parietal cortex. *Nature Communications, 8*(1), 1242. doi:10.1038/s41467-017-01356-z

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage, 17*(2), 825-841. doi:10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782-790. doi:10.1016/j.neuroimage.2011.09.015

Karlsson, M. P., Tervo, D. G. R., & Karpova, A. Y. (2012). Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty. *Science, 338*(6103), 135-139. doi:10.1126/science.1226518

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science, 336*(6077), 95-98. doi:10.1126/science.1216930

Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience, 19*, 1280. doi:10.1038/nn.4382

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868. doi:10.1073/pnas.0600244103

Li, Y. S., Nassar, M. R., Kable, J. W., & Gold, J. I. (2019). Individual neurons in the cingulate cortex encode action monitoring, not selection, during adaptive decision-making. *The Journal of Neuroscience*, 0159-0119. doi:10.1523/jneurosci.0159-19.2019

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron, 84*(4), 870-881. doi:10.1016/j.neuron.2014.10.013

Muller, T. H., Mars, R. B., Behrens, T. E., & O'Reilly, J. X. (2019). Control of entropy in neural models of environmental state. *eLife, 8*, e39404. doi:10.7554/eLife.39404

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage, 59*(3), 2636-2643. doi:10.1016/j.neuroimage.2011.08.076

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *eLife, 8*, e46975. doi:10.7554/eLife.46975

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable Forms of Uncertainty-Driven Representational Change Across the Human Brain. *The Journal of Neuroscience, 39*(9), 1688-1698. doi:10.1523/jneurosci.1713-18.2018

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience, 15*(7), 1040-1046. doi:10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience, 30*(37), 12366-12378. doi:10.1523/JNEUROSCI.0822-10.2010

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences, 110*(38), E3660-E3669. doi:10.1073/pnas.1305373110

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, John P. (2013). The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. *Neuron, 79*(1), 191-201. doi:10.1016/j.neuron.2013.04.037

Purcell, B. A., & Kiani, R. (2016). Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proceedings of the National Academy of Sciences, 113*(31), E4531-E4540. doi:10.1073/pnas.1524685113

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron, 91*(6), 1402-1412. doi:10.1016/j.neuron.2016.08.019

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., . . . Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23 (S1)*, S208-S219. doi:10.1016/j.neuroimage.2004.07.051

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron, 81*(2), 267-279. doi:10.1016/j.neuron.2013.11.005

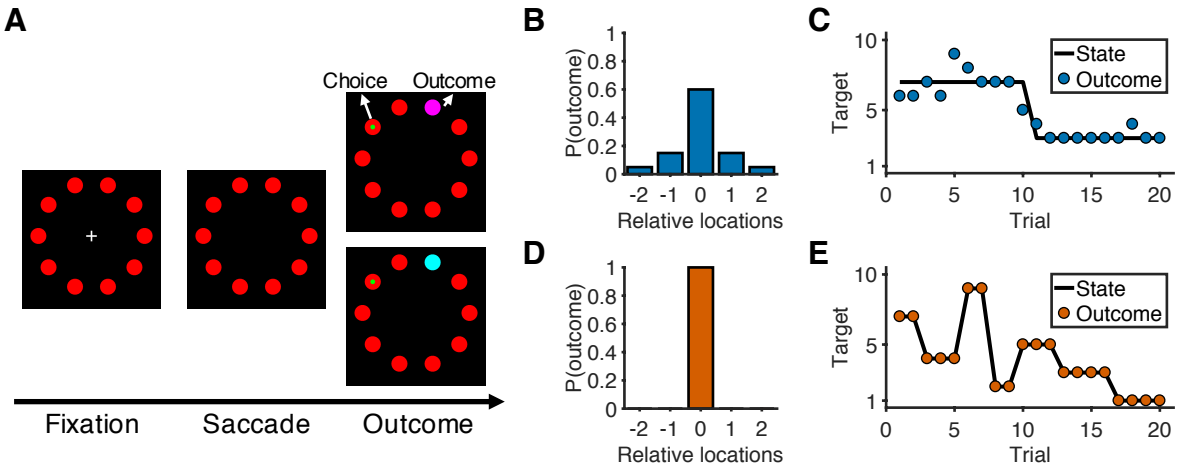


Figure 1

Overview of task and experimental design. (**A**) Sequence of the task. At the start of the trial, participants look at a cross in the center of the screen and maintain fixation for 0.5 sec to initialize the trial. After the cross disappears, participants choose one of 10 targets (red) by looking at it within 1.5 sec and then holding fixation on the chosen target for 0.3 sec. During the outcome phase (1 sec), a green dot inside the target indicates the participants' choice. The rewarded target is shown in purple or cyan to indicate the number of earnable points as 10 or 20, respectively. (**B**) Probability distribution of the rewarded target location in the high-noise condition. Target location is relative to the location of the state (generative mean). The rewarded target probabilities for the relative locations of [-2, -1, 0, 1, 2] are [0.05, 0.15, 0.6, 0.15, 0.05]. (**C**) Example of trials in the high-noise condition. The states change occasionally with a hazard rate of 0.02. (**D**) Probability distribution of the rewarded target location in the unstable condition. Because there is no noise in this condition, the rewarded target is always at the location of the state. (**E**) Example of trials in the unstable condition. The states change frequently with a hazard rate of 0.35.

