Supplementary Information

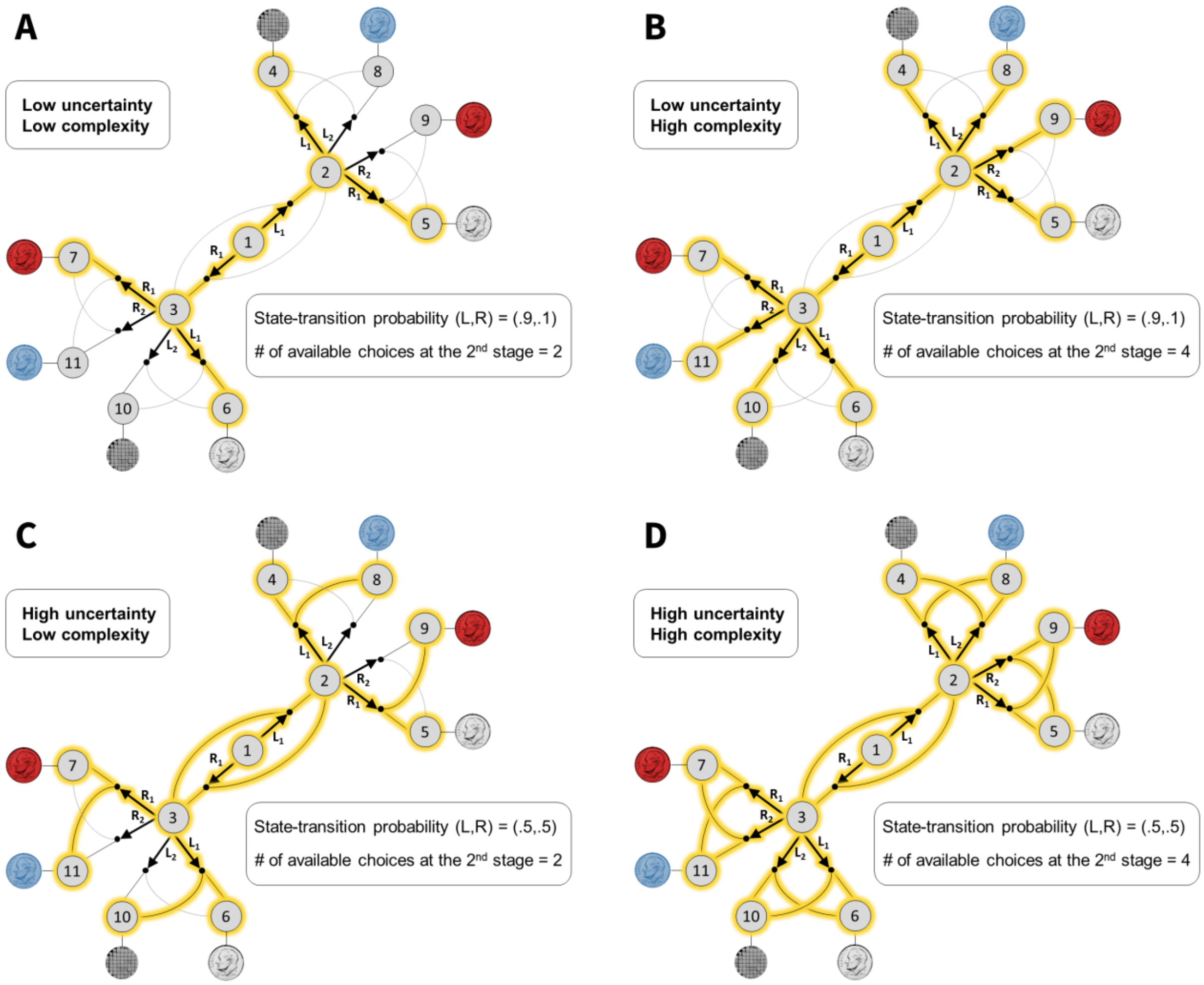

Figure S1. Illustration of four different types of conditions in the task (low/high x uncertainty/complexity)

Neural correlates of goal change

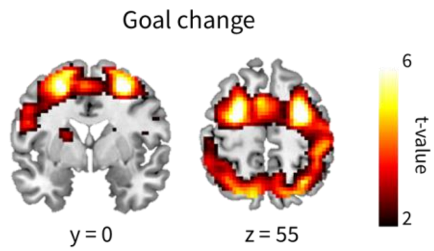

Figure S2. Neural correlates of goal change. Medial frontal gyrus encodes a goal change signal indicating whether the goal needs to be changed from a previous trial. This signal is necessary for goal-driven MF RL. In all the brain images, statistical significancy of effects is illustrated by the heat colormap. Threshold set at p<0.005.

