Supplementary Materials

Supplemental behavioral results

Comparison between decomposed TD and null choice model

To check that the decomposed TD model predicts choice behavior better than chance, we compared it to a "null choice" model in which all possible choices have equal probability. Since the null choice model is a reduced version of the decomposed TD model, we calculated the likelihood ratio test statistic:

$$LR = -2*(LL_{TD} - LL_{null}),$$

Where *LL* denotes the maximized (fixed effects) log-likelihood. For a chi-squared test with 64 degrees of freedom, the likelihood test statistic was 6049.7 and the p-value was astronomically small in favor of the decomposed TD model. Thus, our model predicts choice behavior significantly better than chance.

As an additional comparison, we asked whether subjects earned significantly more reward in the task than would be expected by chance. To do so, we calculated the expected reward under the null choice model, and compared this quantity across subjects to their actual earned rewards. Using a paired-sample t-test, we found that earned reward was significantly higher than would be expected by chance (t_{15}=6.5, p<0.00001).

Finally, to calculate how much variance in the choice behavior was accounted for by the decomposed TD model, we calculated the pseudo-R^2 value, which showed that the model accounted for 47% of the variance in choice behavior.

Logistic regression analyses of choice behavior

We also examined a set of nested logistic regression models as an alternative way to investigate decomposition in choice behavior. These models represent Q-values by a linear combination of predictor variables:

$$Q_t(\mathbf{a}_t) = \mathbf{X}_t\mathbf{w}$$

where $\mathbf{X}_t$ is a C x K design matrix of C joint choice options and K predictor variables (regressors) and $\mathbf{w}$ is a K x 1 vector of regression coefficients. Just as in the TD model, choice probabilities are modeled as a softmax function of the Q-values. In the *full* model, we included 3 regressors:

1. A “joint” reward regressor a particular joint action was rewarded on the previous trial.
2. A “decomposed” reward regressor expressing whether a particular sub-action was rewarded on the previous trial. In other words, each effector’s sub-action gets credit regardless of what the other effector’s sub-action was on the previous trial.
3. A “joint” choice regressor expressing whether a particular joint action was chosen on the previous trial (we could also have included a “decomposed” choice regressor, but we chose to omit this for simplicity).

In the *reduced* model, we removed the decomposed reward regressor. The first key question of interest is whether the regression coefficient for the decomposed reward regressor in the full model was significantly greater than zero, which would indicate that subjects were exploiting the decomposition structure in the rewards. We found this to be the case ($p<0.0001$).

To further support this conclusion, we performed a chi-squared test on the likelihood ratio test statistic (see above) between the full and reduced models. We found $LR=178.2$ ($p<0.00000001$).

Supplemental behavioral experiments

Methods

We performed two additional behavioral experiments to further investigate the predictions of our models. For Experiment 1, twelve subjects participated in the study. For Experiment 2, eleven subjects participated in the study. For both experiments, informed consent was obtained in a manner approved by the New York University Committee on Activities involving Human Subjects.

One major prediction is that the decomposed model should only fit behavior better when the reward structure of the task is actually decomposable into separate effector-specific

components. To test this prediction, in Experiment 1 we designed an “un-decomposable” version of the task which was identical to the decomposed version described in the Methods section, except for 3 differences: (1) each joint action was associated with a unique reward probability; (2) only a single reward was presented on each trial (this is by necessity, since effector-specific rewards no longer exist in this version); and (3) the spatial ordering of the options on the screen were scrambled to discourage subjects from adopting a decomposed learning strategy. We also designed a slightly modified version of the decomposable task described in the Methods (Experiment 2), where a single, summed reward was shown on the screen rather than separate rewards for each effector. We chose to do this so as to equate the decomposable and un-decomposable tasks as much as possible.

One additional complication is that the decomposed TD model described in the Methods section can no longer be applied to these tasks because the separate effector-specific rewards are not available. Thus, we created an alternative decomposed TD model which operates on the summed reward. Note that in the decomposable task when both effectors are rewarded the summed reward will always be $2, and when neither are rewarded the summed reward will be $0. Thus, these two cases have no ambiguity with respect to the effector-specific reward in the decomposable task. As a consequence, it is possible to use the same decomposed TD model for these cases as described in the Methods section. The only difference is that for the case when subjects receive a $1 reward (where there is true ambiguity as to which effector earned the reward), we assumed that subjects divide the reward equally between effectors.

Results

We calculated an approximate Bayes Factor *BF* (see Methods) between the decomposed and joint models for each experiment. A $BF > 4.6$ represents strong evidence in favor of the decomposed model. We found that for the decomposable experiment $BF = 239$, whereas for the un-decomposable experiment $BF = -27$. These results suggest that humans will adopt a decomposed learning strategy only when the reward structure of the task actually admits such a decomposition.