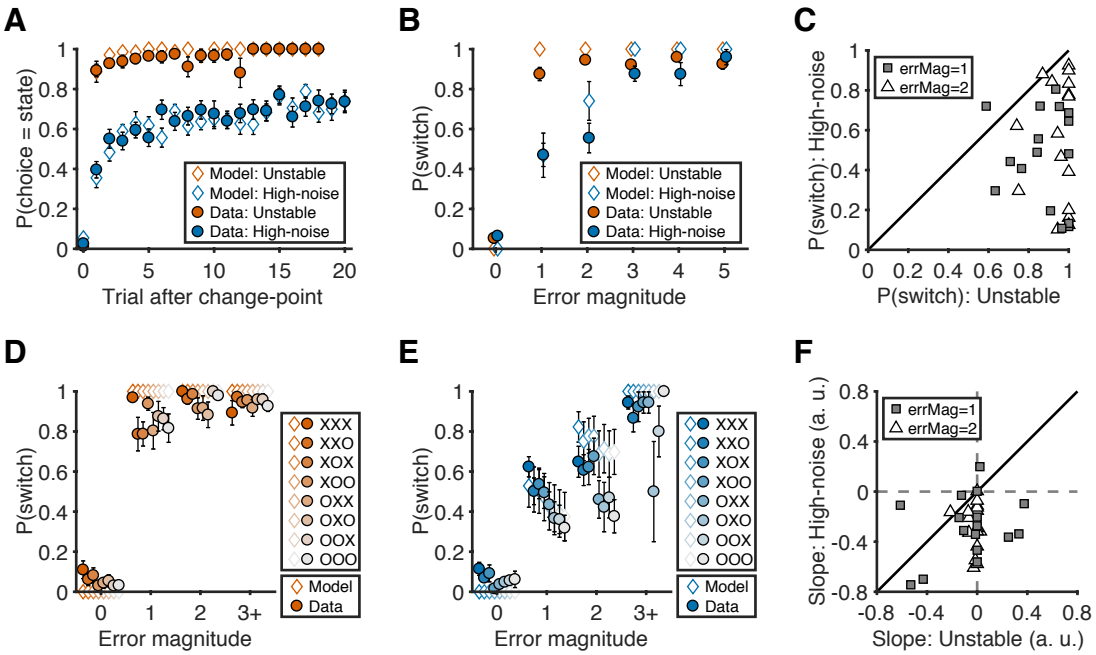


Figure 2

Behavioral results. (**A**) Probability of choosing the best target after change-points. Symbols and error bars are mean±SEM across subjects (solid symbols) or simulations (open symbols). (**B**) Relationship between error magnitude and switch probability. Symbols and error bars are as in **A**. (**C**) The distribution of switch probabilities for small errors (magnitude of 1 or 2) in both conditions. Each data point represents one participant. (**D**) Probability of switch as a function of current error magnitude and error history in the unstable condition. Different colors represent different error histories for the past 3 trials. A correct trial is marked as O, and an error trial is marked as X. For example, XOO implies that trial t-1 was an error trial, and trial t-2 and trial t-3 were correct trials. Symbols and error bars are mean±SEM across subjects. (**E**) Probability of switch as a function of current error magnitude and error history in the high-noise condition. Symbols and error bars are as in **D**. (**F**) The distribution of the slopes of switch probability against error history for small errors (magnitude of 1 or 2) in both conditions. Each data point represents one participant.

