Reinforcement learning performance in at-risk youth

James Waltz[a], Caroline Demro[b], Jason Schiffman[b], Elizabeth Thompson[b], Emily Kline[b], Gloria Reeves[c], Ziye Xu[a], James Gold[a]

[a]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, P.O. Box 21247, Baltimore, MD 21228

[b]Department of Psychology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD, USA, 21250

[c]Division of Child and Adolescent Psychiatry, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, 701 W. Pratt Street, Baltimore, MD, US, 21201

List of Supplementary Materials

Supplementary Methods

Computational model

The gain-loss Q-learning model (Frank et al 2007) is implemented by using separate learning rates α_G and $\alpha_L$for positive and negative feedback. For each trial t, the expected value Q for the chosen stimulus i is updated as a function of learning rate (α_G or α_L) and actual outcome $r(t)$ such that,

$$\text{(1) if } r(t) > 0, \qquad Q_i(t+1) = Q_i(t) + \alpha_G \cdot [r(t) - Q_i(t)]$$

$$\text{(2) if } r(t) = 0, \qquad Q_i(t+1) = Q_i(t) + \alpha_L \cdot [r(t) - Q_i(t)] \text{ ,}$$

where $Q_i(t)$is initialized to 0 and r(t) can be 1 for positive feedback and 0 for negative feedback. According to softmax distribution, the probability of choosing stimulus i over other stimuli in trial t is defined as

$$\text{(3) } P_i\ (t) = \frac{exp[\beta \cdot Q_i(t)]}{\sum_{k=1}^{n} exp[\beta \cdot Q_k(t)]}$$

where n is the total number of choices in each trial and β is the inverse temperature parameter, which reflects the randomness of responses.

The model described above was fit to each participant's training data and a set of best fit learning parameters (α_G, α_L, and β) was found by optimizing log-likelihood estimate (LLE), where LLE is defined as the log product of all probabilities,

$$LLE = log[\prod_t P_i(t)] \text{ .}$$

The model comparison was evaluated by Akaike's Information Criterion (AIC), which estimates the complexity of a given model used to represent the data:

$$AIC = -2 \cdot LLE + 2 \cdot k$$

where k is the number of parameters and the model with the lowest AIC is the best fitting model. The AIC for random model AIC_0 was also calculated using LLE_0, which is the likelihood for random performance. If AIC > AIC_0, the model is penalized for additional parameters and therefore the behavior of subject cannot be represented by the model. Individuals whose data satisfied the fitting criterion ($AIC < AIC_0$) were said to demonstrate "better-than-chance" behavior (because a model of purely random choices was not the best-fitting model). Individuals whose data did *not* satisfy the fitting criterion were said to demonstrate "at-chance" behavior.

To provide another measure of fit, we calculated pseudo-R^2 values, defined as

$$(\mathrm{LLE} - \mathrm{r})/\mathrm{r},$$

where r is the log likelihood of the data under a model of purely random choices ($p = 0.5$ for all choices; Daw et al., 2006). The resulting pseudo-R^2 statistics reveal how well the model fits the data compared to a model predicting chance performance, and is independent of the number of trials to be fit in each set.

We performed second-level statistical analyses on individual estimates of learning parameters (α_G, α_L, and β). In addition, we performed second-level statistical analyses on the products of the inverse temperature parameter and learning rates ($\beta*\alpha_G$ and $\beta*\alpha_L$), because, when viewed separately, inverse temperature and learning rate can have large estimation error and therefore a large standard deviation across subjects. The product of the two parameters, however, is generally more stable. Furthermore, the product of the two parameters is of theoretical interest, in that choice preference, at time t [$P_i(t)$, shown in Equation 3], is a function of the product of inverse temperature and learning rate ($\beta*\alpha$), as $Q_i(t)$ is a function of α. We performed two sets of analyses: 1) Mann-Whitney U-tests, comparing individuals demonstrating better-than-chance behavior and individuals showing at-chance behavior; and 2) Spearman correlation analyses between individual estimates of learning parameters (as well as $\beta*\alpha_G$ and $\beta*\alpha_L$) and clinical symptoms.

