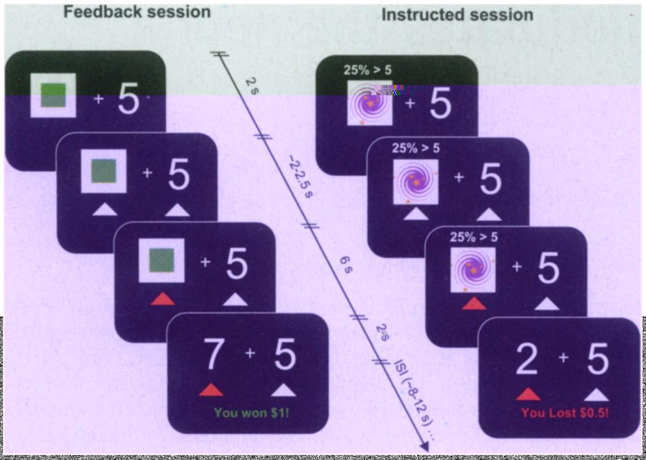
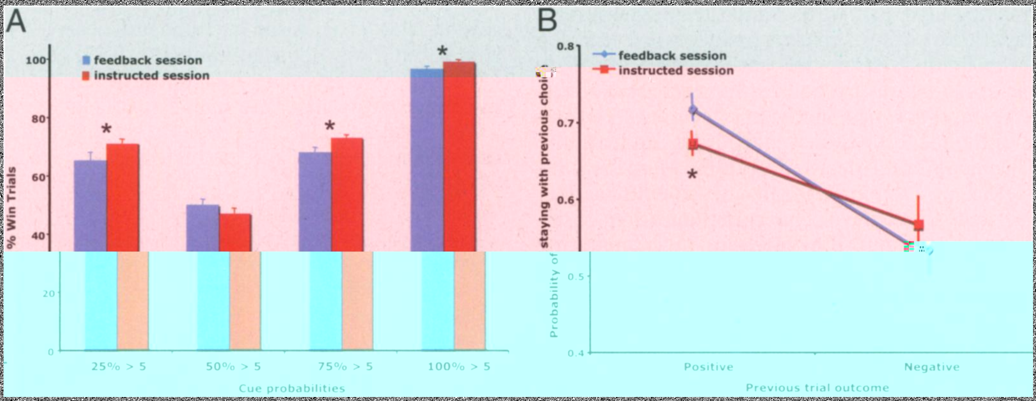
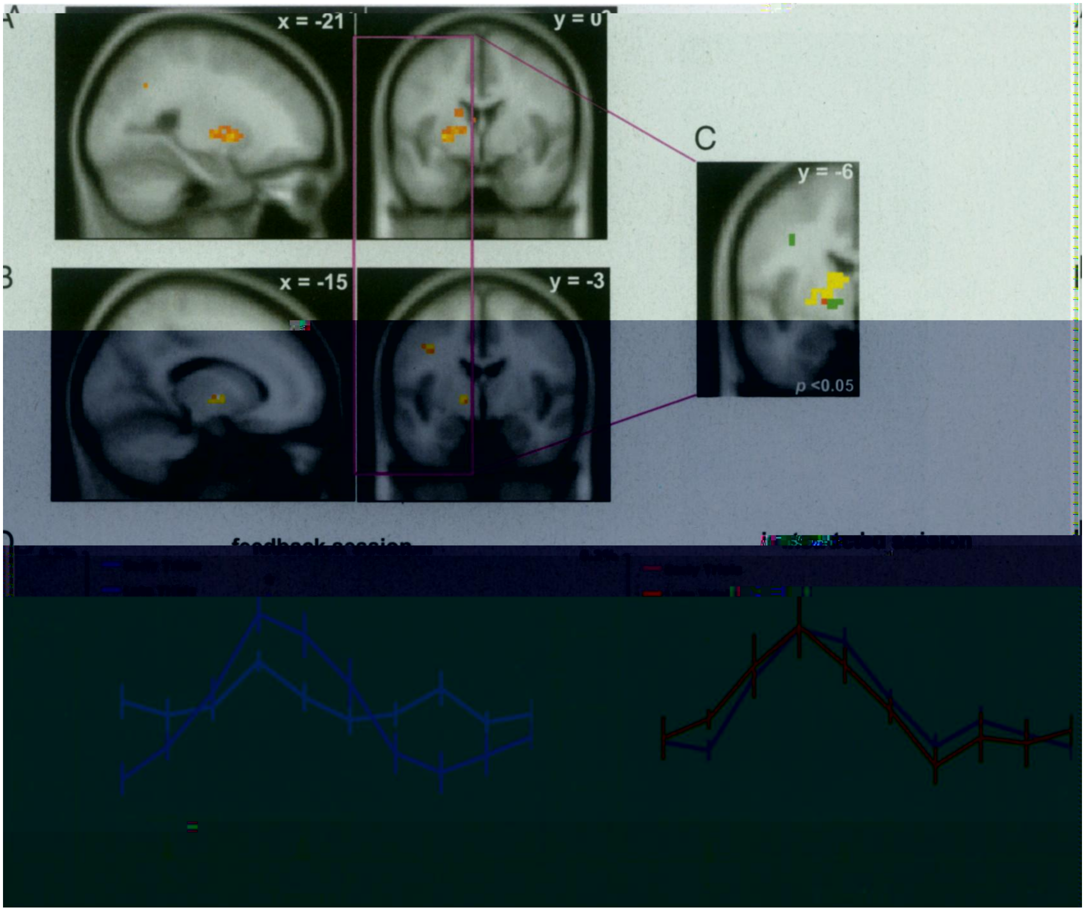
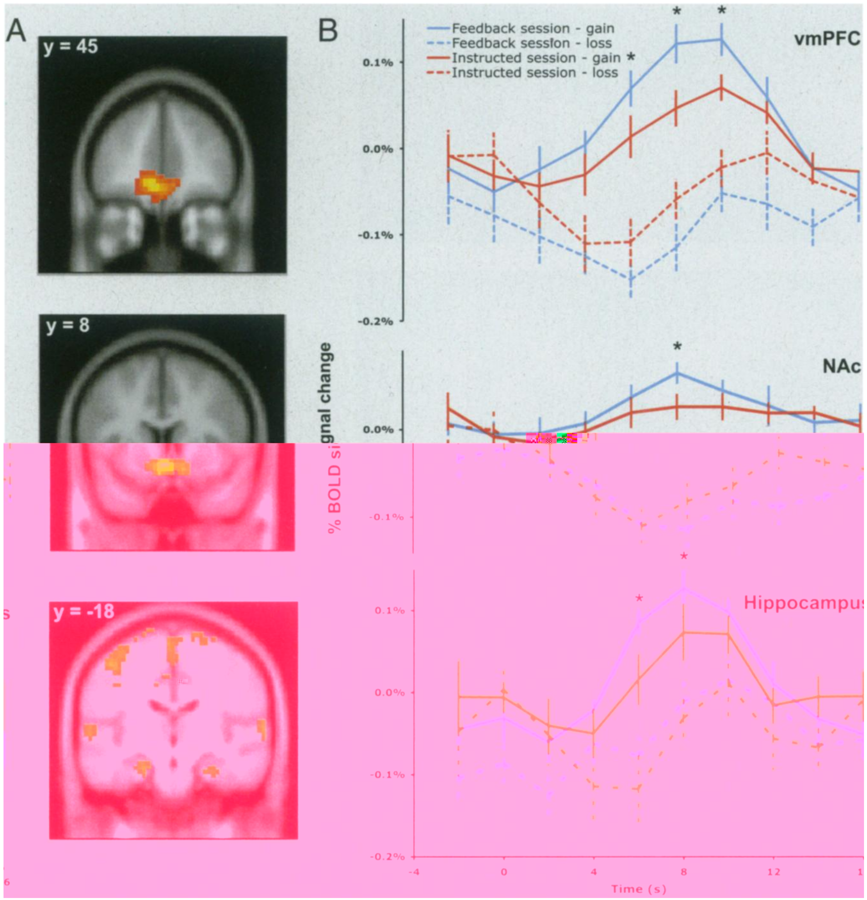


