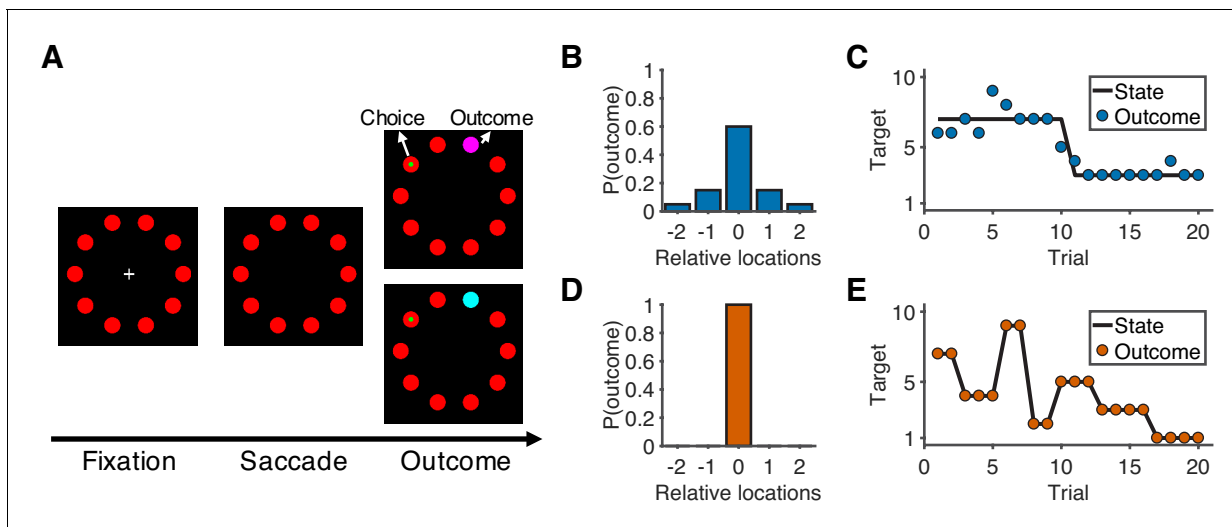


---

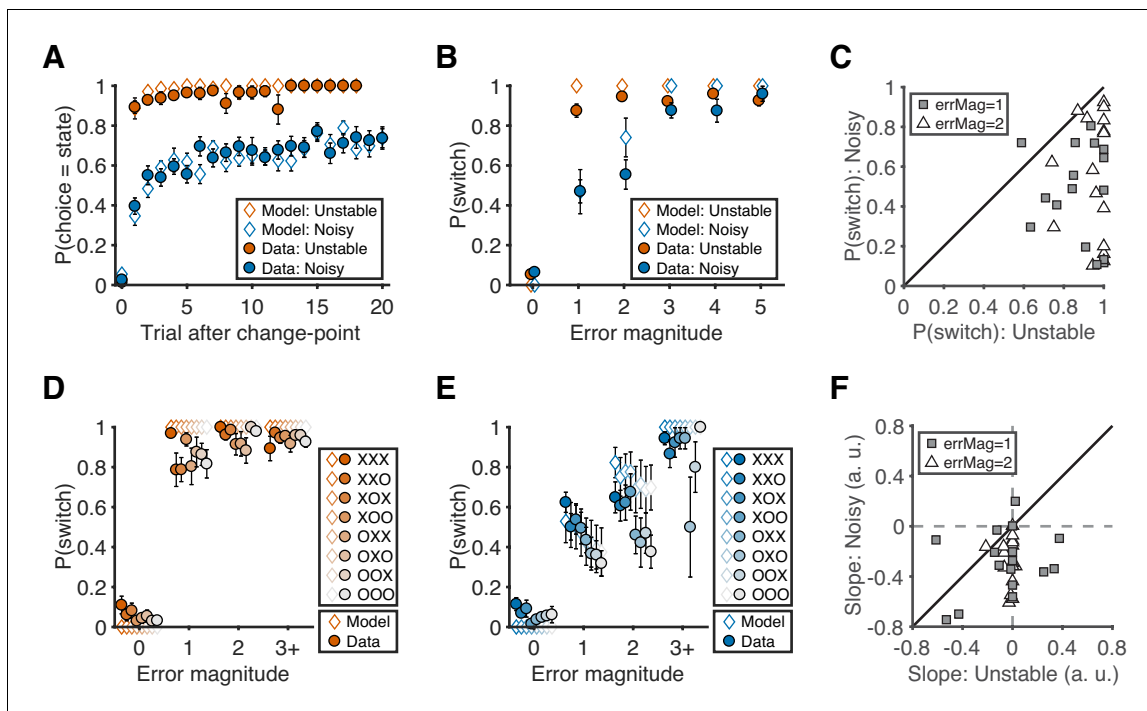
## Figures and figure supplements

Neural encoding of task-dependent errors during adaptive learning

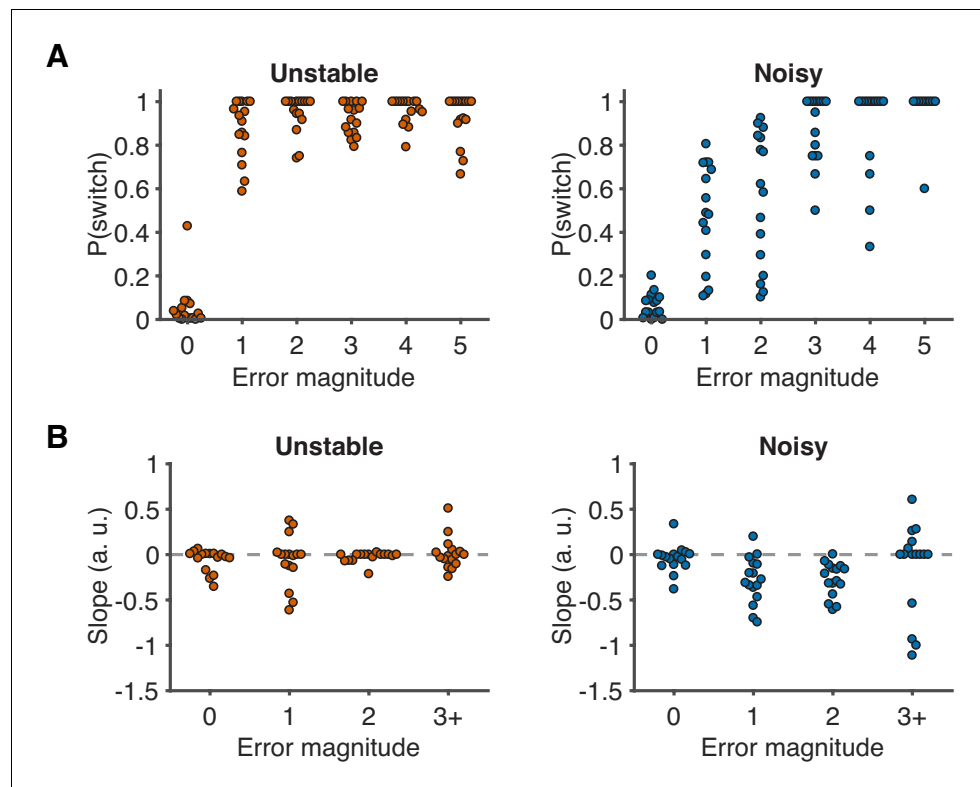
**Chang-Hao Kao et al**



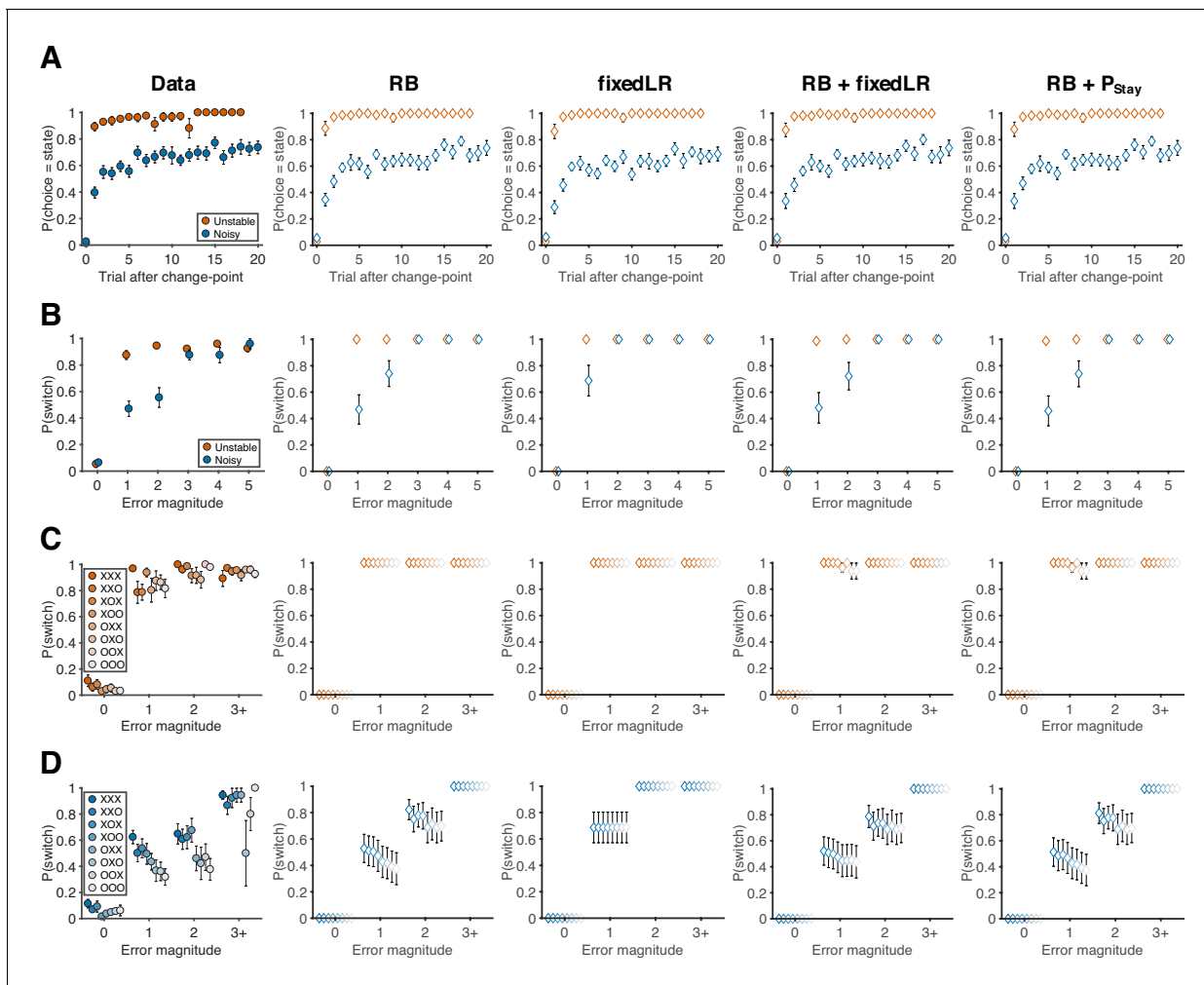
**Figure 1.** Overview of task and experimental design. **(A)** Sequence of the task. At the start of the trial, participants look at a cross in the center of the screen and maintain fixation for 0.5 s to initialize the trial. After the cross disappears, participants choose one of 10 targets (red) by looking at it within 1.5 s and then holding fixation on the chosen target for 0.3 s. During the outcome phase (1 s), 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 noisy 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 noisy 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.