Following the estimation of parameters for each individual participant, we performed simulations using each individual's best fitting parameters. The simulation of each subject's performance was generated by the number of correct responses in each condition in 100 iterations. In order to capture performance during the Early Acquisition Phase (the first two blocks, which were completed by all participants), only the first 120 trials of the simulation were included. The 40 trials in each reinforcement probability condition were divided into 10 bins of 4 trials each, and the mean proportion of optimal responses for each bin across subjects/iterations was plotted for each reinforcement probability condition. Actual and simulated behavioral data are shown for both the entire sample of 70 subjects (Supplementary Figure 1A) and the sample of 57 subjects showing above-chance performance (Supplementary Figure 1B).

Supplementary Results

Computational modeling results

Individual parameter estimates are shown in Supplementary Table 4. The fitting criterion (AIC < AIC_0) was satisfied by 57 participants (showing better-than-chance behavior), and 13 participants were said to exhibit at-chance behavior. Supplementary Table 5 shows mean parameter values across subjects in each of the two groups. The fact that mean values for Temperature, (AIC - AIC_0), and Pseudo-R^2 were close to zero reflects the random nature of the choices in the group of participants showing at-chance behavior. The experimental behavioral data in Supplementary Table 6 (all three measures close to 50%) also clearly demonstrate the random nature of the choices in the group of participants showing at-chance behavior. The fact that mean values for Temperature, (AIC - AIC_0), and Pseudo-R^2 were significantly different from zero reflects the *systematic* nature of the choices in the group of participants showing better-than-chance behavior. In these participants, values for $\beta^*\alpha_G$ and $\beta^*\alpha_L$ were interpretable and indicative of the impact of gains and losses on learning and subsequent choices. Importantly, we observed significant correlations between $\beta^*\alpha_G$, a measure of gain-driven learning, and the severity of negative symptoms and deficits in social function (Supplementary Table 7).

Simulation results

Simulated performance, using mean parameter estimates from computational modeling, is shown in Supplementary Figure 1.

Supplementary Table 1. Partial Spearman correlations between experimental and clinical variables, controlling for IQ.

Measure	Early Acquisition Average	Win-stay Rate	Lose-shift Rate
Positive Sx	-0.225	-0.347*	0.018
Negative Sx	-0.188	-0.240^{+}	-0.141
Disorganization Sx	-0.075	-0.194	-0.026
Global Sx	0.001	0.004	0.109
Global Func: Role	0.191	0.183	0.188
Global Func: Social	0.182	0.293*	0.164

Abbreviations: Sx, Symptoms; Func, Functioning. ** = p < 0.05; + = p < 0.10.

Supplementary Table 2. Comparison of non-psychotic participants taking antipsychotic medications and not, on clinical and experimental variables.

	No APD (N=49)		APD (N=12)			
Measure	Mean	(SD)	Mean	(SD)	Statistic	p
Age	15.5	3.1	17.3	3.2	-1.728[a]	0.089
WASI Estimated IQ	103.6	16.5	104.9	18.5	-0.197[a]	0.844
Positive Sx	6.2	4.8	8.0	5.8	-0.947[b]	0.344
Negative Sx	9.1	6.0	8.1	6.1	-0.427[b]	0.669
Disorganization Sx	3.8	2.7	5.2	2.9	-1.836[b]	0.066
Global Sx	7.3	3.9	9.1	4.9	-1.157[b]	0.247
Global Func: Role	7.0	1.6	6.4	2.4	-0.358[b]	0.721
Global Func: Social	6.8	1.5	6.3	1.4	-1.219[b]	0.223
Early Acquisition Avg	67.1	18.6	64.9	14.7	-0.118[b]	0.906
Win-stay Rate	75.4	19.4	69.9	16.7	-1.179[b]	0.238
Lose-shift Rate	61.2	16.2	65.2	19.4	-0.354[b]	0.723

Notes: [a] = t-test; [b] = z of Mann-Whitney U-test. Abbreviations: APD, Antipsychotic Drug; WASI, Wechsler Abbreviated Scale of Intelligence; Sx, Symptoms; Func, Functioning; Avg, Average.