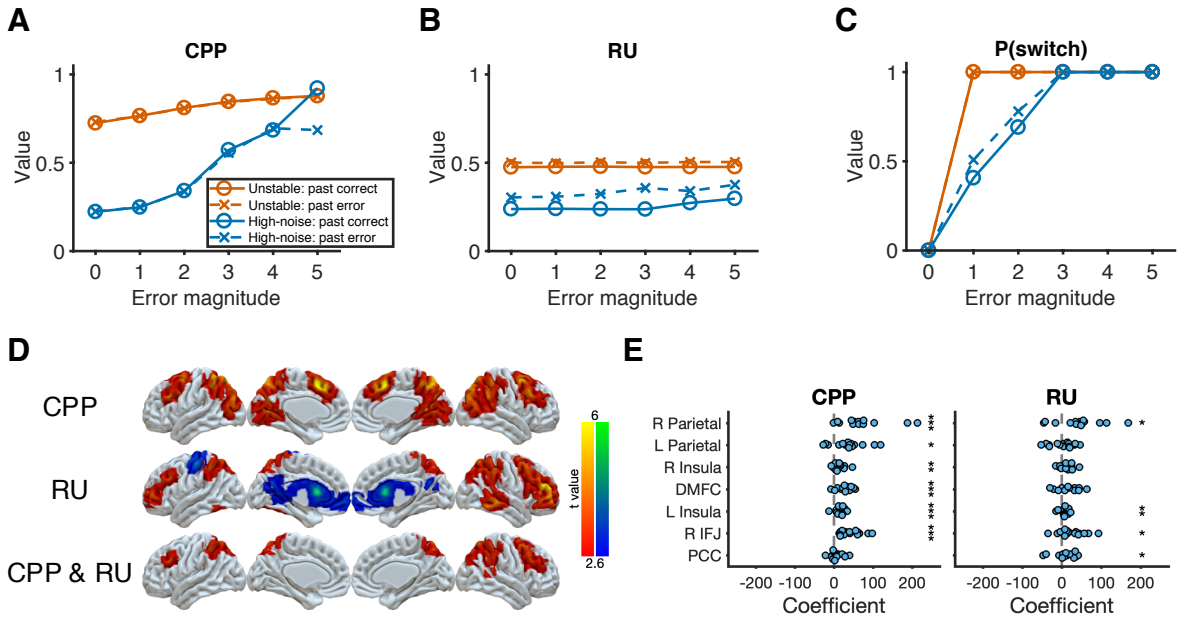


Figure 3

Reduced Bayesian model applied to behavioral and imaging data. (**A**) Model prediction for CPP. We calculated CPP from the reduced Bayesian model using subjective estimates of hazard rate and noise for each condition. The value of CPP increases as the current error magnitude increases in both conditions, but with a stronger dependence on the outcome of the previous trial in the high-noise condition. (**B**) Model prediction for RU. We calculated RU from the reduced Bayesian model using subjective estimates of hazard rate and noise for each condition. The value of RU is minimally affected by the current error magnitude. Instead, a past error tends to increase RU in the high-noise, but not the unstable, condition. (**C**) Model prediction for probability of switching choices. Increasing CPP causes the probability of switching to increase more steeply as the current error magnitude increases in the unstable condition versus in the high-noise condition. For small errors (error magnitude of 1 and 2) in the high-noise condition, the probability of switching is further influence by RU, which is affected by past errors. (**D**) Neural representation of CPP and RU. (**E**) ROI analysis for CPP and RU. These ROIs were selected based on the common regions of CPP, RU, and reward effects in McGuire et al. (2014). Significance was tested by a sign test. *p<0.05, **p<0.01, ***p<0.001.

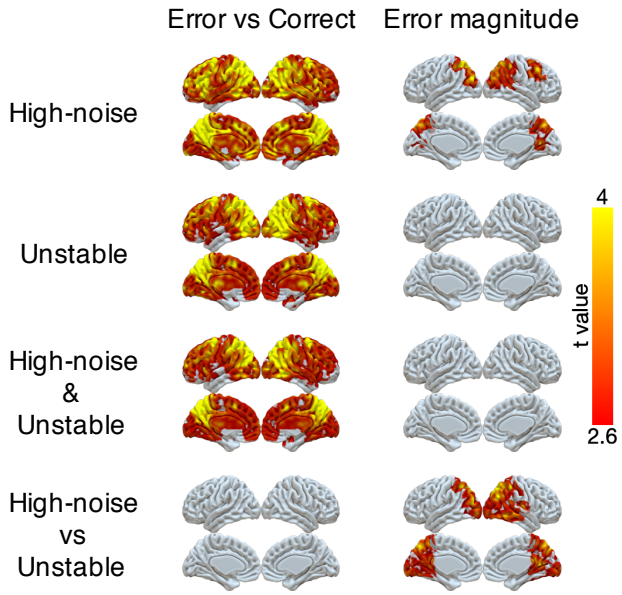


Figure 4

Representations of error and error magnitude. For error versus correct analyses, multi-voxel neural patterns were used to classify whether the response on the current trial was correct or an error. For error magnitude analyses, multi-voxel neural patterns were used to classify different error magnitudes (1, 2, 3+) conditional on the current trial being an error. Accuracies were calculated and compared with the baseline accuracy within each subject and then tested at the group level. The representation of current error magnitude is stronger in parietal cortex in the high-noise condition than the unstable condition.

