Supplemental Figures

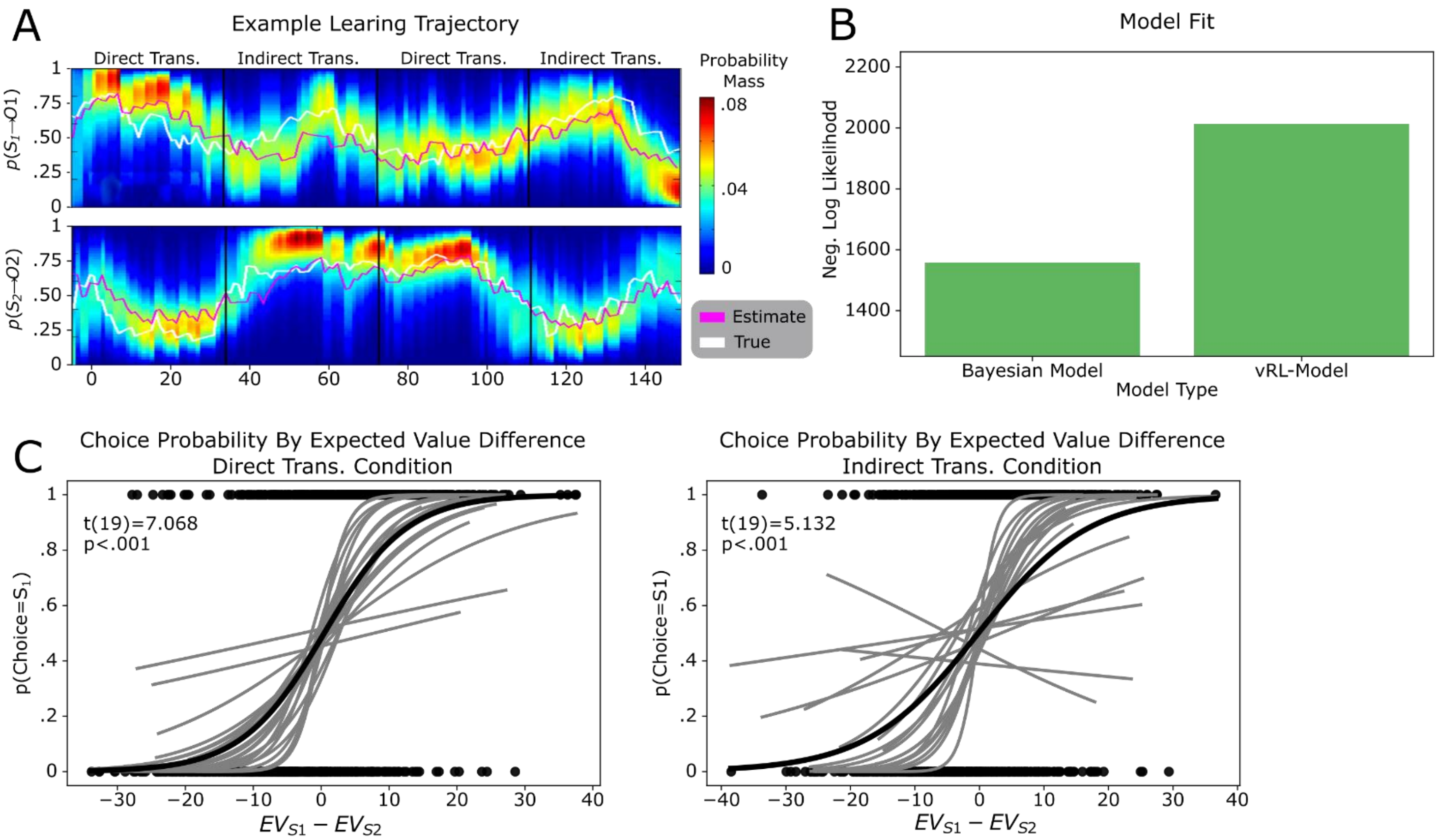

Figure S1. Follow up behavioral analyses

A. Example trajectory across the experiment of the belief estimates generated from the Bayesian learner. Top is the trajectory of S1, and the bottom is the trajectory of S2. While lines represent the true probability trajectory is shown in white and the estimated belief is shown in pink. Color heatmap shows the probability mass for each possible belief in S_x ->O1. B. Comparison of model fits between our Bayesian model and a value-based RL model (vRL) which used an interactive updating procedure to track the value of each shape based on the history of received rewards. The exceedance probability for the Bayesian model was 1, and 0 for the vRL model, suggesting that Bayesian model, which tracked transition probabilities between choices and outcomes, better fit participants actual choices compared to a value tracking model. **C**. Logistic regression curves estimating the change in choice probabilities given the expected value difference between choices. Gray line shows participant specific lines, and the black line shows the effect across groups (associated t-statistics are calculated across participants). The left side shows the effect in the direct transition condition and the right side shows the indirect transition condition.