Supplementary Table 3. Comparison of non-psychotic participants taking stimulant medications and not, on clinical and experimental variables.

	No Stimulant (N=32)		Stimulant (N=29)			
Measure	Mean	SD	Mean	SD	Statistic	p
Age	16.4	3.3	15.2	3.0	1.463[a]	0.149
WASI Estimated IQ	105.4	17.5	101.8	15.7	0.687[a]	0.496
Positive Sx	6.1	5.1	7.0	5.0	-0.862[b]	0.389
Negative Sx	8.6	6.1	9.2	5.9	-0.666[b]	0.506
Disorganization Sx	**3.5**	**2.7**	**4.7**	**2.7**	**-2.097[b]**	**0.036**
Global Sx	7.7	4.0	7.6	4.3	-0.348[b]	0.728
Global Func: Role	7.0	1.9	6.7	1.6	-0.959[b]	0.338
Global Func: Social	6.9	1.6	6.5	1.5	-0.780[b]	0.435
Early Acquisition Avg	63.8	18.5	69.7	16.7	-1.661[b]	0.097
Win-stay Rate	71.0	20.1	78.0	17.1	-1.249[b]	0.212
Lose-shift Rate	59.3	15.2	65.1	18.1	-1.170[b]	0.242

Notes: [a] = t-test; [b] = z of Mann-Whitney U-test. Significant between-group differences in **boldface**. Abbreviations: WASI, Wechsler Abbreviated Scale of Intelligence; Sx, Symptoms; Func, Functioning; Avg, Average.

Supplementary Table 4. Estimates of model parameters in individual participants.