instructed by experimenter, as indicated by the initial Q value associated with different stimuli ($\alpha_{-} = 0$) (*SI Appendix*, Table S2).

Functional MRI Results. ***RL model predicts BOLD signals in the feedback session.*** Our behavioral results suggest that a RL model captures participants' performance in the feedback session, but does not adequately describe learning in the instructed session. To explore if a similar pattern was reflected in the patterns of BOLD responses, we constructed a general linear model (GLM) with the PE regressors generated from the best fitting Q-learning models for both sessions (*SI Appendix*, Tables S1 and S2) and investigated the neural correlates of PE in the feedback session and




A
x = -21
y = 0
C
y = -6
B
x = -15
y = -3
p <0.05
feedback session



A
y = 45
y = 8
y = -18
B
Feedback session - gain
Feedback session - loss
Instructed session - gain
Instructed session - loss
vmPFC
NAc
Hippocampus
% BOLD signal change
0.1%
0.0%
-0.1%
-0.2%
Time (s)

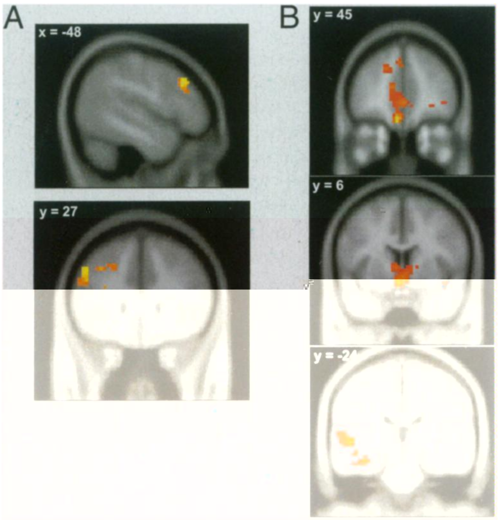


Fig. 5. Left DLPFC activity showed negative functional connectivity to brain structures related to reward valuation. (*A*) Left DLPFC showed relatively greater activation to monetary gains in the instructed than the feedback session ($P <$

ward-related structure circuitry for learning and reward processing, previous neuroeconomic research has outlined a similar circuitry across a range of decision-making tasks in which preexisting reward values that are represented in valuation regions can be modulated based on social processes (51, 52), goals (34), or other cognitive factors (52, 53). Although there have been suggestions that the DLPFC and reward-related regions represent independent systems in the brain competing with each other for the dominance of action selection (27, 28, 53, 54), our results are more consistent with a general role for the DLPFC in modulating the engagement of reward-related regions depending on the relative importance of the information during a learning paradigm.

Our findings also lend neurological evidence to support recent computational approaches to reconcile a broad range of literatures suggesting multiple representation systems in the brain for behavioral control. One such system deploys a model-free method and "learns putatively simpler quantities," such as policies that are sufficient to permit optimal performance through processing action outcomes. It is suggested this computation is carried out in the dorsolateral striatum. The other system, which employs the prefrontal cortex, adopts a model-based method to make use of available or learned rules and derives optimal choice through dynamic programming. The brain arbitrates between different representation systems according to the uncertainty estimated from each system (38,