Supplemental Material

Supplemental Text 1

<u>*Model comparison and model characteristics*</u>

We compared the dynamic learning rate model described in Section 2.5 to a model that allowed for a highly dynamic learning rate (β fixed to .01) and a model with a highly stable learning rate (β fixed to 10), the latter similar to a standard Q-learning model by fitting these models to trial-wise participant choices. -LLH was updated trial-wise by the log of the probability of the observed choice (calculated via a standard softmax function), and best-fitting parameters were identified using fmincon in Matlab v8.0.0.783 (Mathworks, Natick, MA, USA).

The Aikike Information Criterion (AIC), a well-established index of model fit which penalizes for model complexity (Vrieze, 2012), was used to guide model selection. Bayesian Model Selection revealed that our dynamic learning rate was highly favored in both samples, according to the protected exceedance probability (HV: 100%; SZ: 99%) (Rigoux, Stephan, Friston, & Daunizeau, 2014). Model fit (LLH change) was associated with total amount of reversals (Spearman's rho=.59, p<.001), but not with post-reversal perseveration (Spearman's rho=.17, p=.26). Moreover, motivational deficit severity did not correlate with the difference in fit (LLH) between the dynamic and stable learning rate model (Spearman's rho=.10, p=.63), suggesting that motivational deficit severity did not influence the improvement in fit in the dynamic versus stable learning rate model. The predictive probability of the dynamic learning rate model was >.33 in all but one PSZ.

Overall, both groups showed a highly dynamic learning rate: variation in learning rate (SD) was 13% (SD=7.01) in HV and 15.35% (SD=7.19) in PSZ. As expected, the β parameter (which captures the effect of RPE slope m on learning rate α) correlated positively with greater average learning rate (Spearman's rho=.78, $p<0.001$) and decreased variation in the learning rate (Spearman's rho=-.78, $p<0.001$). As discussed in Section 2.4, this confirms that the magnitude of the β parameter directly relates to learning rate dynamics. Inverse temperature (γ) was not associated with LR magnitude (Spearman's rho=.19, p=.23) or variation (Spearman's rho=-.09, p=0.54). However, γ did correlate with the number of reversals achieved across the entire sample (Pearson's r=.72, $p<.0001$), suggesting that participants with a more deterministic choice function performed better on the task.

Although β did not directly correlate with the number of reversals achieved (Spearman's rho=-0.03, $p=0.83$), the product of the average learning rate and the standard deviation of the learning rate (which are both strongly influenced by β) correlated with the number of reversals in an inverted-u fashion (Pearson's r after linear transformation=-0.40, $p=0.009$). Intuitively, this correlation can be understood as better performance for participants without a highly stable/dynamic learning rate and/or very low/high learning rate, which both would negatively affect performance. These analyses suggest that trial-by-trial estimates of learning rate are closely related to actual performance, which motivated our choice to focus on this part of our computational model for further analyses in Section 3.2.