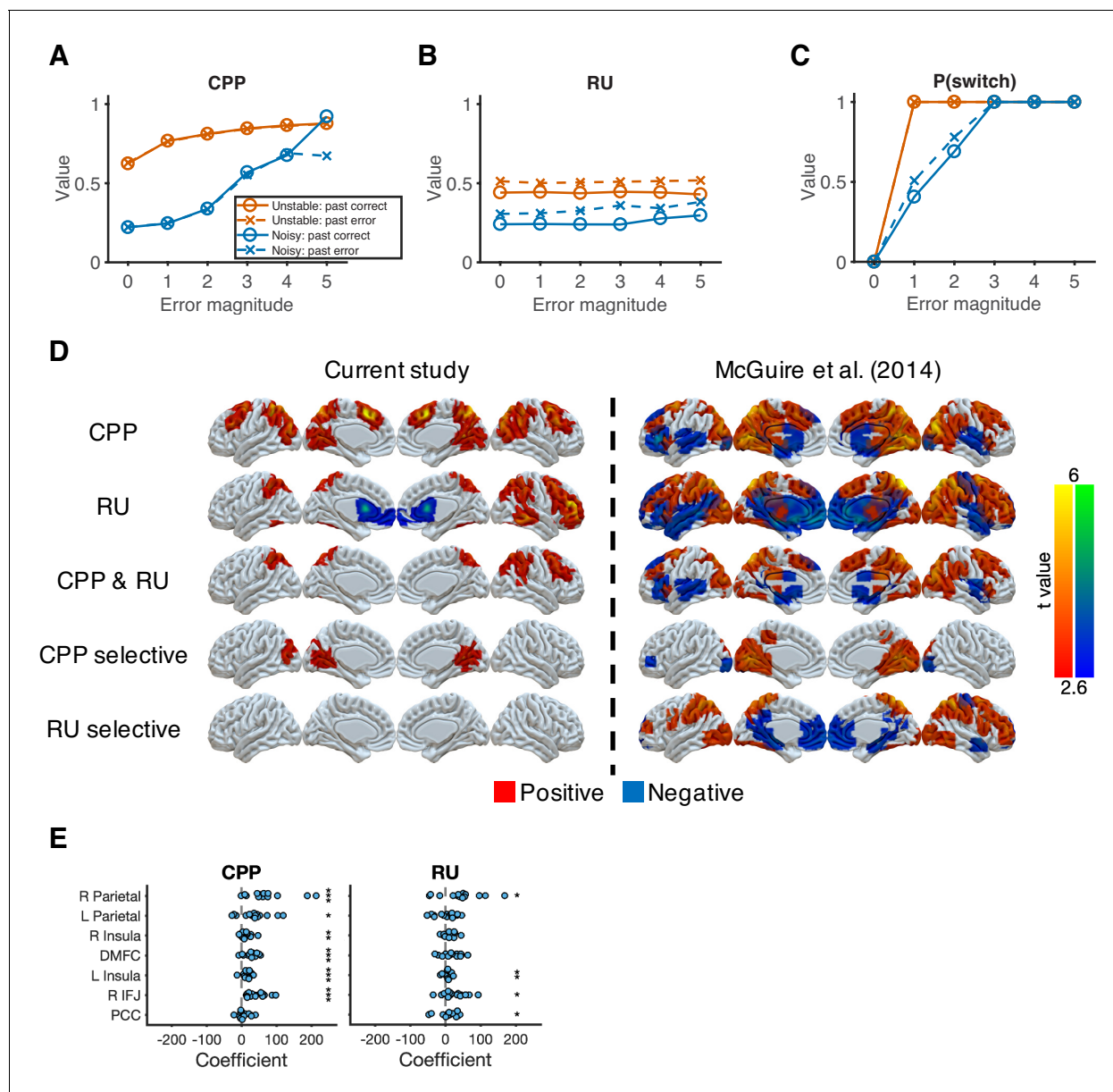
**Figure 2.** Behavioral results. (A) Probability of choosing the best target after change-points. Symbols and error bars are mean  $\pm$  SEM across subjects (solid symbols) or simulations (open symbols). (B) Relationship between error magnitude and switch probability. Symbols and error bars are as 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. Distributions for all error magnitudes are shown in **Figure 2—figure supplement 1**. (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 three 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  $\pm$  SEM across subjects. (E) Probability of switch as a function of current error magnitude and error history in the noisy 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. Distributions for all error magnitudes are shown in **Figure 2—figure supplement 1**.



