How instructed knowledge modulates the neural systems of reward learning

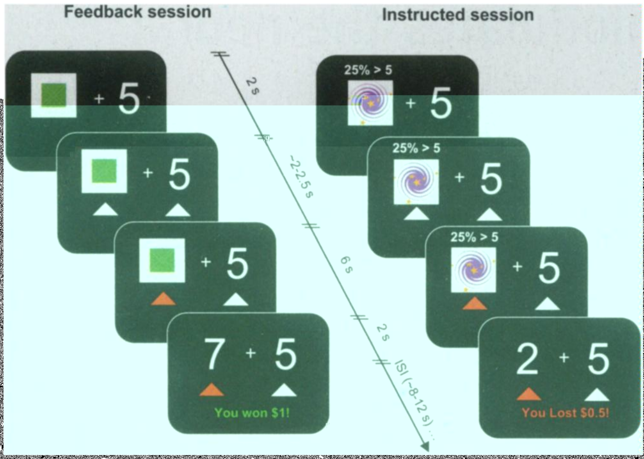
Jian Li[a,b], Mauricio R. Delgado[c], and Elizabeth A. Phelps[a,b,1]

[a]Department of Psychology, [b]Center for Neural Science, New York University, New York, NY 10003; and [c]Department of Psychology, Rutgers University, Newark, NJ 07102

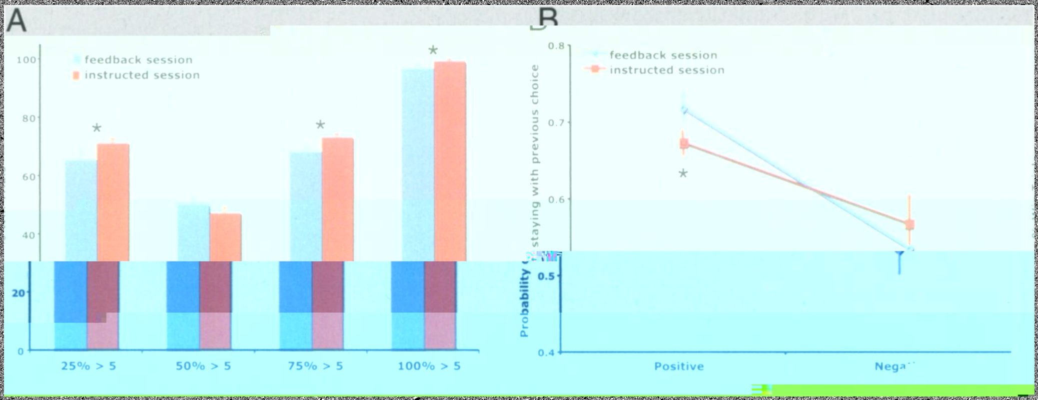
Edited by Edward E. Smith, Columbia University, New York, NY, and approved November 19, 2010 (received for review October 5, 2010)

Recent research in neuroeconomics has demonstrated that the reinforcement learning model of reward learning captures the patterns of both behavioral performance and neural responses during a range of economic decision-making tasks. However, this

knowledge on choice selection (12, 13). We designed an experiment with two sessions. In the "feedback" session, participants' choices were only based on the win/loss feedback, and in the "instructed" session participants could also incorporate the



instructed by experimenter, as indicated by the initial Q value associated with different stimuli ($\alpha = 0$) (*SI Appendix*, Table S2).