Posterior predictions

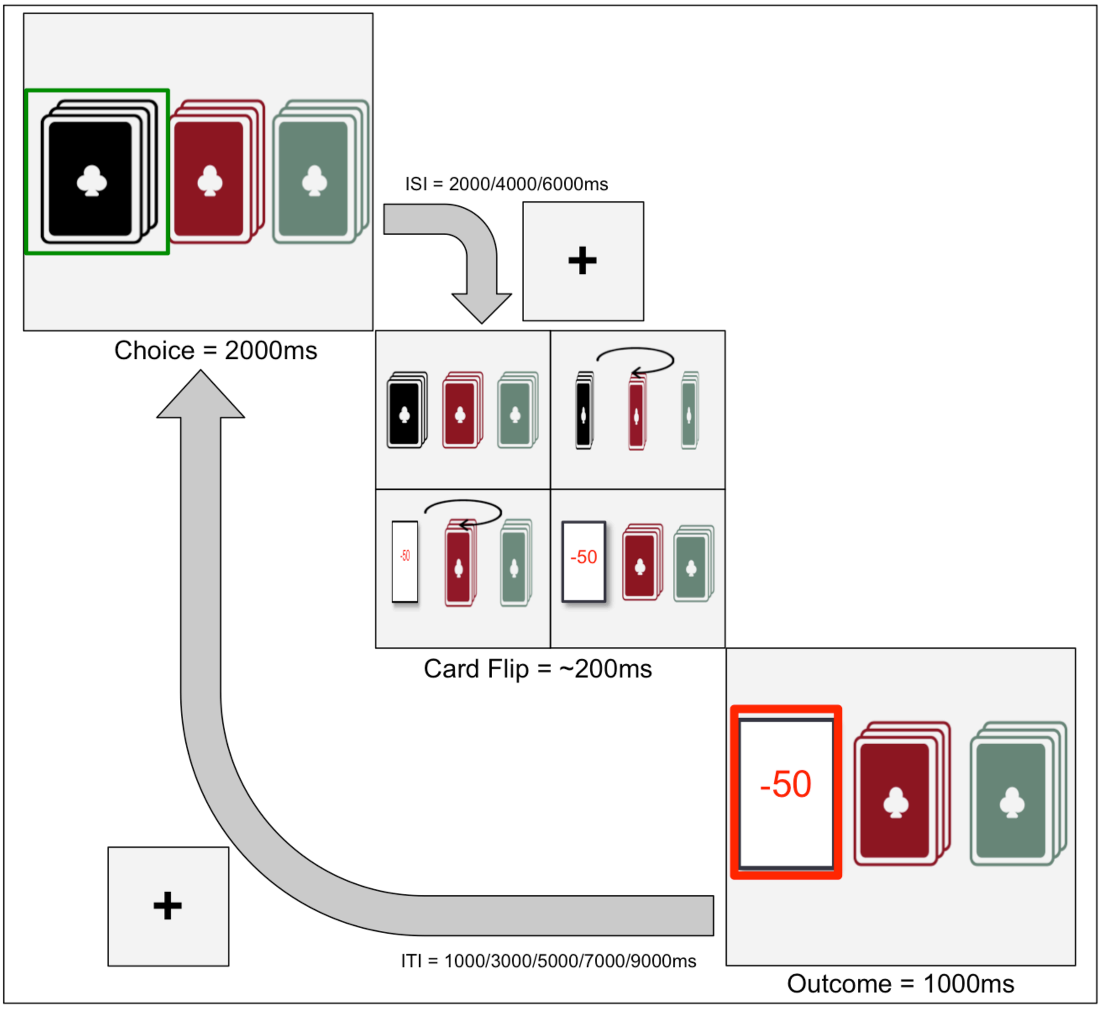
We additionally simulated data using individual parameters to ensure that the winning model could also account for performance in both groups. Because of the correlation between inverse temperature and performance, we fixed γ to 1, thereby eliminating inter-individual differences in the stochasticity of the choice function, thus highlighting the effect of the β parameter on simulated performance. A challenge with simulating reversal learning data is the reproduction of a sequence of correct answers, rather than simply recovering average performance, which is typically done in probabilistic learning tasks. Even if simulations favor the optimal deck, the probabilistic nature of simulations can lead to an underestimation of reversals. In order to highlight the effect of different β values (which control learning rate dynamics) on performance, we simulated data with a more deterministic (greedy) choice function, meaning that the deck with the highest expected value was more deterministically chosen in simulations. The average number of predicted reversals by the winning model was highly similar to the actual data (HV: M=6.74, SD=2.18, SZ: M=5.77, SD=1.48; see Section 3.2 for the actual data). Moreover, the number of actual reversals correlated significantly with simulated reversals in the entire sample (Spearman's rho=0.31, p=0.03). This correlation was trend-significant in the model with a highly dynamic learning rate (Spearman's rho =0.28, p=0.06) and absent in the model with a highly stable learning rate (Spearman's rho=.12, p=.42). Simulated data, as well as the effect of different β values on simulated performance, are reported for 8 typical participants (4HV, 4PSZ) in Supplemental Figure 2. All in all, model comparison and posterior predictions favored the dynamic learning rate model described in Section 2.4.