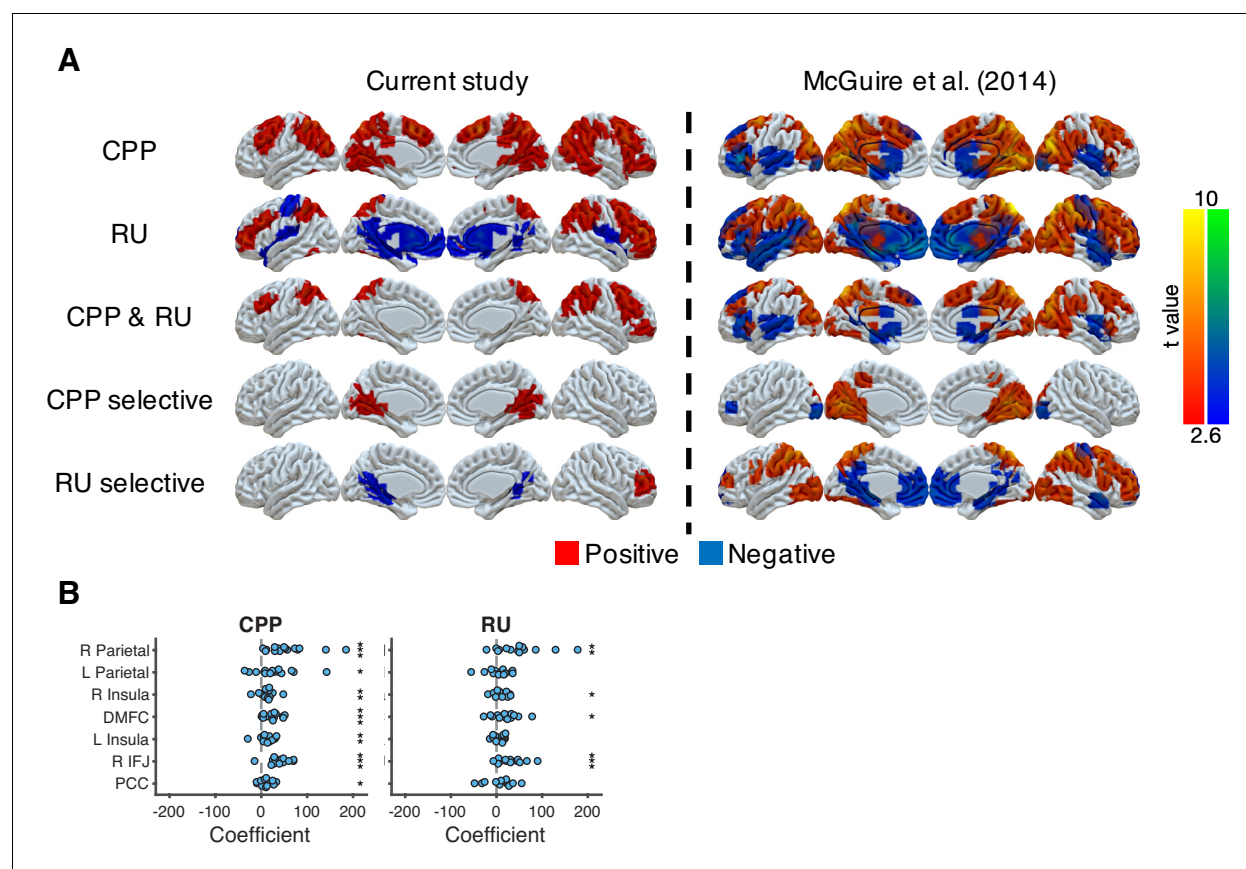
**Figure 2—figure supplement 1.** Distributions of behavior as a function of error magnitude. (A) Distributions of switch probability as a function of error magnitude. Each data point represents one participant. (B) Distributions of slopes of switch probability against error history as a function of error magnitude. Each data point represents one participant.