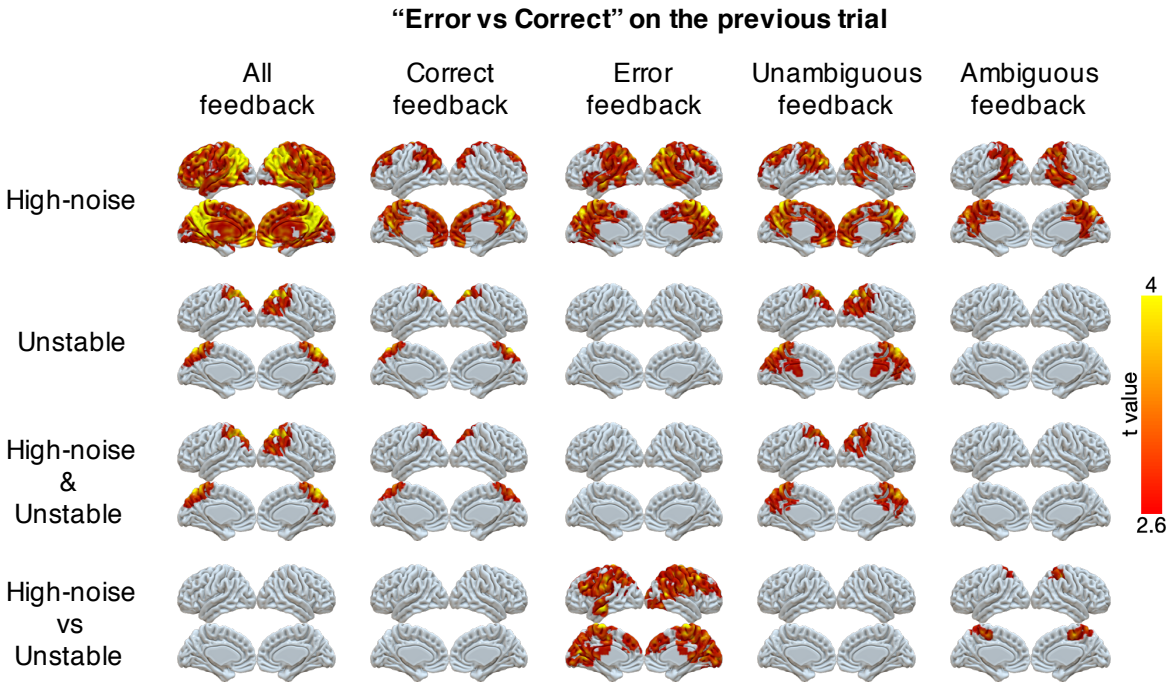
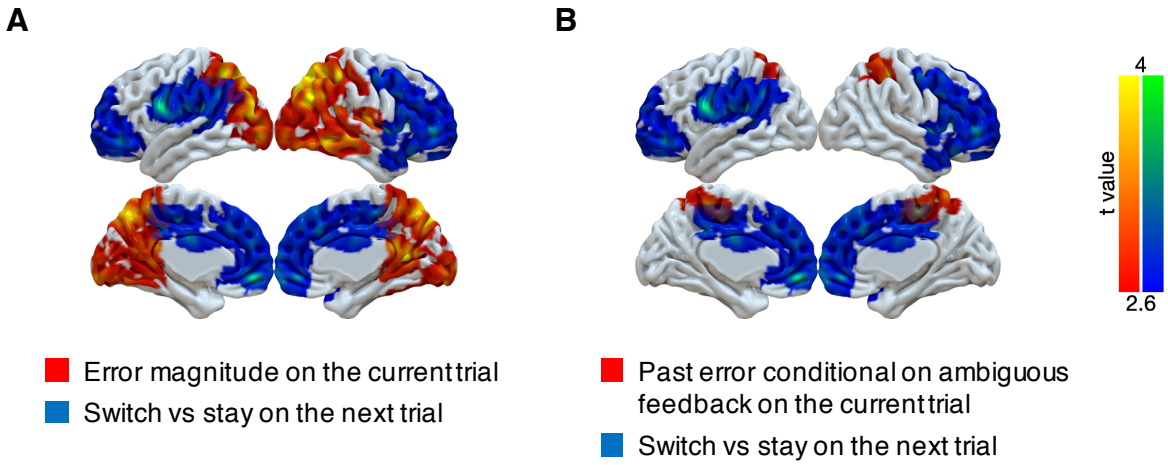


Figure 5

Representations of errors on the previous trial conditional on different types of current trials (columns). Multi-voxel neural patterns were used to classify correct responses versus errors on the previous trial. This analysis was repeated for different types of current trials: all feedback, correct feedback, error feedback, unambiguous feedback (error magnitudes are 0/3+), and ambiguous feedback (error magnitudes are 1/2). The representation of past errors is stronger in parietal cortex in the high-noise condition than the unstable condition when the current trial is an error or provides ambiguous feedback.

823



824

825 **Figure 6**

826 Representations of subsequent behavioral choices (switch versus stay) after ambiguous small errors in the
827 high-noise condition. (**A**) Overlap of results for switch versus stay on the next trial and error magnitude
828 on the current trial. Multi-voxel neural patterns were used to classify whether participants switch their
829 choice to another target or stay on the same target on the next trial. We focused on the most ambiguous
830 errors (error magnitude of 1 or 2 in the high-noise condition). Above-chance classification performance
831 was found throughout much of the frontal lobe. (**B**) Overlap of results for switch versus stay on the next
832 trial and past error conditional on ambiguous feedback on the current trial.