References

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. *Neuroimage, 84*, 971-985. doi: 10.1016/j.neuroimage.2013.08.065

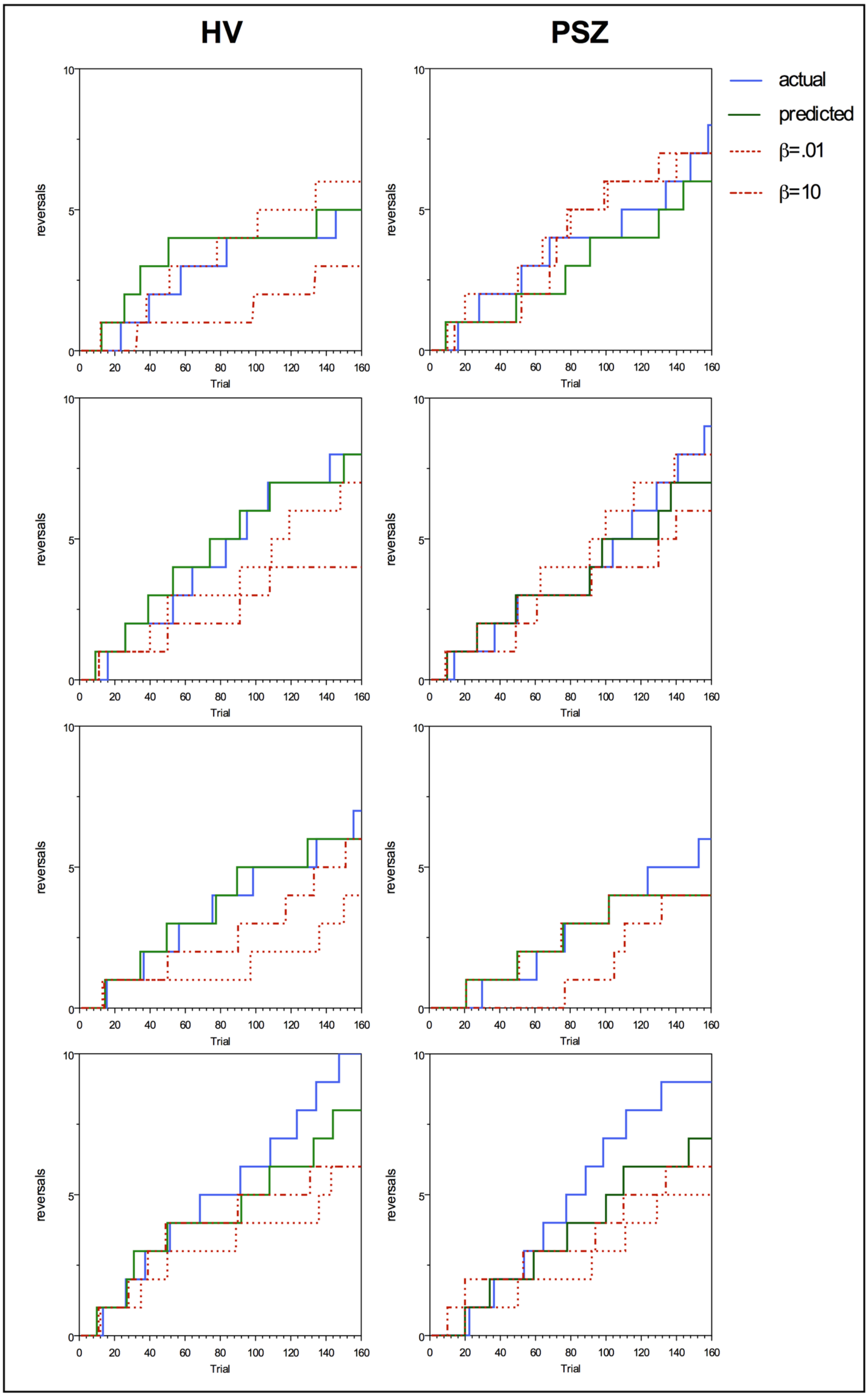
Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods, 17*(2), 228-243. doi: 10.1037/a0027127