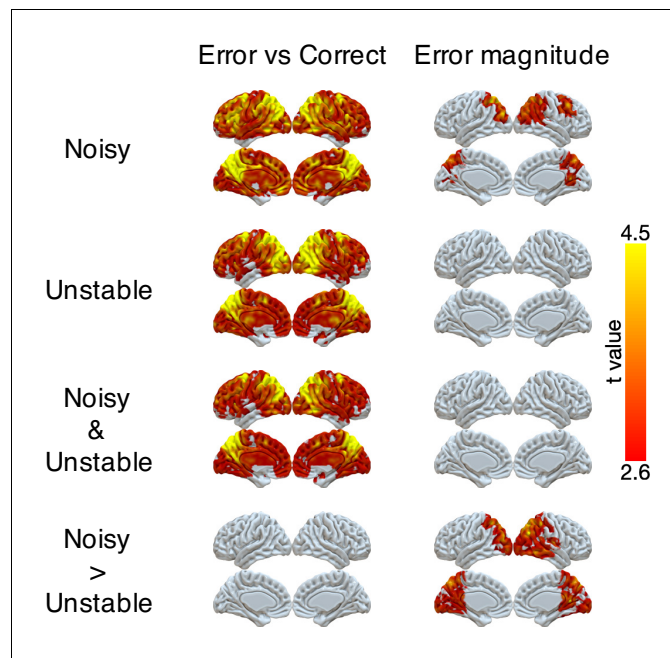
**Figure 2—figure supplement 2.** Behavioral data and predictions from different models. (A) Probability of choosing the best target after change-points. (RB: reduced Bayesian; fixedLR: fixed learning rate;  $P_{\text{stay}}$ : fixed tendency to stay) (B) The relationship between error magnitude and 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 noisy condition. Symbols and colors are as in **Figure 2**.



**Figure 2—figure supplement 3.** Reduced Bayesian model applied to behavioral and imaging data. (A) Model prediction for CPP. We calculated CPP from the fitted reduced Bayesian model, which incorporates 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 noisy condition. (B) Model prediction for RU. We calculated RU from the fitted reduced Bayesian model, which incorporates 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. (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 noisy condition. For small errors (error magnitude of 1 and 2) in the noisy condition, the probability of switching is further influenced by RU, which is affected by past errors. (D) Neural representation of CPP and RU. CPP selective effect represents the conjunction of  $CPP > 0$  and  $CPP > RU$ . RU selective effect represents the conjunction of  $RU > 0$  and  $RU > CPP$ . The results were thresholded based on uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ). (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. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

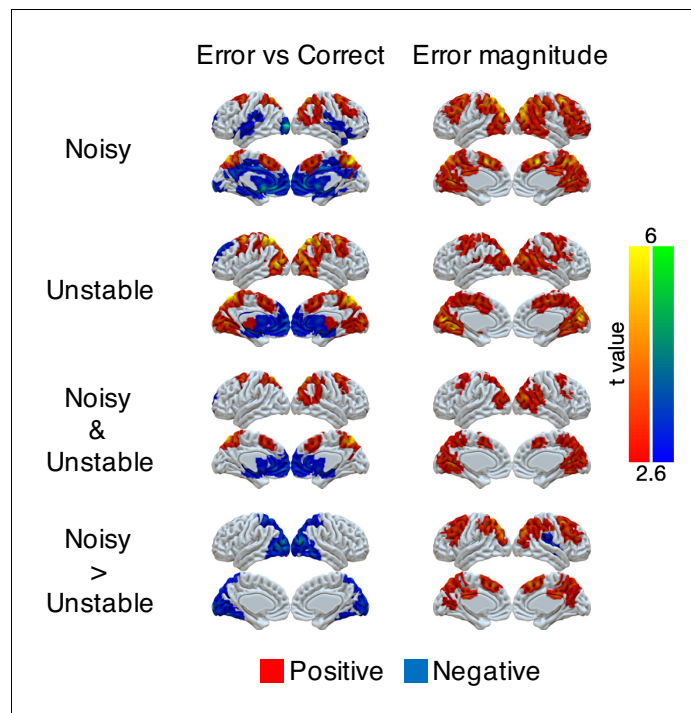


**Figure 2—figure supplement 4.** Neural representations of CPP and RU from the approximately ideal observer, which is the reduced Bayesian model with true hazard rate and noise, for direct comparison to analyses in *McGuire et al., 2014*, which used covariates constructed from the ideal rather than the fitted model. **(A)** 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  $RU > CPP$ . The results were thresholded based on uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ). **(B)** 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. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