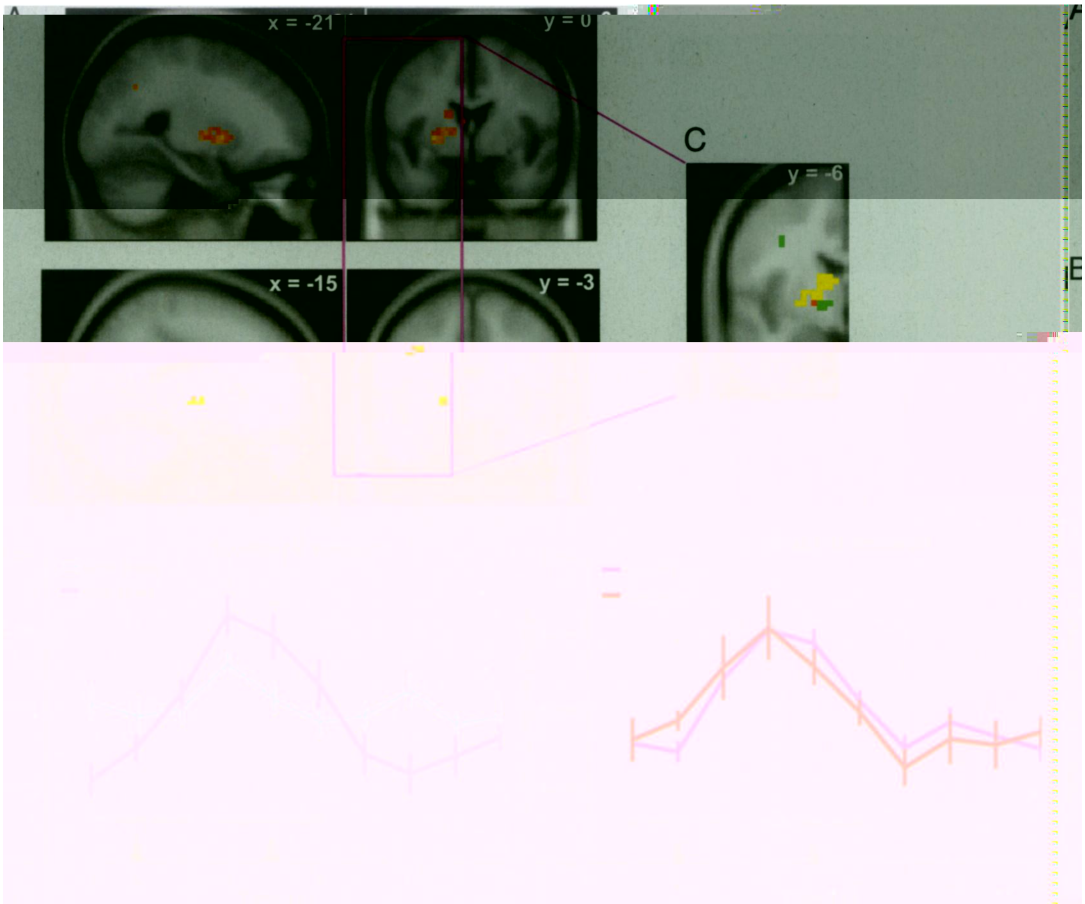


Fig. 3. BOLD responses for prediction errors in both sessions. (*A*) Activity of the striatum showed significant correlation to the PE signal in the feedback session ($P < 0.05$, corrected). Such correlations were not observed in the above structures in the instructed session ($P < 0.01$, uncorrected). (*B*) A two-way ANOVA showed an interaction between session (feedback and instructed) and learning phase (early and late) in the left striatum. (*C*) Striatal activation identified in the PE (*A*, yellow) and session × learning phase interaction (*B*, green) analyses, and the overlapping region (red). (*D*) BOLD response patterns in