Subject	beta_alphaG	beta_alphaL	temperature	alphaG	alphaL	LLE	LLE0	LLE - LLE0	AIC - AIC0	Chance Perf
1	0.03100	0.00000	0.09766	0.31746	0.00000	-31.06313	-249.53299	218.46985	-430.93970	0
2	0.03596	0.01360	0.06764	0.53161	0.20098	-68.74916	-249.53299	180.78383	-355.56765	0
3	0.00491	0.00000	1.00000	0.00491	0.00000	-107.66202	-249.53299	141.87097	-277.74193	0
4	0.00165	0.00000	1.00000	0.00165	0.00000	-167.86577	-249.53299	81.66722	-157.33443	0
5	0.15273	0.02681	0.15273	1.00000	0.17556	-10.28473	-83.17766	72.89294	-139.78587	0
6	0.09824	0.00901	0.09824	1.00000	0.09174	-11.66206	-83.17766	71.51560	-137.03119	0
7	0.03191	0.03259	1.00000	0.03191	0.03259	-14.84110	-83.17766	68.33656	-130.67313	0
8	0.02007	0.02404	1.00000	0.02007	0.02404	-16.74873	-83.17766	66.42893	-126.85786	0
9	0.04938	0.00186	0.08321	0.59351	0.02237	-19.68238	-83.17766	63.49528	-120.99056	0
10	0.01336	0.00992	1.00000	0.01336	0.00992	-21.91322	-83.17766	61.26444	-116.52889	0
11	0.06425	0.00141	0.06425	1.00000	0.02193	-190.37135	-249.53299	59.16164	-112.32328	0
12	0.02102	0.02710	0.23844	0.08817	0.11368	-25.27694	-83.17766	57.90072	-109.80144	0
13	0.01209	0.00207	0.12528	0.09649	0.01656	-26.42915	-83.17766	56.74851	-107.49702	0
14	0.01318	0.01370	1.00000	0.01318	0.01370	-26.48458	-83.17766	56.69308	-107.38616	0
15	0.01398	0.00236	0.06783	0.20609	0.03483	-70.48799	-124.76649	54.27850	-102.55701	0
16	0.00790	0.00355	1.00000	0.00790	0.00355	-30.50929	-83.17766	52.66837	-99.33674	0
17	0.02872	0.00115	0.08028	0.35774	0.01430	-32.96333	-83.17766	50.21433	-94.42866	0
18	0.01194	0.02767	0.21656	0.05514	0.12777	-36.51286	-83.17766	46.66480	-87.32960	0
19	0.03235	0.00530	0.06813	0.47480	0.07781	-39.23459	-83.17766	43.94308	-81.88615	0
20	0.00586	0.00310	1.00000	0.00586	0.00310	-40.45708	-83.17766	42.72058	-79.44115	0
21	0.03081	0.00148	0.05283	0.58323	0.02794	-41.13962	-83.17766	42.03804	-78.07608	0
22	0.01126	0.00115	0.10224	0.11012	0.01129	-41.71209	-83.17766	41.46557	-76.93114	0
23	0.00633	0.00527	1.00000	0.00633	0.00527	-41.99201	-83.17766	41.18565	-76.37130	0
24	0.02913	0.00275	0.06674	0.43646	0.04123	-43.47519	-83.17766	39.70247	-73.40494	0
25	0.02656	0.00497	0.05164	0.51429	0.09629	-43.91726	-83.17766	39.26040	-72.52080	0
26	0.00558	0.00474	1.00000	0.00558	0.00474	-45.88057	-83.17766	37.29709	-68.59418	0
27	0.00981	0.00438	0.07555	0.12983	0.05803	-46.06106	-83.17766	37.11660	-68.23321	0
28	0.00296	0.00000	0.04726	0.06270	0.00000	-214.80007	-249.53299	34.73292	-63.46583	0
29	0.04395	0.00000	0.04395	1.00000	0.00000	-216.40721	-249.53299	33.12578	-60.25156	0
30	0.00306	0.00408	0.02042	0.14988	0.19985	-218.09766	-249.53299	31.43533	-56.87065	0
31	0.00428	0.00317	0.21228	0.02017	0.01493	-52.38399	-83.17766	30.79367	-55.58735	0
32	0.01998	0.00145	0.04562	0.43809	0.03171	-52.39404	-83.17766	30.78363	-55.56725	0
33	0.01402	0.00000	0.04371	0.32070	0.00000	-138.24792	-166.35532	28.10740	-50.21480	0
34	0.00151	0.00000	1.00000	0.00151	0.00000	-142.65074	-166.35532	23.70458	-41.40917	0