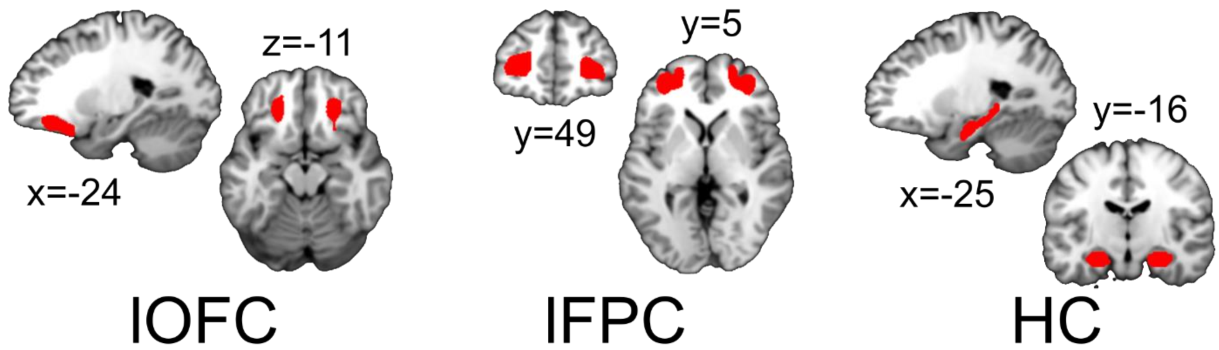

Figure S2. Pre-selected anatomical ROIs

Illustrations of pre-selected anatomical ROIs taken from Neubert et al, 2015. The lOFC ROI corresponds to index 9 and 30, lFPC corresponds to indexes 14 and 35. The HC ROI was defined in Yushkevich et al., 2015.