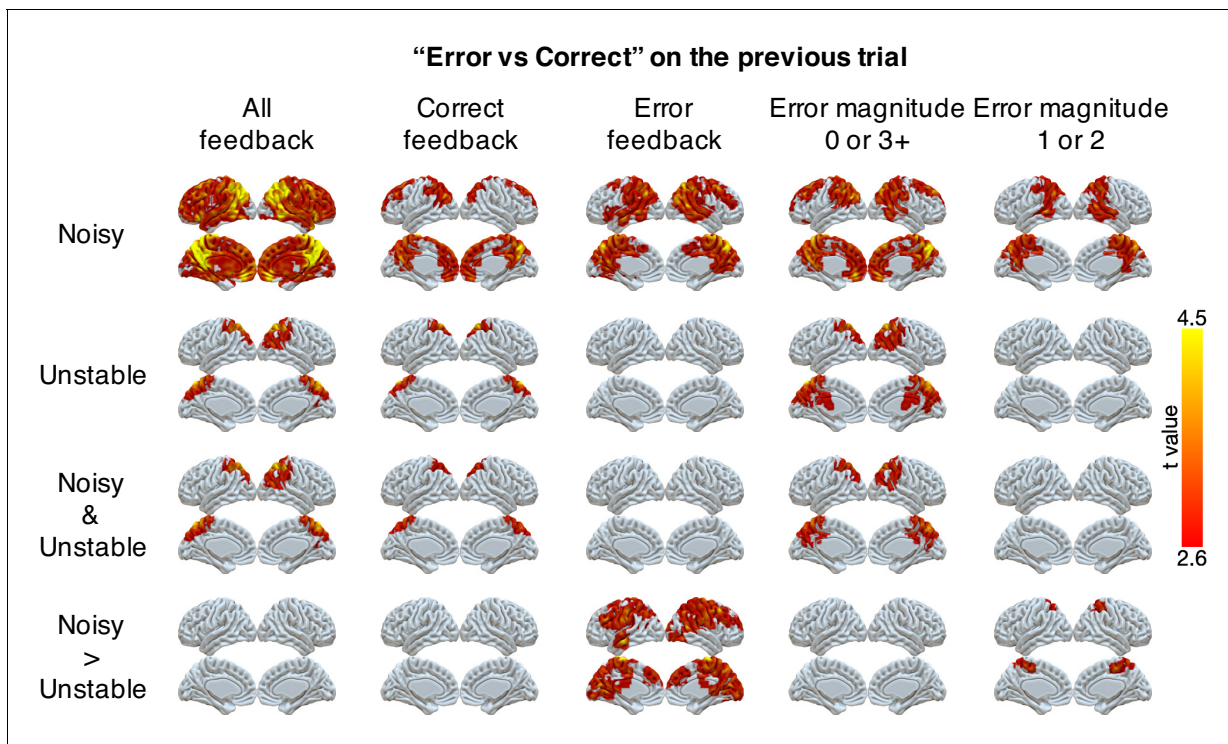


**Figure 3.** Representations of error and error magnitude. For error versus correct analyses, multi-voxel neural patterns were used to classify whether the response on the current trial was correct or an error. For error magnitude analyses, multi-voxel neural patterns were used to classify different 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 noisy condition than the unstable condition. The cluster-forming threshold was an uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ), with cluster mass corrected for multiple comparisons using non-parametric permutation tests.