35	0.00399	0.00152	1.00000	0.00399	0.00152	-60.83258	-83.17766	22.34508	-38.69017	0
36	0.01361	0.01587	0.02748	0.49540	0.57736	-61.64879	-83.17766	21.52887	-37.05774	0
37	0.00941	0.00773	0.00941	1.00000	0.82171	-231.56199	-249.53299	17.97100	-29.94199	0
38	0.00069	0.00000	1.00000	0.00069	0.00000	-231.92696	-249.53299	17.60602	-29.21204	0
39	0.00060	0.00000	0.07733	0.00775	0.00000	-236.58622	-249.53299	12.94677	-19.89354	0
40	0.00156	0.00000	0.03290	0.04737	0.00000	-237.12065	-249.53299	12.41234	-18.82467	0
41	0.01578	0.00002	0.02865	0.55085	0.00059	-239.94910	-249.53299	9.58388	-13.16777	0
42	0.00701	0.00701	0.00701	1.00000	1.00000	-199.44503	-208.63730	9.19228	-12.38455	0
43	0.00232	0.00309	0.05917	0.03923	0.05222	-74.18936	-83.17766	8.98830	-11.97660	0
44	0.01633	0.00516	0.01633	1.00000	0.31598	-74.73123	-83.17766	8.44643	-10.89287	0
45	0.00448	0.02005	0.02376	0.18856	0.84364	-75.44906	-83.17766	7.72860	-9.45720	0
46	0.00049	0.00451	1.00000	0.00049	0.00451	-76.53101	-83.17766	6.64665	-7.29331	0
47	0.00000	0.00213	0.03326	0.00000	0.06405	-243.85918	-249.53299	5.67380	-5.34760	0
48	0.00042	0.01356	0.03876	0.01087	0.34979	-244.27735	-249.53299	5.25564	-4.51128	0
49	0.02598	0.00000	0.02598	1.00000	0.00000	-244.67310	-249.53299	4.85989	-3.71977	0
50	0.00040	0.00000	1.00000	0.00040	0.00000	-244.84539	-249.53299	4.68760	-3.37520	0
51	0.00109	0.01322	0.01626	0.06683	0.81294	-244.88836	-249.53299	4.64463	-3.28925	0
52	0.00026	0.01255	0.03068	0.00852	0.40904	-244.91993	-249.53299	4.61305	-3.22611	0
53	0.00081	0.00571	1.00000	0.00081	0.00571	-78.95957	-83.17766	4.21809	-2.43618	0
54	0.00790	0.00368	0.01097	0.72057	0.33572	-95.73079	-99.81319	4.08240	-2.16480	0
55	0.02121	0.00000	0.02214	0.95814	0.00000	-204.27560	-207.94415	3.66855	-1.33711	0
56	0.01494	0.00011	0.02126	0.70240	0.00531	-245.96075	-249.53299	3.57224	-1.14448	0
57	0.00112	0.00073	0.00870	0.12901	0.08419	-163.32848	-166.35532	3.02684	-0.05369	0
58	0.00000	0.00296	0.02297	0.00000	0.12864	-246.95425	-249.53299	2.57874	0.84253	1
59	0.00000	0.03218	0.03218	0.00000	1.00000	-248.07747	-249.53299	1.45552	3.08897	1
60	0.00000	0.01293	0.02086	0.00000	0.62018	-248.30206	-249.53299	1.23093	3.53815	1
61	0.00000	0.00474	0.01852	0.00000	0.25572	-248.31709	-249.53299	1.21589	3.56821	1
62	0.00208	0.00208	0.00208	1.00000	1.00000	-248.36577	-249.53299	1.16722	3.66557	1
63	0.00919	0.00000	0.00919	1.00000	0.00000	-248.73581	-249.53299	0.79718	4.40564	1
64	0.00000	0.00027	0.74477	0.00000	0.00037	-165.60787	-166.35532	0.74745	4.50510	1
65	0.00175	0.00121	0.00175	1.00000	0.68944	-248.88734	-249.53299	0.64564	4.70871	1
66	0.00004	0.02198	0.02198	0.00179	1.00000	-165.82959	-166.35532	0.52574	4.94853	1
67	0.00000	0.00715	0.00753	0.00000	0.94941	-207.47507	-207.94415	0.46909	5.06182	1
68	0.00000	0.00012	0.28990	0.00000	0.00041	-249.11422	-249.53299	0.41877	5.16246	1
69	0.00105	0.00057	0.00105	1.00000	0.53981	-249.35697	-249.53299	0.17602	5.64797	1
70	0.00000	0.00267	0.00267	0.00000	1.00000	-249.49963	-249.53299	0.03336	5.93328	1