833

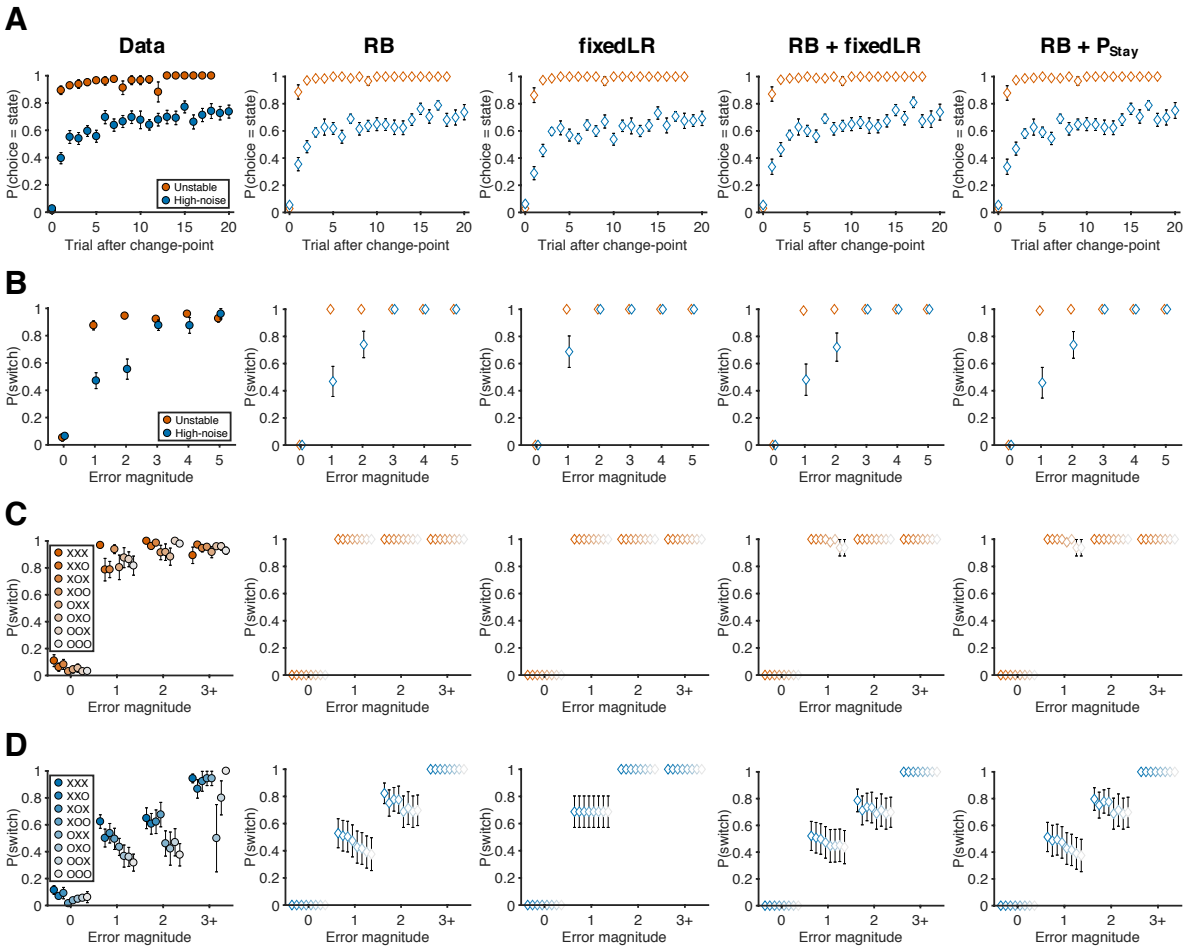


Figure 2 - Figure supplement 1

Behavioral data and predictions from different models. (**A**) Probability of choosing the best target after change-points. (RB: reduced Bayesian; fixedLR: fixed learning rate; P_{stay}: fixed tendency to stay) (**B**) The relationship between error magnitude and switch probability. (**C**) Probability of switch as a function of current error magnitude and error history in the unstable condition. (**D**) Probability of switch as a function of current error magnitude and error history in the high-noise condition. Symbols and colors are as in Figure 2.

