PNAS

Supporting Information for

The consequences of AI training on human decision making

Lauren S. Treiman, Chien-Ju Ho, Wouter Kool

Lauren S. Treiman
Email: ltreiman@wustl.edu

This PDF file includes:

Supporting Text
Tables S1 to S15

Supporting Information Text

Experiment 4 Session 2 Exploratory Analysis. In Experiment 4, we ran a pre-registered mixed-effects model to determine whether people previously assigned to the AI training condition rejected more unfair offers than those previously assigned to the control condition. This mixed-effects model found neither a main effect of training condition nor an interaction effect between offer amount and training condition, demonstrating that there was no difference in acceptance rates between these two conditions (see main text for further description). These results differ from our findings in Experiments 1 and 2, which showed that people in the AI training condition continued to reject unfair offers more than those in the control condition even after AI training.

While it is possible that habit formation requires more salience than we used in this study (especially in the short experimental paradigms used here), visual inspection of Figure 5e, suggested that people previously in the AI training conditions were more punitive than those in the control condition for unfair offers. An exploratory follow-up analyses provided provisional evidence for this conjecture. We ran an exploratory mixed-effects model using the control condition as a reference level and conditioning on only *unfair* offers. The results of the mixed effects model once again revealed a main effect of offer amount ($b = 1.91,\ \mathrm{SE} = 0.18,\ p < 0.01$). Additionally, we found a main effect of training condition for the AI training for others condition ($b = -1.72,\ \mathrm{SE} = 0.81, p = 0.034$) but not for the AI training self condition ($b = -1.00,\ \mathrm{SE} = 0.85,\ p = 0.24$). Interestingly, the mixed effects model indicated a three-way interaction between partner type, training condition for the AI training self condition, and offer amount ($b = 0.40,\ \mathrm{SE} = 0.20,\ p = 0.048$). To investigate this interaction, we ran two more exploratory mixed effects models conditioning on each partner type. When only considering human partners, we found main effects of offer amount ($b = 2.08,\ \mathrm{SE} = 0.26,\ p < 0.001$) and training condition for only the AI training for others condition ($b = -2.13,\ \mathrm{SE} = 0.98,\ p = 0.03$). When conditioning on AI partners, we once again found main effects of offer amount ($b = 1.64,\ \mathrm{SE} = 0.27,\ p < 0.001$) and training condition for AI training for others ($b = -1.76,\ \mathrm{SE} = 0.89,\ p = 0.048$). Interestingly, we found an interaction effect between offer amount and training condition for the AI training for self condition ($b = 0.92,\ \mathrm{SE} = 0.44,\ p = 0.034$). Specifically, the difference in acceptance rates between the AI training for self and control conditions decreased as the offer amount became fairer. There were no additional significant effects in any of the three exploratory mixed effects models ($ps = 0.24$).

Table S1. ANOVA Results: Experiment 1 Session 1

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*
fairness	69.620	1	69.620	592.091	< .001
fairness ✻ training condition	1.055	1	1.055	8.969	0.003
Residuals	24.810	211	0.118		
opponent	0.034	1	0.034	1.825	0.178
opponent ✻ training condition	8.329×10^{-4}	1	8.329×10^{-4}	0.045	0.832
Residuals	3.896	211	0.018		
fairness ✻ opponent	0.003	1	0.003	0.405	0.525
fairness ✻ opponent ✻ training condition	9.117×10^{-5}	1	9.117×10^{-5}	0.011	0.915
Residuals	1.677	211	0.008		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*
training condition	0.763	1	0.763	4.232	0.041
Residuals	38.061	211	0.180		

Note. Type III Sum of Squares

Table S2. Mixed-Effects Regression Results: Experiment 1 Session 1

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	1.802	0.201	< .001 ***
opponent AI	-0.106	0.054	0.051 .
offer amount	1.874	0.059	<.001 ***
AI training	-0.368	0.199	0.064 .
opponent AI ✻ AI training	-0.048	0.039	0.224
opponent AI ✻ offer amount	0.021	0.054	0.703
offer amount ✻ AI training	0.170	0.054	0.002 **
opponent AI ✻ offer amount ✻ AI training	0.015	0.039	0.698

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S3. ANOVA Results: Experiment 2 Session 1

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*
fairness	142.232	1	142.232	1562.250	< .001
fairness ✻ training condition	2.081	2	1.040	11.428	< .001
Residuals	30.135	331	0.091		
opponent	1.451×10^{-4}	1	1.451×10^{-4}	0.011	0.915
opponent ✻ training condition	0.033	2	0.016	1.299	0.274
Residuals	4.196	331	0.013		
opponent ✻ fairness	0.022	1	0.022	2.009	0.157
opponent ✻ fairness ✻ training condition	0.025	2	0.012	1.166	0.313
Residuals	3.548	331	0.011		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*
training condition	4.795	2	2.397	17.325	< .001
Residuals	45.800	331	0.138		