Abbreviations: beta_alphaG, inverse temperature * learning rate for gains; beta_alphaL, inverse temperature * learning rate for losses; alphaG, learning rate for gains; alphaL, learning rate for losses; LLE, final log likelihood estimate; LLE0, initial log likelihood estimate; LLE - LLE0, change in log likelihood estimate, as a consequence of fitting; AIC - AIC0, change in Akaike's Information Criterion, as a consequence of fitting; Chance Perf, subject was deemed to exhibit chance performance.

Supplementary Table 5. Mean estimates of model parameters in participants showing better-than-chance behavior and in participants showing at-chance behavior.

	Better Than Chance (N=57)		At Chance (N=13)			
Measure	Mean	SD	Mean	SD	z of U	p
Temperature (β)	**0.327**	**0.427**	**0.090**	**0.211**	**-3.097**	**0.002**
α_G	0.308	0.365	0.308	0.480	-1.124	0.261
α_L	**0.126**	**0.238**	**0.553**	**0.425**	**-2.036**	**0.042**
$\beta*\alpha_G$	**0.018**	**0.025**	**0.001**	**0.003**	**-3.888**	**<0.001**
$\beta*\alpha_L$	0.006	0.008	0.007	0.010	-0.705	0.481
AIC - AIC_0	**-72.673**	**81.917**	**4.237**	**1.333**	**-4.968**	**<0.001**
Pseudo-R^2	**0.337**	**0.288**	**0.004**	**0.001**	**-4.968**	**<0.001**

Abbreviations: α_G, Learning rate for gains (Go-learning); α_L, Learning rate for losses (NoGo-learning); AIC = Akaike's Information Criterion. (AIC - AIC_0) and Pseudo-R^2 are both measures of model fit. Significant between-group differences are **bolded**.

Supplementary Table 6. Comparison of participants showing better-than-chance behavior and not, on clinical and experimental variables.

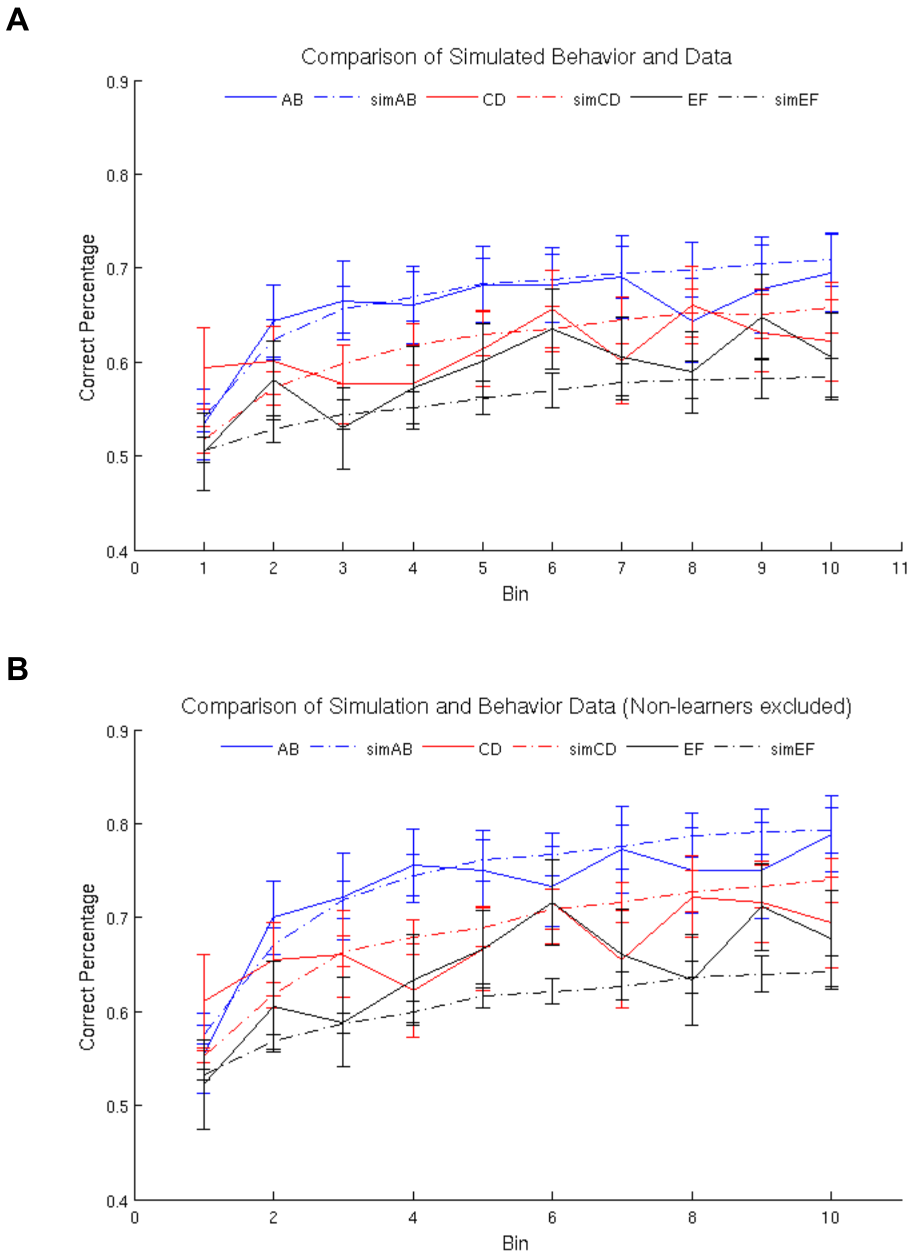
	Better Than Chance (N=57)		At Chance (N=13)			
Measure	Mean	SD	Mean	SD	Statistic	p
Age	15.9	3.2	15.9	2.8	0.027[a]	0.978
WASI Estimated IQ	**107.6**	**15.8**	**91.5**	**10.5**	**3.169[a]**	**0.003**
Positive Sx	7.4	5.7	10.3	7.6	-0.498[b]	0.618
Negative Sx	9.1	6.3	13.8	7.1	-1.318[b]	0.188
Disorganization Sx	4.5	3.1	5.5	4.7	-0.641[b]	0.522
Global Sx	7.8	4.1	8.4	4.8	-0.323[b]	0.747
Global Func: Role	6.8	1.8	6.1	2.0	-0.424[b]	0.671
Global Func: Social	6.7	1.5	6.0	1.7	-0.903[b]	0.367
Early Acquisition Avg	**69.9**	**16.8**	**48.6**	**5.7**	**-3.683[b]**	**<0.001**
Win-stay Rate	**78.0**	**17.0**	**48.7**	**12.6**	**-3.916[b]**	**<0.001**
Lose-shift Rate	**64.7**	**16.1**	**51.4**	**10.2**	**-2.650[b]**	**0.008**