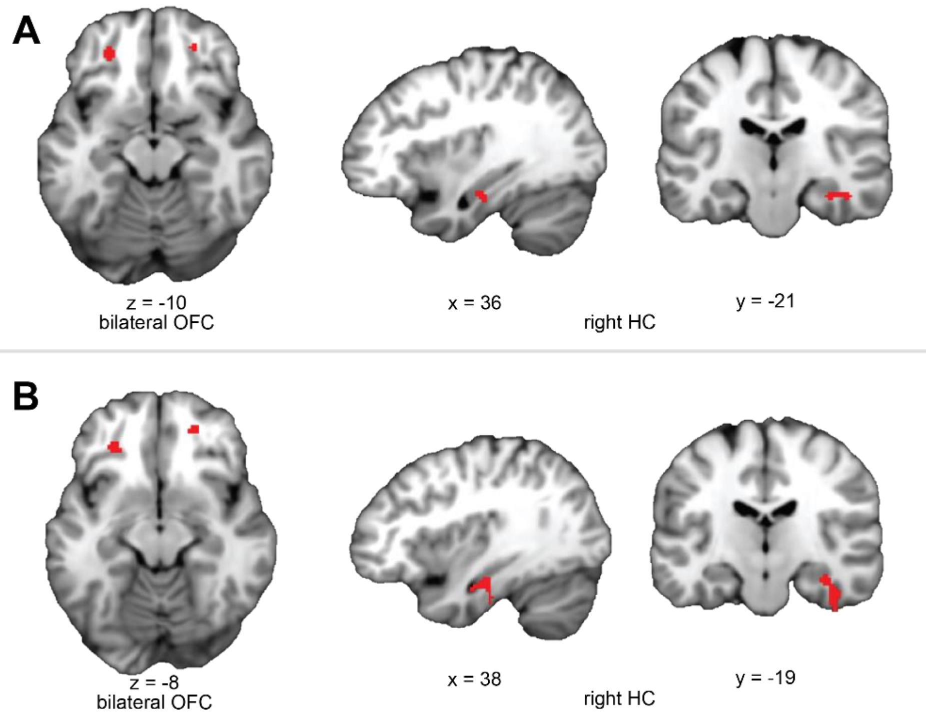

Figure S3. Functionally defined ROIs for in the direct transitions condition.
A) Despite having *a priori* defined anatomical ROIs for our decoding analysis of the causal choice, we wanted to test whether our results depended on these ROI definitions by using a data-driven approach. Here, we trained an SVM classifier to decode representations of the causal choice in run 1 of the direct transition condition, then tested the decoder on run 2 to find regions of the orbitofrontal cortex (OFC) and hippocampus (HC) that significantly decoded causal choice representations at a significance level of $t(19) > 2.54$, $p < .01$, uncorrected. We then used these regions as ROIs for a separate analysis which trained the classifier in run 1 and tested the classifier in run 2. B) Shows ROIs generated from the same procedure as described in A, but the use of each run for training and testing are switched.

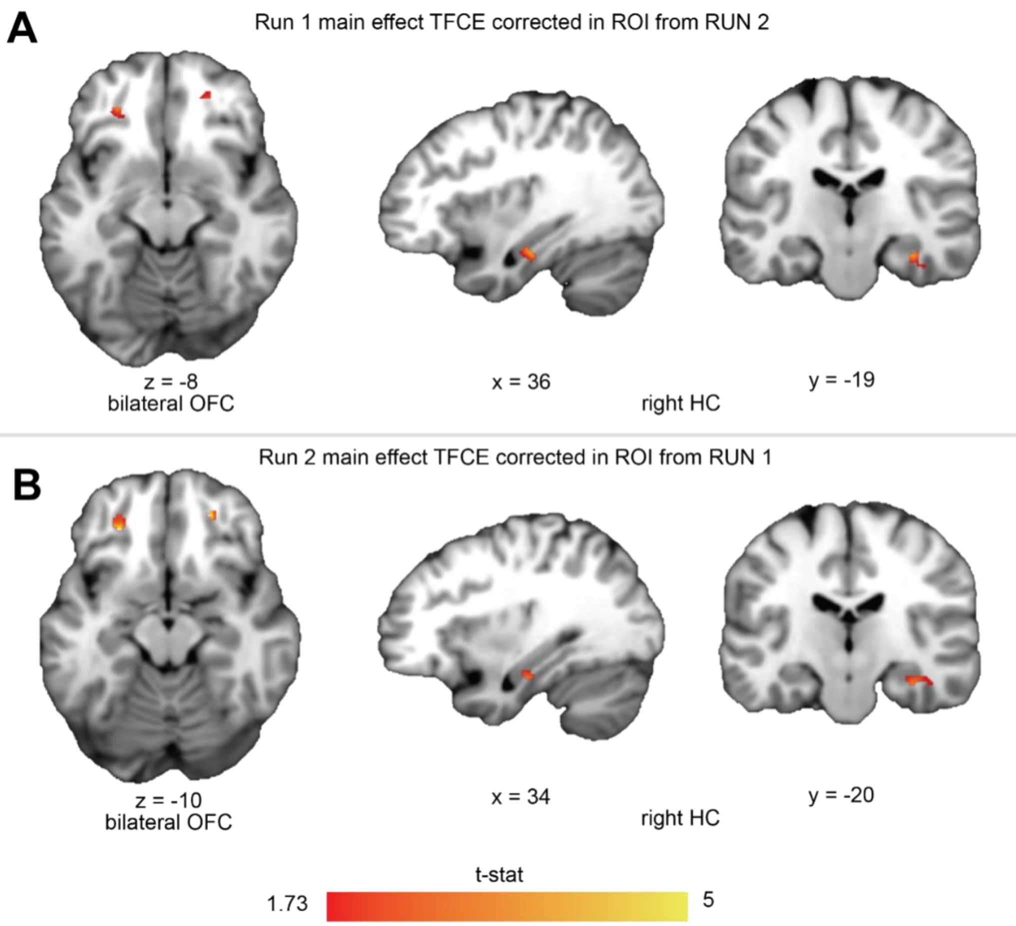

Figure S4. Main effect of choice decoding accuracy at the time of feedback TFCE corrected in each run of the direct transition condition

A. Regions of the OFC showing significant decoding of the causal choice in run 1 of the direct transition condition. Significance was tested using TFCE correction over voxels with the ROI generated from run 2, using the procedure described above (Fig.S1). For illustration, we show voxels that survive at threshold to $t(19)=1.73$, $p<.05$ uncorrected. B. Shows the same as A but for voxels in run 2, using the ROI generated from run 1.