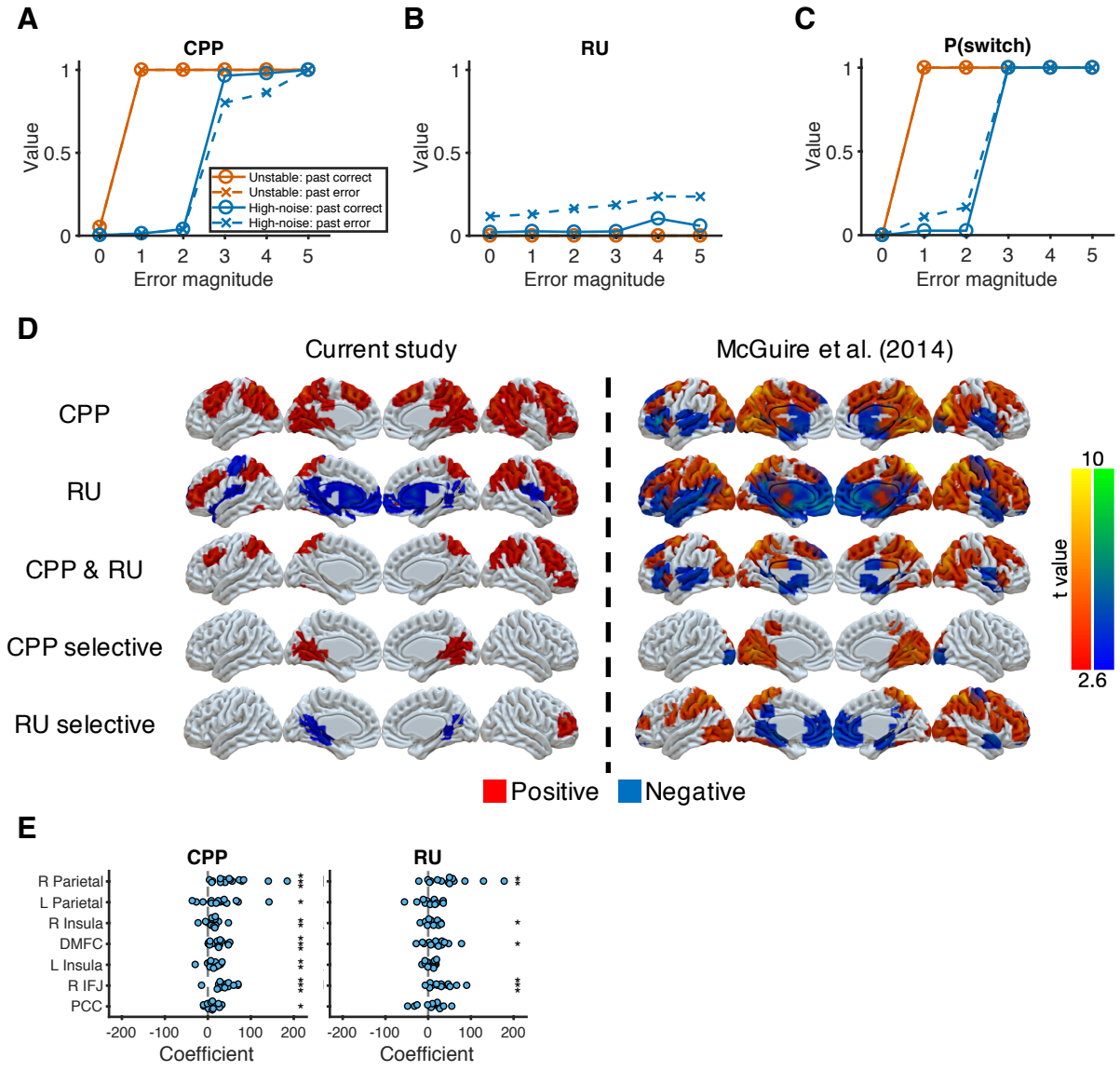


Figure 3 - Figure supplement 1

Reduced Bayesian model with true hazard rate and noise applied to behavioral and imaging data. (**A**) Model prediction for CPP. We calculated CPP from the reduced Bayesian model using the true hazard rate and noise for each condition. (**B**) Model prediction for RU. We calculated RU from the reduced Bayesian model using the true hazard rate and noise for each condition. (**C**) Model prediction for probability of switch. Probability of switch is influenced by both CPP and RU. (**D**) Neural representation of CPP and RU in the current study and in McGuire et al. (2014). CPP selective effect represents the conjunction of CPP>0 and CPP>RU. RU selective effect represents the conjunction of RU>0 and RU>CPP. (**E**) ROI analysis for CPP and RU. These ROIs were selected based on the common regions of CPP, RU and reward effects in McGuire et al. (2014). Significance was tested by a sign test. *p<0.05, **p<0.01, ***p<0.001.

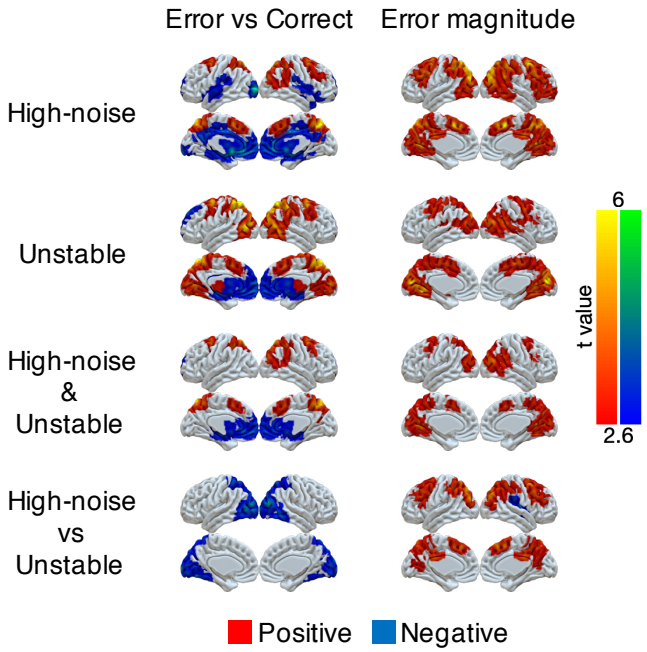


Figure 4 - Figure supplement 1

Univariate representations of error and error magnitude. A GLM was implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). The trial-by-trial regressors of interest that were included in the GLM were: onset of correct trials, earnable value on correct trials, onset of error trials, error magnitude on error trials, switch or stay on error trials and earnable value on error trials. We focused on the effects of error (which is the difference between the onset of error trials and the onset of correct trials) and error magnitude. Group *t*-values are shown. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel $p<0.005$.