Supplemental Figure 1. Reinforcement Learning task

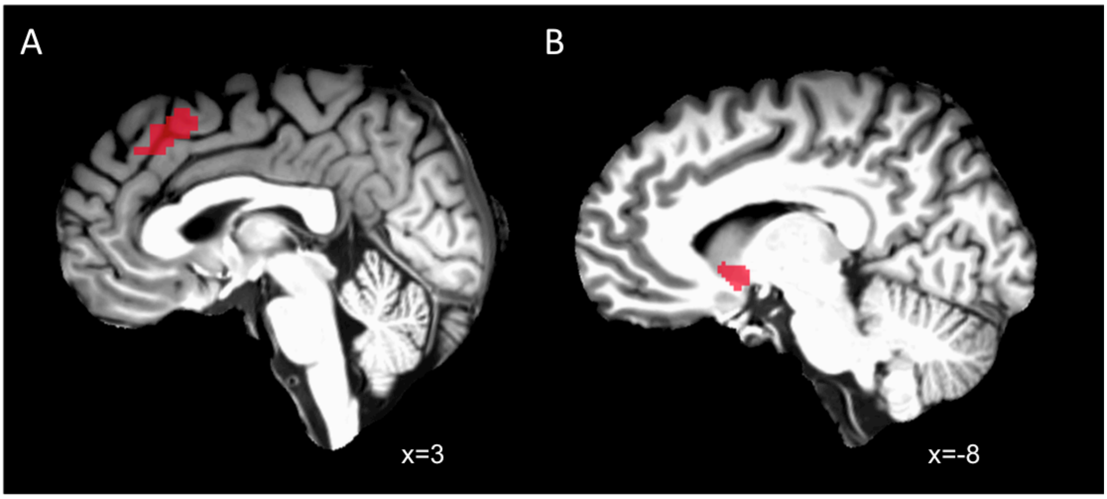


Graphical overview of the RL task. The RL task consisted of a choice, card-flip and outcome phase. Inter-trial (ITI) and inter-stimulus (ISI) intervals were pseudo-randomized. Numbers below each phase refer to event duration.

Supplemental Figure 2. Model simulations and the effect of β on simulated reversal learning in eight typical participants