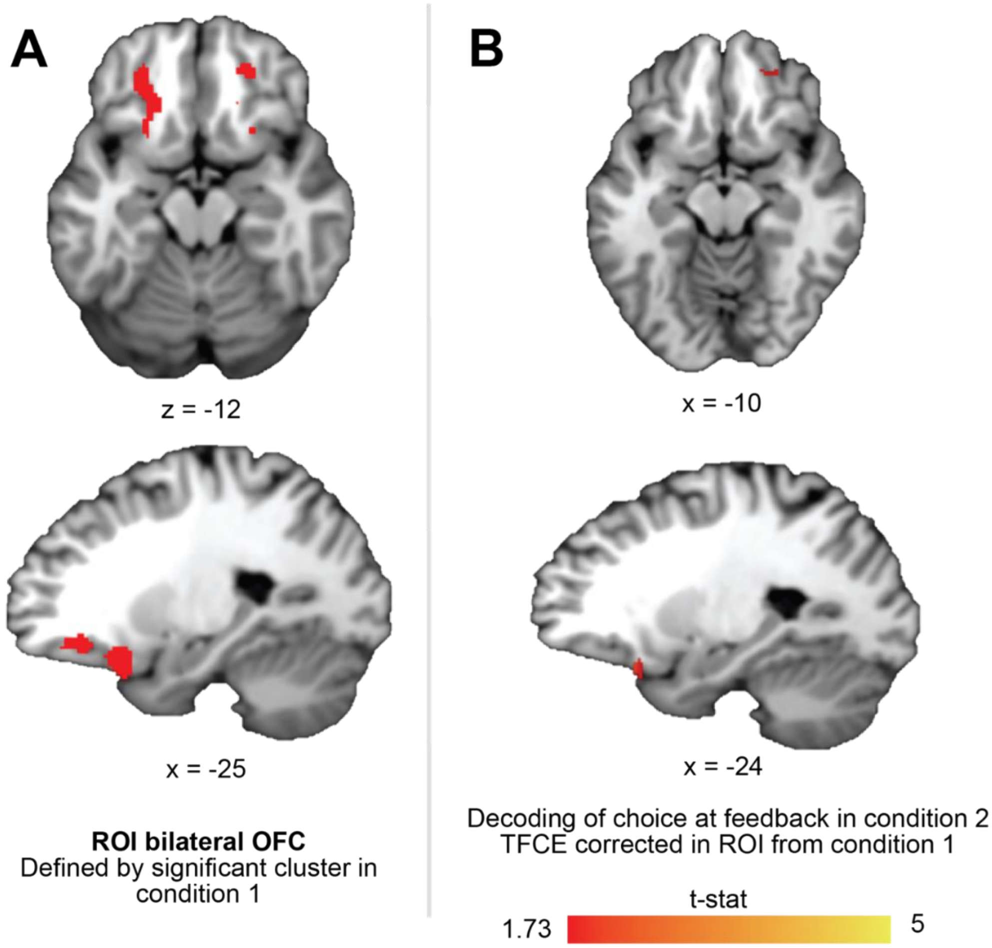

Figure S5. Significnant informaton connectivity between lFPC and OFC in functionally defined ROI from direct transition condition

A. We did not observe signficiant decoding of the causal choice a in bilateral OFC ROI defined by significant cluseter in in the idirected transition condition. Thus, we used the accuracy map for decoding choices at feedback during the direct transition condition ($t(19) > 1.73$; $p < .05$) in the OFC, averaged across runs. B) We then used those cluster as ROI for TFCE correction for regions of the lOFC that showed significant information connectivity with lFPC. We did this by testing for significant correlations between the trial-by-trial fidelity of pending representations in the lFPC and causal choice representation during feedback in lOFC (see Methods).