Table S1. Estimated parameter values of the model (the best version according to the model comparison). Related to Figure 3B.

Subject \ Parameter	1	2	3	4	5	6
1	0.3039	0.1476	1.1474	4.9426	0.3998	0.1467
2	0.4099	0.3898	1.7325	7.2658	0.1251	0.0983
3	0.3273	0.0641	1.0149	9.9897	0.2474	0.1500
4	0.4531	0.3963	6.1229	1.6000	0.3540	0.0500
5	0.6669	0.1782	8.5423	5.8020	0.1001	0.0501
6	0.5085	0.1553	2.4221	1.6583	0.3363	0.0953
7	0.5298	0.1970	8.1945	1.0182	0.2462	0.0675
8	0.5934	0.1596	1.1625	4.2577	0.4000	0.0722
9	0.5148	0.0760	6.6298	1.3031	0.3999	0.1500
10	0.3015	0.2655	9.6187	1.0402	0.3979	0.0581
11	0.3840	0.1493	1.0913	3.1196	0.3603	0.1211
12	0.5837	0.1333	1.8027	6.7621	0.1137	0.1498
13	0.3841	0.0540	1.0013	9.9992	0.2381	0.1497
14	0.4858	0.1710	2.0061	2.7555	0.2732	0.1309
15	0.6117	0.3467	3.8635	1.9669	0.3958	0.0533
16	0.6662	0.1181	3.1976	1.2406	0.1345	0.1498
17	0.7322	0.4402	1.0022	9.3621	0.3817	0.0739
18	0.4333	0.1671	6.2510	1.2109	0.4000	0.1341
19	0.4260	0.1986	1.8499	2.6964	0.2915	0.1179
20	0.5091	0.2434	1.3119	1.9792	0.3996	0.0660
21	0.4707	0.3337	5.4166	1.2503	0.1446	0.0917
22	0.7432	0.0500	1.0114	8.0928	0.1650	0.1500
23	0.4739	0.1642	1.0013	9.9274	0.3972	0.1361
24	0.4397	0.2298	3.4788	3.4961	0.2193	0.1070

Parameter: 1- the threshold for defining zero state prediction error, 2- learning rate for the estimate of absolute reward prediction error, 3- the amplitude of a transition rate function (MB→MF), 4- the amplitude of a transition rate function (MF→MB), 5- inverse softmax temperature, and 6- learning rate of the model- based and the model-free, respectively.

Table S2. Neural signatures of the model-based, the model-free, and the arbitration system signals.

x	y	z	Peak in region	Hemi	p	# of voxels in the cluster	Z score	T score
			State prediction error (SPE)					
48	17	28	lPFC	R	0.000	73	5.62*	8.98
-36	14	28	lPFC	L	0.003	23	5.27*	7.92
33	20	1	Insula	R	0.013	8	4.97*	7.11
-30	17	1	Insula	L	0.069	-	2.76[1]	3.09
			Reward prediction error (RPE)					
-9	5	-8	Ventral striatum	L	0.010	-	3.52[2]	4.21
15	5	-8	Ventral striatum	R	0.046	-	2.42[3]	2.64
			Goal change					
-24	-4	52	MFG	R	0.009	13	4.98*	7.16
27	-1	55	MFG	L	0.013	14	4.89*	6.92
			Max reliability					
45	23	-11	ilPFC	R	0.000	197	4.55+	6.13
6	38	46	FPC	R	0.003	133	4.46+	5.94
-42	26	-2	ilPFC	L	0.021	84	4.53+	6.09
			Complexity (negative correlation)					
-18	5	58	SMA/MFG	L	0.039	98	3.58+	4.30
			Interaction of complexity and Max reliability – negative correlation					
54	23	7	ilPFC	R	0.003	157	4.75+	6.59
57	-40	4	STG	R	0.021	98	4.43+	5.88
-45	23	4	ilPFC	L	0.023	-	3.30[4]	3.86
			Chosen value of the goal-driven model-free (Q_{MF})					
-27	-1	61	SMA	L	0.000	123	5.80*	9.55
21	2	55	SMA	R	0.000	61	5.85*	9.73