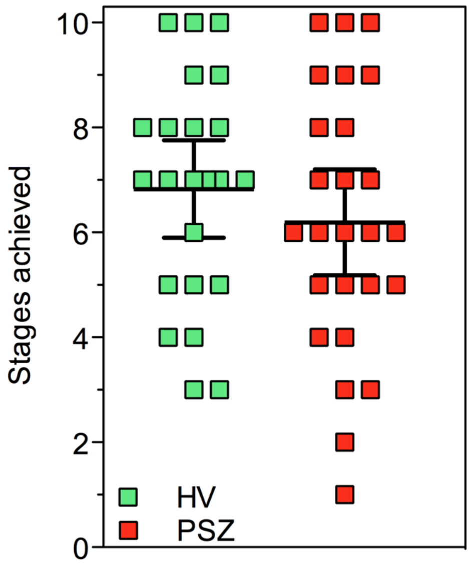


Supplemental Figure 3. Regions of Interest



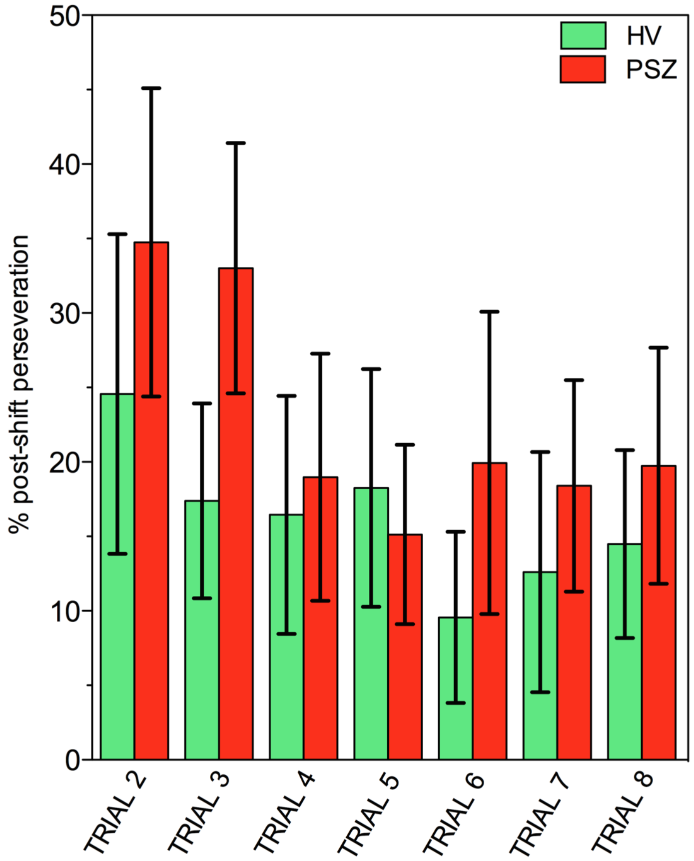
Depiction of the two ROIs that were selected for fMRI analyses; the dorso-medial prefrontal cortex (4A) and ventral striatum (4B). Origin, peak voxel and volume of the ROIs are reported in the main text.

Supplemental Figure 4. Reinforcement Learning task performance



Total amount of sudden shifts in reinforcement contingencies ('stages') achieved per group (no significant group difference, p>.05). Bars represent 95% confidence intervals.

Supplemental Figure 5. Trial-wise post-shift perseveration in PSZ



Trial-by-trial post-shift perseveration for HV and PSZ (based on the first 5 stages). On almost all trials post-reversal, PSZ showed an increased tendency to choose the previous-best card deck. Bars represent 95% confidence intervals.

Supplemental Tables

Supplemental Table 1 Antipsychotic Medication Type and Haloperidol Equivalents

Participant	AP1	AP2	AP3	Haloperidol Equiv.
SZ1	QUETIAPINE	-	-	21.97
SZ2	CLOZAPINE	-	-	19.21
SZ3	CLOZAPINE	-	-	12.59
SZ4	ZIPRASIDONE	THIOTHIXENE	-	5.94
SZ5	CLOZAPINE	-	-	3.98
SZ6	PALIPERIDONE	-	-	24.20
SZ7	CLOZAPINE			3.98
SZ8	CLOZAPINE	-	-	9.51
SZ9	OLANZAPINE	-	-	7.70
SZ10	QUETIAPINE	QUETIAPINE	HALOPERIDOL	13.57
SZ11	RISPERIDONE	-	-	19.74
SZ12	RISPERIDONE	ARIPIPRAZOLE	-	8.45
SZ13	RISPERIDONE	-	-	6.73
SZ14	CLOZAPINE	-	-	9.51
SZ15	OLANZAPINE	-	-	3.98
SZ16	CLOZAPINE	-	-	6.63
SZ17	FLUPHENAZINE	FLUPHENAZINE	-	9.97
SZ18	ARIPIPRAZOLE	-	-	5.47
SZ19	RISPERIDONE	-	-	2.98
SZ20	CLOZAPINE	-	-	6.63
SZ21	OLANZAPINE	-	-	18.22
SZ22	FLUPHENAZINE	QUETIAPINE	-	9.25
SZ23	CLOZAPINE	-	-	12.59
SZ24	PALIPERIDONE	-	-	25.34
SZ25	QUETIAPINE	PALIPERIDONE	-	26.52
SZ26	CLOZAPINE	-	-	9.51