Brain region	X	Y	Z	Z-score
Premotor Cortex	24	8	46	6.28
Inferior Frontal Gyrus	-54	4	34	4.79
Intraparietal Sulcus	-34	-38	38	5.69
Angular Gyrus	48	-28	42	5.32
Angular Gyrus	60	-48	40	5.2
Cerebellum	8	-12	2	4.94
Parahippocampal Gyrus	-12	-24	4	5.27

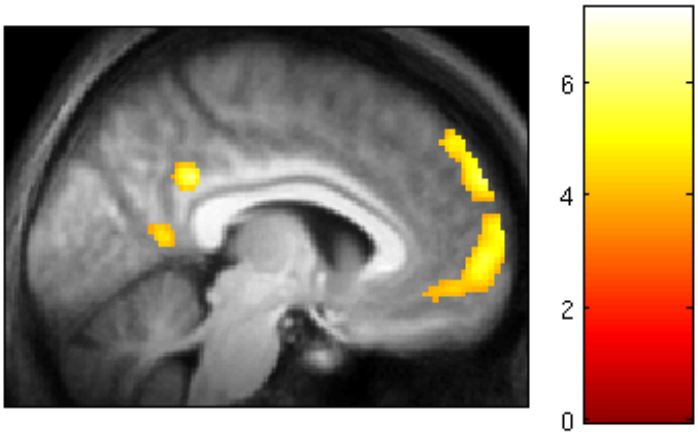
Supplemental Table 1: Voxels displaying a negative correlation with average chosen value, $p<0.05$, FWE-corrected. Note that no voxels survived this threshold for the positive correlation with average value.

Brain region	X	Y	Z	Z-score
Cuneus	34	-84	0	6.02
V1	-32	-90	-2	5.93

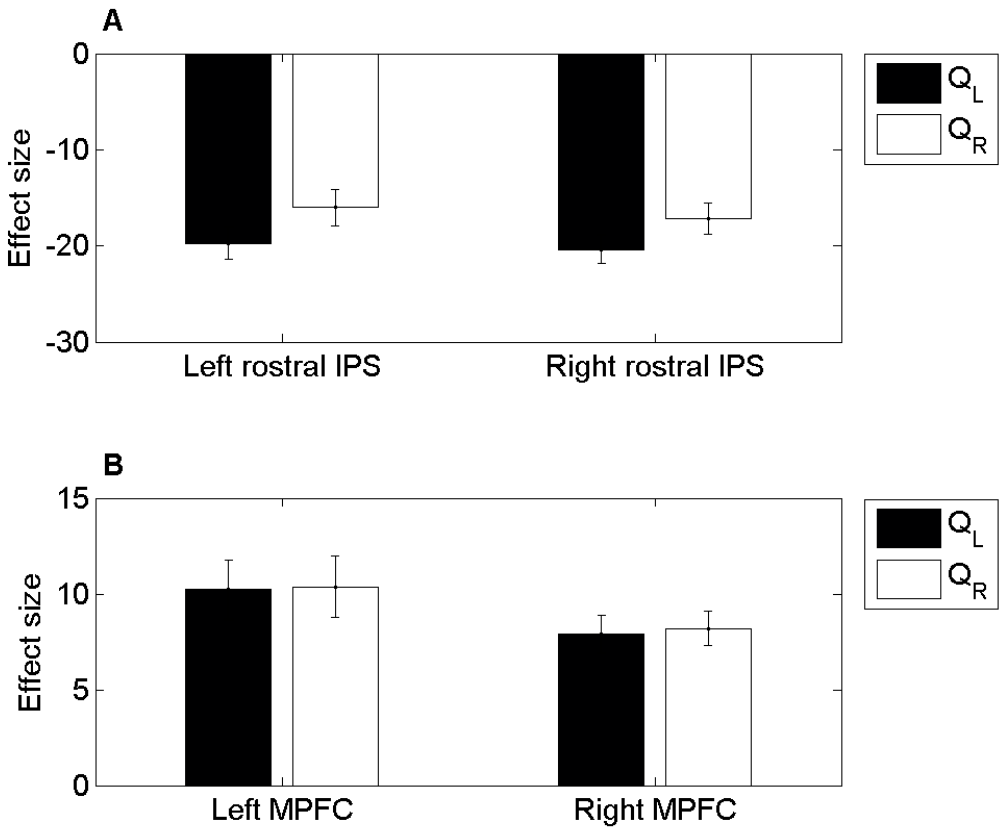
Supplemental Table 2: Voxels displaying positive correlation with average prediction error, $p<0.05$, FWE-corrected.

Brain region	Q_L	Q_R
Left caudal IPS	-4.50	-3.46
Right caudal IPS	-3.53	-4.53
Left rostral IPS	-4.59	-4.14
Right rostral IPS	-5.04	-3.21
Left mPFC	2.29	2.42
Right mPFC	2.05	2.80
	δ_L	δ_R
Left ventral striatum	2.92	2.67
Right ventral striatum	3.37	3.67

Supplemental Table 3: Z-values for chosen value or prediction error effects in the voxels of interest, presented separately for each effector. Note that these statistics are uncorrected for multiple comparisons (using the correlated contrasts, Q_L+Q_R or $\delta_L+\delta_R$) used to select these voxels.



Supplemental Figure S1: Sagittal slice (x=-5) of MPFC activation for the Q_L+Q_R contrast, thresholded at $p<0.001$ (uncorrected).



Supplemental Figure S2: *Parameter estimates in functional VOIs*. (a) Responses in rostral IPS to the left and right value regressors, separated by left (-52, -28, 42) and right (48, -28, 42) hemisphere. (b) Responses in MPFC to the left and right value regressors, separated by left (-34, -38, -38) and right (38, -34, 44) hemisphere.