Notes: [a] = t-test; [b] = z of Mann-Whitney U-test. Abbreviations: WASI, Wechsler Abbreviated Scale of Intelligence; Sx, Symptoms; Func, Functioning; Avg, Average. Significant between-group differences are **bolded**.

Supplementary Table 7. Spearman correlations between clinical variables and modeling parameters from individual participants.

	Temperature (β)	α_G	α_L	$\beta*\alpha_G$	$\beta*\alpha_L$
Positive Sx	-0.163	0.099	0.133	-0.096	-0.001
Negative Sx	-0.072	-0.077	-0.001	-0.270*	-0.023
Disorganization Sx	-0.203	0.124	0.050	-0.102	-0.085
Global Sx	0.084	-0.057	0.141	-0.098	0.141
Global Func: Role	0.113	-0.023	0.025	0.145	0.049
Global Func: Social	0.110	0.162	0.055	0.282*	0.119

Abbreviations: α_G, Learning rate for gains (Go-learning); α_L, Learning rate for losses (NoGo-learning); Sx, Symptoms; Func, Functioning. * = $p < 0.05$.



Supplementary Figure 1. Simulations of participant behavior across first two Acquisition blocks (first 120 Acquisition trials) using parameters estimated from computational modeling. (A) Comparison of actual and simulated performance of all 70 participants, including those exhibiting at-chance behavior, on all three Acquisition pairs (AB = 80%/20%; CD = 70%/30%; EF = 60%/40%). Actual performance plotted using solid lines; simulated performance plotted using dashed lines. “Correct Percentage” = proportion of choices of more-frequently-rewarded stimulus. Each Bin represents 12 trials. (B) Comparison of actual and simulated performance of 57 participants showing better-than-chance performance (excluding those exhibiting at-chance behavior), on all three Acquisition pairs.

References

Akaike, H (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* 19(6):716-723.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A* 104(41):16311-6. Epub 2007 Oct 3. PubMed PMID: 17913879; PubMed Central PMCID: PMC2042203.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441(7095):876-879.