Supplemental Table 2 Performance measures using all trials

Performance measure	HV (n=23)	PSZ (n=26)	t	*p*
Stages achieved	6.82 (2.15)	6.19 (2.50)	0.95	0.35
Total earnings (in $)	8.47 (0.93)	8.10 (1.12)	1.22	0.24
Switching (%)	32.66 (13.29)	35.72 (14.48)	0.77	0.45
Win-Stay (%)	60.57 (11.25)	57.72 (13.17)	0.81	0.42
Lose-Shift (%)	23.07 (6.16)	24.59 (6.02)	0.87	0.39

Supplemental Table 3 Individual Parameters and Model Fit indices

Participant	α(1)	β	γ	AIC change	LLH change
HV1	0.84	0.01	0.92	103.52	54.76
HV2	0.90	10	0.97	158.63	82.32
HV3	0.60	0.06	1.85	180.63	93.31
HV4	0.73	0.88	1.44	167.59	86.80
HV5	0.74	1.00	2.06	210.38	108.19
HV6	0.90	4.60	3.66	260.21	133.10
HV7	0.78	1.05	1.95	214.12	110.06
HV8	0.90	10	1.72	216.83	111.42
HV9	0.36	1.69	2.68	242.44	124.22
HV10	0.81	.086	1.80	207.09	106.55
HV11	0.79	10	2.70	258.80	132.40
HV12	0.03	1.88	0.84	69.57	37.78
HV13	0.28	2.00	1.16	166.19	86.10
HV14	0.90	10	1.71	216.17	111.09
HV15	0.90	10	1.94	212.41	109.21
HV16	0.90	6.38	0.59	95.88	50.94
HV17	0.87	0.79	2.22	223.33	114.67
HV18	0.39	0.01	2.07	225.70	115.85
HV19	0.43	1.69	1.86	192.90	99.45
HV20	0.35	1.59	1.16	179.52	92.76
HV21	0.58	1.04	1.44	165.37	85.68
HV22	0.81	1.18	1.77	204.48	105.24
HV23	0.90	10	1.99	224.12	115.06
SZ1	0.12	0.87	1.30	143.80	74.90
SZ2	0.65	1.00	1.73	213.66	109.83
SZ3	0.46	0.69	1.37	138.87	72.43
SZ4	0.90	0.99	1.71	191.21	98.60
SZ5	0.82	0.94	2.19	208.74	107.37
SZ6	0.86	10	1.71	209.63	107.82
SZ7	0.52	1.22	1.89	219.9	112.95
SZ8	0.45	0.20	0.15	48.07	27.03
SZ9	0.34	1.44	0.88	92.60	49.30
SZ10	0.90	10	1.61	210.63	108.32
SZ11	0.90	10	1.36	214.53	110.26
SZ12	0.54	0.13	2.51	208.26	107.13

SZ13	0.73	1.05	1.58	182.48	94.24
SZ14	0.81	0.92	2.08	201.60	103.80
SZ15	0.71	0.95	1.80	195.20	100.60
SZ16	0.90	10	1.57	190.75	98.37
SZ17	0.73	10	1.12	168.18	87.09
SZ18	0.30	0.15	1.90	189.86	97.93
SZ19	0.87	10	1.47	187.68	96.84
SZ20	0.90	10	0.95	108.14	57.07
SZ21	0.82	0.09	0.98	109.21	57.61
SZ22	0.21	0.01	1.39	170.07	88.04
SZ23	0.90	10	1.57	189.43	97.71
SZ24	0.34	2.74	1.34	142.76	74.38
SZ25	0.90	0.05	1.96	186.80	96.40
SZ26	0.13	2.77	0.75	100.36	53.18