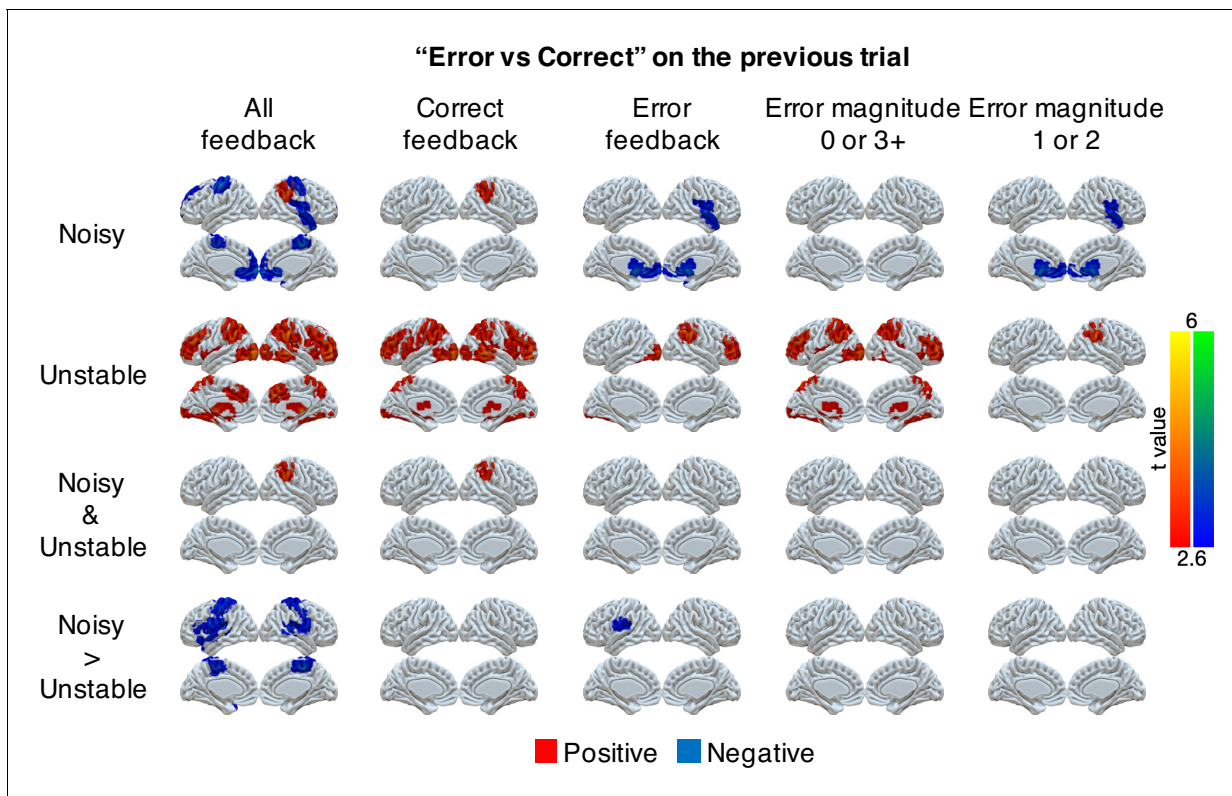




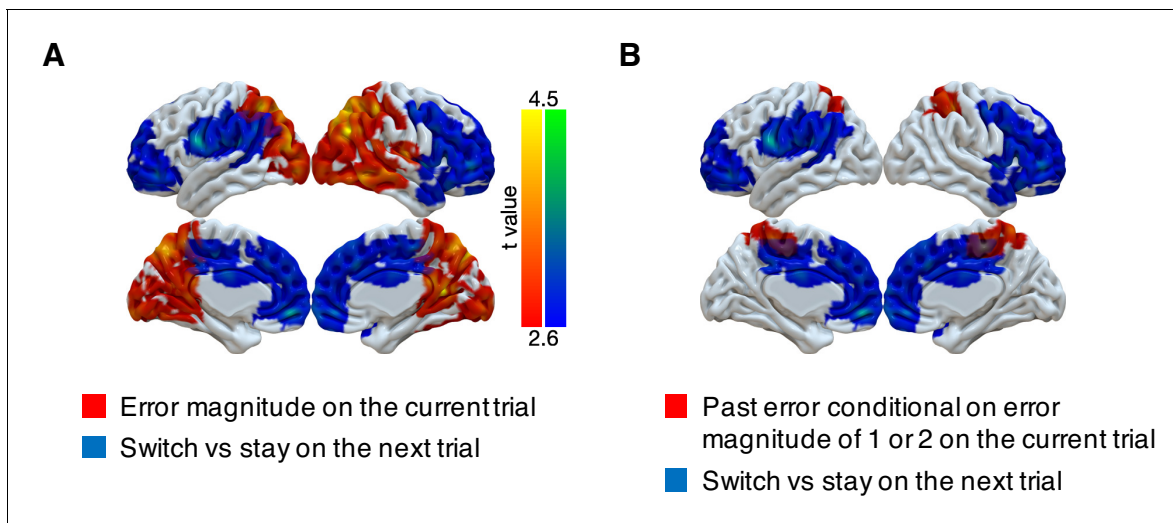
**Figure 3—figure supplement 1.** Univariate representations of error and error magnitude. A GLM was implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). The trial-by-trial regressors of interest that were included in the GLM were: onset of correct trials, earnable value 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 5000 iterations. The cluster-forming threshold was uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ).



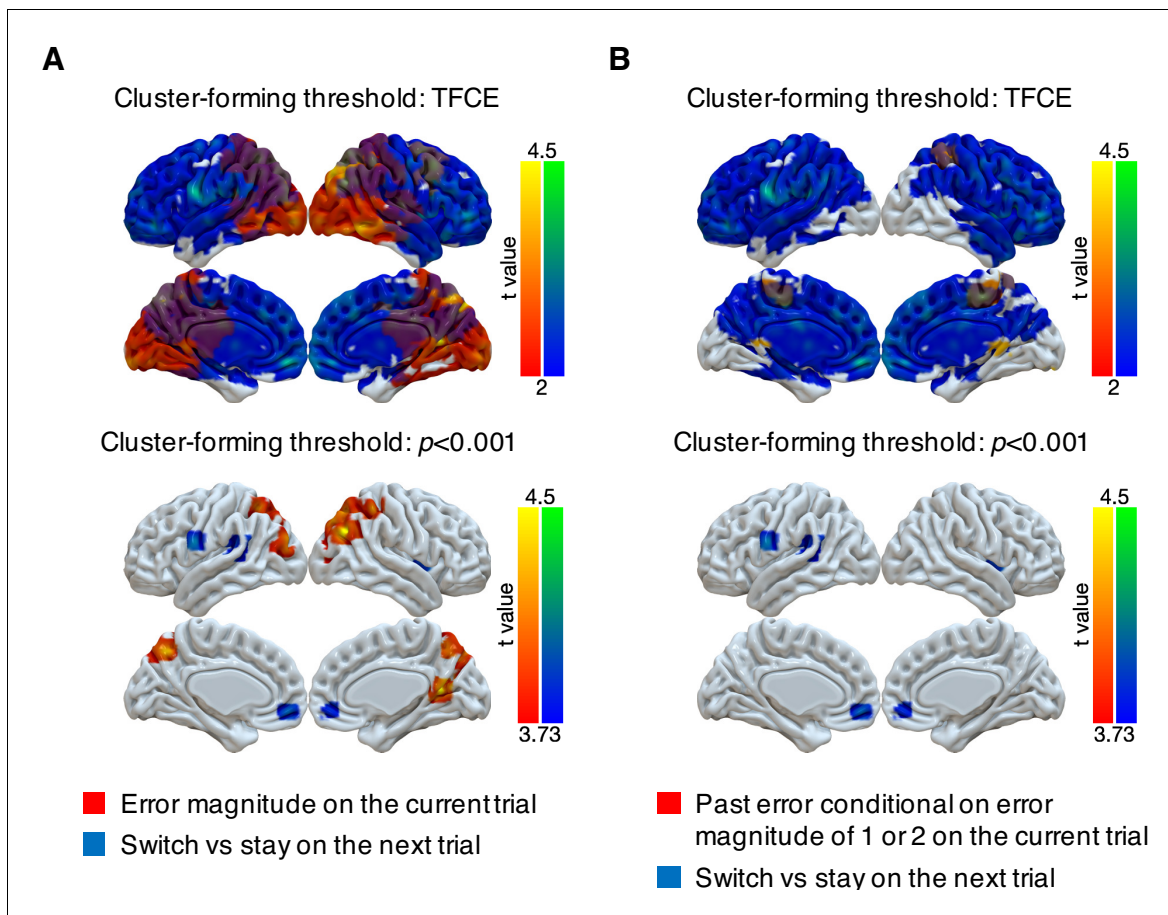
**Figure 4.** Representations of errors on the previous trial conditional on different types of current trials (columns). Multi-voxel neural patterns were used to classify correct responses versus errors on the previous trial. This analysis was repeated for different types of current trials: all feedback, correct feedback, error feedback, error magnitude of 0 or 3+, and error magnitude of 1 or 2. The representation of past errors is stronger in parietal cortex in the noisy condition than the unstable condition when the current trial is an error or the current error magnitude is 1 or 2. The cluster-forming threshold was an uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ), with cluster mass corrected for multiple comparisons using non-parametric permutation tests.