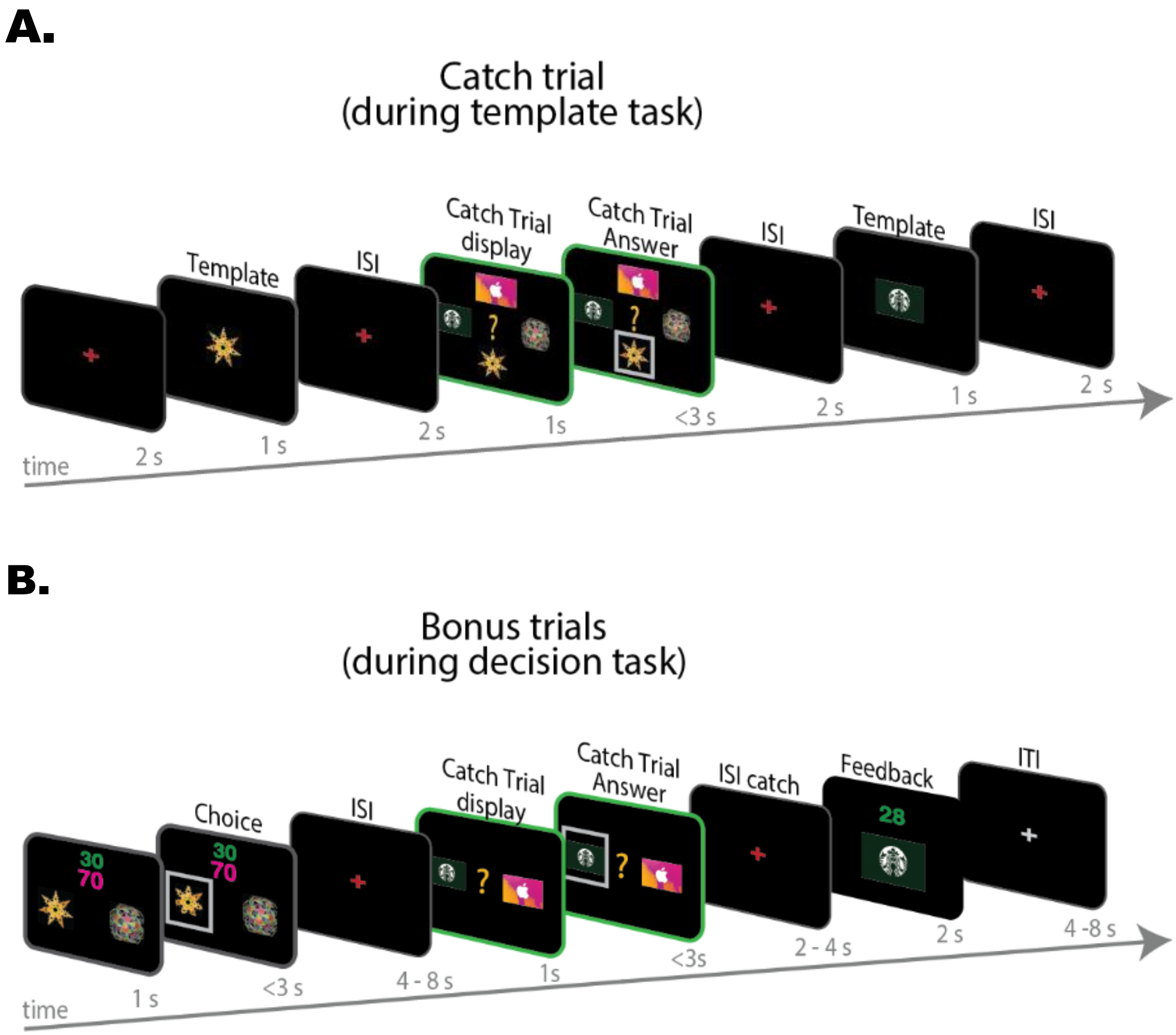

Figure S6. Depiction of catch trials

A. To ensure that participants where we included valuable *catch trials* in the passive observing "template task". Participants were asked to report which image out of the four (2 gift cards and 2 stimuli) was the last one presented on the screen. They were endowed an extra £10 from which we removed £1 for every incorrect response. There were four catch trials per template run. B. The decision task included "bonus trials" in which participants could predict which gift card they expected to see on the subsequent feedback screen given their choice. They were given 3£ extra on the final gift card that was given to them for every correct answer. The first run of the direct transition condition had two catch trials; the second run had one. Both runs of the indirect transition condition had one catch trial each.