-33	41	25	lPFC	L	0.002	31	5.34+	8.13
30	44	31	lPFC	R	0.000	225	4.68+	6.43
-36	-4	1	Posterior putamen	L	0.032	-	3.20[5]	3.71
Chosen value of the goal-driven model-free (Q_{MB})								
0	8	55	SMA	L/R	0.009	24	5.07*	7.37
-27	2	61	MFG	L	0.001	26	5.51*	8.63
27	8	49	MFG	R	0.016	6	4.93*	7.02
-30	53	13	lPFC	L	0.000	429	4.86+	6.86
Value difference of the arbitration system (chosen – unchosen)								
-12	23	-5	vmPFC	L	0.041	-	3.11[6]	3.57

* : threshold $p < 0.05$ FWE (voxel-level), the number of voxels in the cluster counted with the voxel-level threshold

+ : survives after whole-brain correction at the cluster-level (height threshold T = 3.55, threshold $p < 0.05$ FWE (cluster-level)), the number of voxels in the cluster counted with the cluster-level threshold

[1, 2, 3, 4, 5, 6] : survives after small-volume correction within the coordinate for each of the relevant contrasts from our original paper. The number of voxels in the cluster is not indicated here since we are using voxel-based small-volume correction.

[1] : survives after small-volume correction within a 10-mm sphere centered on the peak co-ordinate from this same contrast in this region from our original Lee et al., 2014 study (-30, 20, -2) (Lee et al., 2014)

[2] : survives after small-volume correction within a 10-mm sphere centered on the peak co-ordinate from this same contrast in this region from our original Lee et al., 2014 study (-9, 2, -8) (Lee et al., 2014)

[3] : survives after small-volume correction within a 10-mm sphere centered coordinate (9, 5, -8) (Lee et al., 2014)

[4] : survives after small-volume correction within a 10-mm sphere centered coordinate (-54, 38, 3) (Lee et al., 2014)

[5] : survives after small-volume correction within a 10-mm sphere centered coordinate (-27, -4, 1) (Lee et al., 2014)

[6] : within a 10-mm sphere centered on the peak co-ordinate from this same contrast in this region from our original Lee et al., 2014 study (-9, 29, -11) (Lee et al., 2014)

lPFC : lateral prefrontal cortex, MFG : medial frontal gyrus, ilPFC : inferior lateral prefrontal cortex, FPC : frontopolar cortex, SMA : supplementary motor area, STG : superior temporal gyrus, vmPFC : ventromedial prefrontal cortex

Supplementary Methods

GLM design. A general linear model (GLM) was used to generate voxelwise statistical parametric maps (SPMs) from the fMRI data. We created subject-specific design matrices containing the following regressors:

(R1) regressors encoding the average BOLD response at two choice states and one outcome states, (R2,R3) two parametric regressors encoding the model-derived prediction error signals – state prediction error (SPE) of MB and reward prediction error (RPE) of MF, (R4) a regressor encoding the average BOLD response at the start of each choice state (the time of presentation of the values of each token in the first stage and the time of the state presentation in the second stage), (R5) a parametric regressor encoding the goal change; it is a binary variable indicating whether the type of a coin associated with the largest value is different from the one in the previous trial. (R6) a parametric regressor encoding max or separate reliability of MB and MF, (R7) a parametric regressor encoding complexity, (R8) a parametric regressor encoding complexity x max reliability, (R8, R9) two parametric regressors encoding the chosen value of the model-free and the model-based system, respectively (QMF and QMB), (R10) and one parametric regressor encoding the chosen minus the unchosen value , a weighted sum of the QMB and QMF values according to the output of the arbitration system (QArb). For value signals of the arbitration output, we also in a separate model tested for the effects of both the chosen values alone instead of the effect of chosen minus unchosen value, but as found previously in our 2014 paper, we found that the chosen minus unchosen value signal showed a more robust effect in vmPFC, hence we used chosen vs unchosen value for the arbitration value signal in our main fMRI analysis. For each GLM run at the single subject level, orthogonalization of the regressors was disabled. Finally, we implemented a second-level random effects analysis for each regressor of interest, and applied correction for multiple comparisons. Our primary means of correction was small volume correction using 10mm spheres centered on the co-ordinates for the relevant computational signals from our 2014 study (Lee et al., 2014), given we had strong a-priori hypotheses about the location of each of the computational variables based on our original study. However, we also reported if the clusters survived more stringent correction at the whole brain level, cluster corrected at $p<0.05$ FWE (extent threshold at $p<0.001$), or the more stringent again whole brain voxel-level correction at $p<0.05$ FWE.

Reference

Lee, S.W., Shimojo, S., and O'Doherty, J.P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. Neuron *81*, 687–699.