Note. Type III Sum of Squares

Table S4. Mixed-Effects Regression Results: Experiment 2 Session 1

Reference level: control condition

Effect	Estimate	SE	*p*	
Fixed Effects				
Intercept	2.890	0.333	< .001	***
opponent AI	0.066	0.099	0.505	
offer amount	2.250	0.105	<.001	***
AI training for self	-1.918	0.432	<.001	***
AI training for others	-1.763	0.449	<.001	***
opponent AI ✻ AI training for self	-0.153	0.124	0.215	
opponent AI ✻ AI training for others	-0.109	0.131	0.406	
opponent AI ✻ offer amount	-0.057	0.070	0.417	
offer amount ✻ AI training for self	0.303	0.140	0.0304	*
offer amount ✻ AI training for others	0.545	0.155	<.001	***
opponent AI ✻ offer amount ✻ AI training for self	0.018	0.094	0.849	
opponent AI ✻ offer amount ✻ AI training others	-0.026	0.106	0.804	

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: AI training others condition

Effect	Estimate	SE	*p*	
Fixed Effects				
Intercept	1.128	0.305	< .001	***
opponent AI	-0.043	0.085	0.617	
offer amount	2.795	0.123	<.001	***
AI training for self	-0.155	0.411	0.706	
control condition	1.763	0.449	<.001	***
opponent AI ✻ AI training for self	-0.045	0.113	0.691	
opponent AI ✻ control condition	0.109	0.131	0.406	
opponent AI ✻ offer amount	-0.083	0.079	0.295	
offer amount ✻ AI training for self	-0.243	0.152	0.110	
offer amount ✻ control condition	-0.545	0.155	<.001	***
opponent AI ✻ offer amount ✻ AI training for self	0.044	0.101	0.663	
opponent AI ✻ offer amount ✻ control condition	0.026	0.106	0.804	

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S5. ANOVA Results: Experiment 3

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	*p*
fairness	66.152	1	66.152	463.542	< .001
fairness ✻ training condition	0.292	1	0.292	2.044	0.154
Residuals	30.826	216	0.143		
opponent	4.697×10^{-5}	1	4.697×10^{-5}	0.006	0.940
opponent ✻ training condition	0.046	1	0.046	5.503	0.020
Residuals	1.787	216	0.008		
fairness ✻ opponent	0.005	1	0.005	0.626	0.430
fairness ✻ opponent ✻ training condition	0.034	1	0.034	4.400	0.037
Residuals	1.685	216	0.008		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
training condition	0.718	1	0.718	3.682	0.056
Residuals	42.102	216	0.195		

Note. Type III Sum of Squares

Table S6. Exploratory ANOVA Results: Experiment 3

Conditioned on AI partners

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
fairness	32.510	1	32.510	444.495	< .001
fairness ✻ training condition	0.263	1	0.263	3.597	0.059
Residuals	15.798	216	0.073		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
training condition	0.562	1	0.562	5.428	0.021
Residuals	22.378	216	0.104		

Note. Type III Sum of Squares

Conditioned on Human partners

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
fairness	33.647	1	33.647	434.881	< .001
fairness ✻ training condition	0.063	1	0.063	0.814	0.368
Residuals	16.712	216	0.077		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
Training condition	0.201	1	0.201	2.016	0.157
Residuals	21.512	216	0.100		

Note. Type III Sum of Squares

Table S7. Mixed-Effects Regression Results: Experiment 3

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	2.294	0.247	< .001 ***
opponent AI	-0.022	0.057	0.699
offer amount	1.988	0.067	<.001 ***
AI training	-0.590	0.241	0.014 *
opponent AI ✻ AI training	-0.074	0.057	0.199
opponent AI ✻ offer amount	-0.059	0.043	0.172
offer amount ✻ AI training	-0.187	0.062	0.003 **
opponent AI ✻ offer amount ✻ AI training	0.059	0.043	0.168

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S8. ANOVA Results: Experiment 4 Session 1

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
fairness	112.461	1	112.461	834.178	< .001
fairness ✻ training condition	1.312	2	0.656	4.866	0.008
Residuals	46.107	342	0.135		
opponent	0.004	1	0.004	0.386	0.535
opponent ✻ training condition	0.005	2	0.003	0.242	0.785
Residuals	3.679	342	0.011		
fairness ✻ opponent	5.204×10^{-5}	1	5.204×10^{-5}	0.006	0.940
fairness ✻ opponent ✻ training condition	0.011	2	0.006	0.634	0.531
Residuals	3.093	342	0.009		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
training condition	2.504	2	1.252	6.147	0.002
Residuals	69.667	342	0.204		