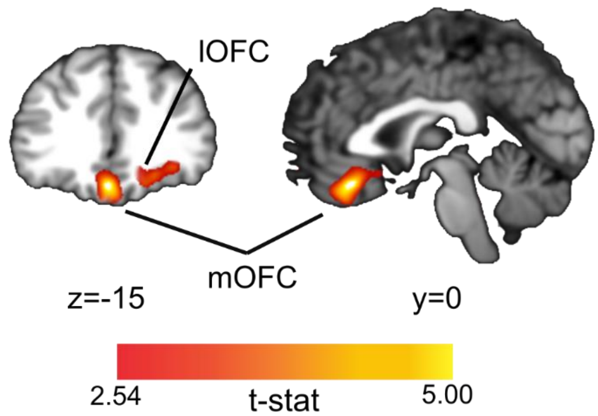

Figure S7. Exploratory information connectivity analysis for "Indirect transition condition" To ascertain whether additional regions maybe involved in credit assignment beyond those that formed the focus of our study, we repeated the analysis described in Figure 2 and Figure 4 but using a whole brain search light procedure. All aspects of the analyses were the same as those previously conducted except that we corrected for multiple comparisons at the whole brain level using TFCE. For the "direct transition condition", we found no additional regions that showed high decoding for the causal choice at the time of outcome. However, for the "indirect transition condition" we identified a region of medial OFC (mOFC) which showed information about the causal choice that was predicted by pending representation in lFPC (pTFCE<.05). The left panel shows a coronal slice through a t-statistic map, thresholded using the same conventions as Figure 4; the right panel shows a sagittal slice through the same map. These results suggest a potential role for mOFC in credit assignment uniquely during the "indirect transition condition".

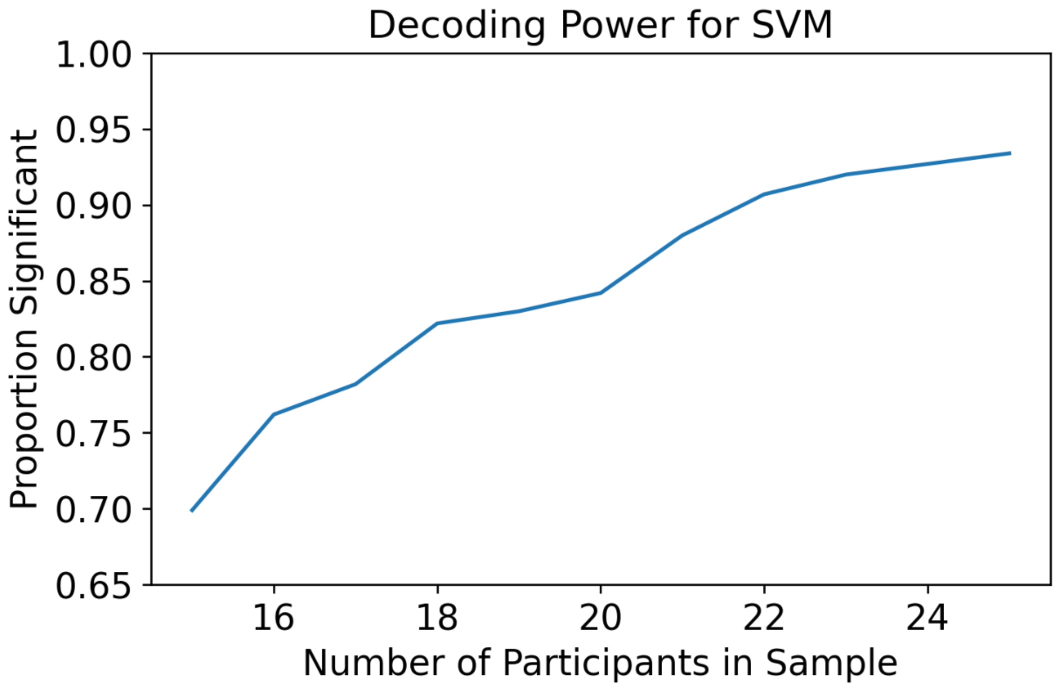

Figure S8. Power analysis for Reinstatement Effect in the lOFC
Power analysis using an independent data set. Twenty-eight participants competed an associative learning task, in which they learned the causal associations between four different choices, and two food rewards. We estimated voxel activity at the time of the outcome for each trial and tested for multivariate patterns of the causal choice in the lOFC, using the same procedures described in the main text (see Methods). We began by drawing 1000 samples of participants of size N, with replacement, for values of N ranging from 15 to 25. We then tested for significant decoding of the causal choice within each subset using small-volume TFCE correction. Finally, we calculated the proportion of these samples that were at or below a significance level of *pTFCE* <.05.