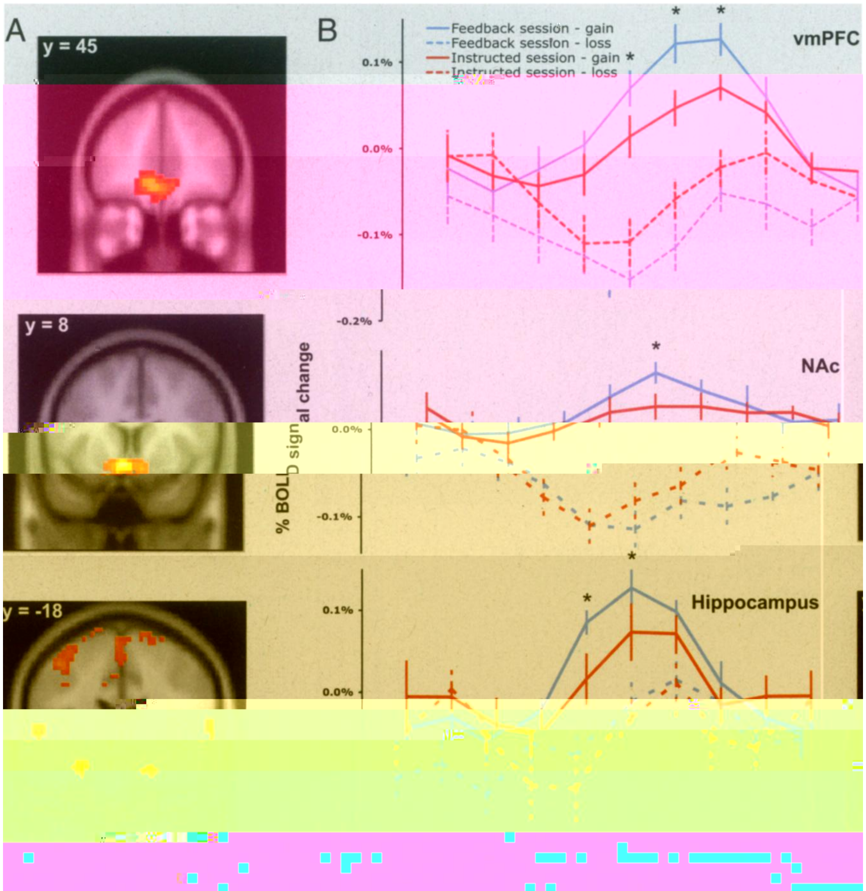
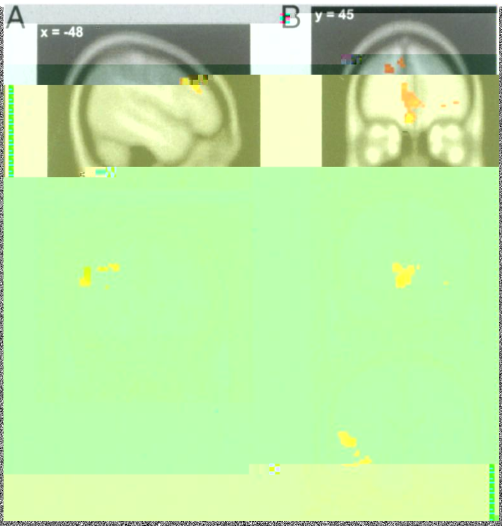


Fig. 4. BOLD responses discriminating win and loss for both sessions. (*A*) A whole-brain analysis revealed greater activation in the NAc, vmPFC, and bilateral hippocampal complex for win than loss trials across both sessions ($P < 0.05$, corrected). (*B*) BOLD time course of activation in the NAc, vmPFC, and bilateral hippocampal complex for win and loss trials in the feedback and instructed sessions (*, significant difference of time points near activation peaks, $P < 0.05$; ± SEM).

showed a greater BOLD response to win outcomes during the instructed session (Fig. 5*A* and *SI Appendix*, Table S5).

Functional connectivity between DLPFC and reward-related brain structures. The DLPFC has previously been implicated in decision-selection (17, 38, 39). When feedback is the only source of information, choice-dependent outcomes can be evaluated and fed back to valuation systems to provide a better approximation of action values and guide individuals toward choices that maximize


A
x = -48
B
y = 45

our model against others suggested in the literature based on behavioral data with similar tasks (27, 28) using the Bayesian information criterion as a criterion for model selection. For the feedback session, the simple RL with one learning rate (α) for both positive and negative prediction errors fits participants' behavior better. However, RL with different learning rates (α_+ and α_-) for positive and negative (δ_+ and δ_-) PEs fits participants' choices the best in the instructed session (see *SI Appendix* for details).

Imaging Analysis. We first regressed PEs that were generated for both the feedback and instructed sessions using the best-fitting parameters to the whole-brain BOLD signals at the revelation of monetary outcome to identify the brain areas whose activities were correlated with the calculation of PE. Monetary outcomes were also included as dummy regressors to account for the effect of the magnitude of the reward value.

The finite impulse response from time 0 to ~12 s (TR0 to ~TR6) was generated by resampling the BOLD time series of each voxel in the brain and averaging across 40 trials each for the early and late learning phases in both sessions. Because canonical hemodynamic response function typically peaks at 6 to ~8 s after the stimulus onset, the two-way ANOVA was performed on both TR3 (6 s) and TR4 (8 s). These whole-brain analyses were performed on each voxel to identify brain regions that showed a significant interaction effect with time (i.e., early vs. late learning) and session (i.e., feedback vs. instructed session).

Finally, we conducted a PPI analysis to investigate the connectivity between brain regions that may modulate the impact of instructed knowledge on RL learning signals (see *SI Appendix* for technical details).

ACKNOWLEDGMENTS. We thank K. Sanzenbach and the Center for Brain