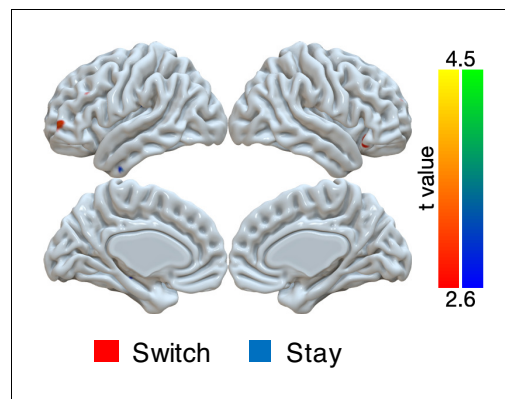
**Figure 4—figure supplement 1.** Univariate representations of error on the previous trial conditional on different types of current trials (columns). Several GLMs were implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). First, we examined errors on the previous trial 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 error magnitudes of 0 or 3+ or error magnitudes of 1 or 2. The trial-by-trial regressors of interest that were included in the GLM were: onset of current trials with error magnitudes of 0 or 3+, errors on trial  $t-1$ ,  $t-2$  or  $t-3$  conditional on the current trial error magnitude of 0 or 3+, onset of current trials with error magnitudes of 1 or 2, errors on trial  $t-1$ ,  $t-2$  or  $t-3$  conditional on the current trial error magnitude of 1 or 2. We focused on the effects of errors on trial  $t-1$  conditional on the current trials error magnitude of 0 or 3+ or error magnitude of 1 or 2. Group  $t$ -values are shown. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5000 iterations. The cluster-forming threshold was uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ).



**Figure 5.** Representations of subsequent behavioral choices (switch versus stay) after ambiguous small errors in the noisy condition. **(A)** Overlap of results for switch versus stay on the next trial and error magnitude on the current trial. Multi-voxel neural patterns were used to classify whether participants switch their choice to another target or stay on the same target on the next trial. We focused on the most ambiguous errors (error magnitude of 1 or two in the noisy condition). Above-chance classification performance was found in a large cluster encompassing the frontal lobe. The cluster-forming threshold was an uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ), with cluster mass corrected for multiple comparisons using non-parametric permutation tests. **(B)** Overlap of results for switch versus stay on the next trial and past error conditional on error magnitude of 1 or two on the current trial.



**Figure 5—figure supplement 1.** Representations of subsequent behavioral choices (switch versus stay) thresholded via threshold-free cluster enhancement (TFCE) or with a cluster-forming threshold of  $p < 0.001$ . (A) Overlap of results for switch versus stay on the next trial and error magnitude on the current trial. We implemented two types of cluster-forming approaches: TFCE and uncorrected voxel  $p < 0.001$ . First, significance testing was implemented through permutation tests with threshold-free cluster enhancement (FSL's randomize), which does not require a pre-defined cluster-forming threshold. The result of switch versus stay showed little spatial specificity. For the purpose of display, the results were thresholded based on uncorrected voxel  $p < 0.03$  ( $t = 2$ ). Second, we used a cluster-forming threshold of uncorrected voxel  $p < 0.001$  ( $t = 3.73$ ) and tested the significance of the formed cluster via one-sample cluster-mass permutation tests with 5000 iterations. The results showed high spatial specificity and several previously identified regions were still significant: middle cingulate cortex [14, -8, 30], right insula [38, 4, 2], medial OFC [-4, 50, -10], left premotor cortex [-62, 2, 24] and left superior temporal gyrus [-50, -32, 12]. (B) Overlap of results for switch versus stay on the next trial and past error conditional on error magnitude of 1 or 2 on the current trial. The two types of cluster-forming approaches are shown.



**Figure 5—figure supplement 2.** Univariate GLM for switch versus stay on small error trials (magnitudes of 1 or 2) in the noisy 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 trials with error magnitudes of 1 or two followed by switching, onset of trials with error magnitudes of 1 or two followed by staying. We tested the effects of the difference between switch and stay for small errors. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5000 iterations. The cluster-forming threshold was uncorrected voxel  $p < 0.01$  ( $t = 2.6$ ). There were no significant clusters. For the demonstration, the results were shown with uncorrected voxel  $p < 0.01$ .