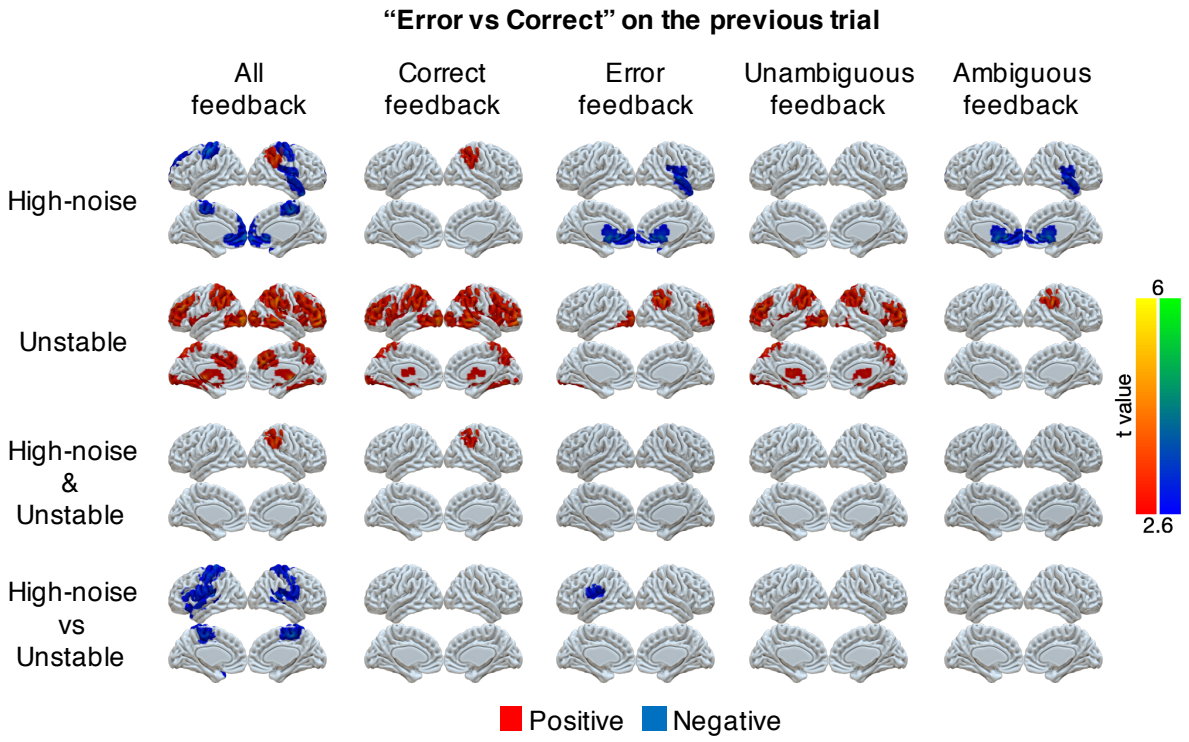


Figure 5 - Figure supplement 1

Univariate representations of error on the previous trial conditional on different types of current trials (columns). Several GLMs were implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). First, we examined errors on the previous trial across all trials. The trial-by-trial regressors of interest that were included in the GLM were: onset of trials, error on trial t, error on trial t-1, error on trial t-2, and error on trial t-3. We focused on the effect of error on trial t-1. Second, we separated the analysis of past errors conditional on the current trial being correct or an error. The trial-by-trial regressors of interest that were included in the GLM were: onset of current correct trials, errors on trial t-1, t-2, or t-3 conditional on the current trial being correct, onset of current error trials, errors on trial t-1, t-2, or t-3 conditional on the current trial being an error. We focused on the effects of error on trial t-1 conditional on the current trial being correct or an error. Third, we separated errors conditional on unambiguous (error magnitudes are 0/3+) or ambiguous feedback (error magnitudes are 1/2). The trial-by-trial regressors of interest that were included in the GLM were: onset of current trials with unambiguous feedback, errors on trial t-1, t-2 or t-3 conditional on the current trial providing unambiguous feedback, onset of current trials with ambiguous feedback, errors on trial t-1, t-2 or t-3 conditional on the current trial providing ambiguous feedback. We focused on the effects of errors on trial t-1 conditional on the current trials providing unambiguous or ambiguous feedback. Group *t*-values are shown. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel $p<0.005$.

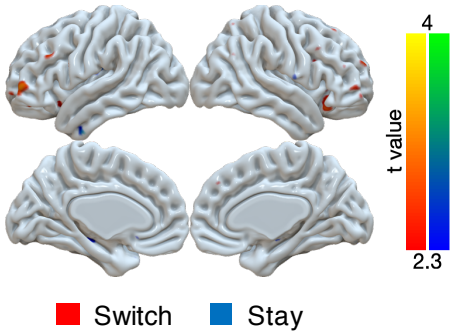


Figure 6 - Figure supplement 1

Univariate GLM for switch versus stay on ambiguous feedback in the high-noise condition. A GLM was implemented with several trial-by-trial regressors of interest: onset of trials with error magnitude of 0, onset of trials with error magnitude of 3+, onset of ambiguous feedback (error magnitudes are 1/2) followed by switching, onset of ambiguous feedback followed by staying. We tested the effects of the difference between switch and stay after ambiguous feedback. Group *t*-values are shown. The results were thresholded based on uncorrected voxel $p<0.01$.

Table 1

BIC of behavior models

Model	Condition	BIC improvement by RB model
Reduced Bayesian model (RB)	Unstable	
	High-noise	
Fixed learning rate model (fixedLR)	Unstable	5.26 [3.97, 5.71]**
	High-noise	-22.22 [-77.36, 0.03]†
RB + fixedLR	Unstable	-9.85 [-10.97, -7.65]***
	High-noise	-4.60 [-9.93, 0.98]
RB + P_{stay}	Unstable	-5.20 [-5.66, -3.76]**
	High-noise	-5.57 [-5.68, -2.60]**

The values were shown as median [IQR]. A negative value means that RB model performed better than the compared model. Significance was tested by a sign test. $^{\dagger}p<0.08$, $^{**}p<0.01$, $^{***}p<0.001$.

Table 2

Parameters of behavior models

Model	Parameter	Unstable	High-noise	Unstable > High-noise
RB	H	0.88 [0.68, 0.92]	0.33 [0.11, 0.50]	0.46 [0.25, 0.72]***
	K	0.59 [0.03, 2.22]	1.86 [1.22, 2.32]	-0.35 [-1.98, 0.71]
fixedLR	α_{fixed}	0.96 [0.86, 0.97]	0.63 [0.37, 0.73]	0.33 [0.19, 0.49]***
RB + fixedLR	H	0.27 [0.00, 0.84]	0.02 [0.00, 0.13]	0.14 [0.00, 0.64]†
	K	7.76 [2.18, 9.99]	3.83 [2.36, 9.02]	0.26 [-0.68, 5.77]
	α_{fixed}	0.97 [0.75, 1.00]	0.93 [0.23, 1.00]	0.02 [-0.10, 0.56]
	w	0.41 [0.15, 0.79]	0.65 [0.40, 0.87]	-0.23 [-0.57, 0.24]
RB + P_{stay}	H	0.86 [0.73, 0.94]	0.35 [0.11, 0.60]	0.47 [0.18, 0.72]**
	K	4.54 [0.12, 9.99]	2.09 [1.51, 4.71]	0.02 [-1.33, 6.41]
	P_{stay}	0.00 [0.00, 0.05]	0.01 [0.00, 0.13]	0.00 [-0.11, 0.02]

Parameter values were shown as median [IQR]. Difference of parameter values between the two conditions was tested by a sign test. $^{\dagger}p<0.08$, $^{**}p<0.01$, $^{***}p<0.001$.

914 Table 3

915 Summary of fMRI results: error magnitude and past error

Cluster index	#Voxels	Region	Peak *t*	Peak *x*	Peak *y*	Peak *z*
Error magnitude: high-noise versus unstable						
1	21032	R Precuneus	5.22	16	-56	12
		R Angular gyrus	5.17	44	-70	32
		L Precuneus	5.08	-18	-58	20
		Occipital pole	5.07	2	-98	-2
		L Superior parietal lobule	4.91	-10	-66	48
		R Occipital cortex	4.69	26	-76	18
		L Occipital cortex	4.54	-38	-86	26
		R Superior parietal lobule	4.44	44	-44	54
		Posterior cingulate cortex	4.43	2	-46	20
Past error on ambiguous feedback: high-noise versus unstable						
1	1881	Posterior cingulate cortex	4.79	12	-24	52
		R Superior parietal lobule	4.04	32	-38	54
		R Precuneus	3.58	6	-54	70
		L Superior parietal lobule	3.54	-16	-54	62
Conjunction: Error magnitude & Past error on ambiguous feedback						
1	304	R Superior parietal lobule	3.41	38	-40	52
2	103	R Precuneus	3.02	2	-58	70
3	81	L Superior parietal lobule	3.23	-18	-56	72

916

Table 4

Summary of fMRI results: behavior change

Cluster index	#Voxels	Region	Peak *t*	Peak *x*	Peak *y*	Peak *z*
Switch versus stay on ambiguous feedback in the high-noise condition						
1	12042	Middle cingulate cortex	4.35	14	-8	30
		R Insula	4.33	38	4	2
		Medial orbitofrontal cortex	4.24	-4	50	-10
		R Frontal pole	4.11	40	46	0
		R Inferior frontal gyrus	4.11	48	26	10
		L Frontal pole	4.01	-24	52	-2
		Dorsomedial frontal cortex	3.96	0	26	34
		Posterior cingulate cortex	3.93	2	-28	50
		R Precentral gyrus	3.91	48	-6	50
		Anterior cingulate cortex	3.51	0	48	20
2	3134	L Precentral gyrus	4.43	-62	2	24
		L Superior temporal gyrus	4.28	-50	-32	12
		L Postcentral gyrus	3.61	-50	-26	44