Note. Type III Sum of Squares

Table S9. Mixed-Effects Regression Results: Experiment 4 Session 1

Reference level: control condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	2.545	0.312	< .001 ***
opponent AI	-0.011	0.078	0.885
offer amount	1.900	0.084	<.001 ***
AI training for self	-1.167	0.448	0.009 **
AI training for others	-1.454	0.420	<.001 ***
opponent AI ✻ AI training for self	-0.056	0.111	0.615
opponent AI ✻ AI training for others	0.006	0.102	0.952
opponent AI ✻ offer amount	-0.021	0.053	0.692
offer amount ✻ AI training for self	0.138	0.117	0.239
offer amount ✻ AI training for others	-0.056	0.105	0.596
opponent AI ✻ offer amount ✻ AI training for self	0.007	0.080	0.929
opponent AI ✻ offer amount ✻ AI training others	0.034	0.072	0.632

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: AI training others condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	1.091	0.284	< .001 ***
opponent AI	-0.005	0.064	0.934
offer amount	1.844	0.068	<.001 ***
AI training for self	0.287	0.430	0.505
control condition	1.454	0.420	<.001 ***
opponent AI ✻ AI training for self	-0.062	0.101	0.540
opponent AI ✻ control condition	-0.006	0.102	0.952
opponent AI ✻ offer amount	0.013	0.048	0.783
offer amount ✻ AI training for self	0.193	0.108	0.075 .
offer amount ✻ control condition	0.056	0.105	0.596
opponent AI ✻ offer amount ✻ AI training for self	-0.027	0.077	0.722
opponent AI ✻ offer amount ✻ control condition	-0.034	0.072	0.632

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S10. ANOVA Results: Experiment 5

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
fairness	62.549	1	62.549	467.284	< .001
fairness ✻ training condition	0.013	1	0.013	0.098	0.754
Residuals	27.574	206	0.134		
opponent	0.007	1	0.007	0.645	0.423
opponent ✻ training condition	0.017	1	0.017	1.543	0.216
Residuals	2.329	206	0.011		
fairness ✻ opponent	0.003	1	0.003	0.330	0.566
fairness ✻ opponent ✻ training condition	0.012	1	0.012	1.427	0.234
Residuals	1.756	206	0.009		

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
training condition	0.064	1	0.064	0.318	0.574
Residuals	41.457	206	0.201		

Note. Type III Sum of Squares

Table S11. Mixed-Effects Regression Results: Experiment 5

Pre-registered mixed-effects regression results:

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	1.961	0.235	< .001 ***
opponent AI	-0.066	0.055	0.233
offer amount	1.864	0.062	<.001 ***
AI training	-0.284	0.230	0.218
opponent AI ✻ AI training	-0.015	0.055	0.791
opponent AI ✻ offer amount	-0.071	0.041	0.080 .
offer amount ✻ AI training	-0.188	0.057	0.001 **
opponent AI ✻ offer amount ✻ AI training	0.095	0.041	0.019 *

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Exploratory mixed effects results: conditioned on AI training condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	1.560	0.258	< .001 ***
opponent AI	-0.077	0.071	0.28
offer amount	1.616	0.070	<.001 ***
opponent AI ✻ offer amount	0.0216	0.051	0.67

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Exploratory mixed effects results: conditioned on control condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	2.433	0.411	< .001 ***
opponent AI	-0.053	0.086	0.538
offer amount	2.156	0.107	<.001 ***
opponent AI ✻ offer amount	-0.182	0.065	0.0053 **

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S12. Mixed-Effects Regression Results: Experiment 1 Session 2

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	4.450	0.428	< .001 ***
opponent AI	-0.105	0.084	0.21
offer amount	2.681	0.122	<.001 ***
AI training	0.484	0.395	0.22
opponent AI ✻ AI training	-0.046	0.058	0.43
opponent AI ✻ offer amount	0.045	0.084	0.59
offer amount ✻ AI training	0.751	0.113	<.001 ***
opponent AI ✻ offer amount ✻ AI training	0.053	0.058	0.37

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S13. Mixed-Effects Regression Results: Experiment 2 Session 2

Reference level: control condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	5.082	0.512	< .001 ***
opponent AI	0.036	0.129	0.781
offer amount	1.979	0.113	<.001 ***
AI training for self	-0.991	0.658	0.132
AI training for others	-1.317	0.692	0.057 .
opponent AI ✻ AI training for self	0.029	0.166	0.861
opponent AI ✻ AI training for others	0.002	0.170	0.991
opponent AI ✻ offer amount	-0.132	0.079	0.093 .
offer amount ✻ AI training for self	0.766	0.174	<.001 ***
offer amount ✻ AI training for others	0.552	0.172	0.001 **
opponent AI ✻ offer amount ✻ AI training for self	0.004	0.108	0.968
opponent AI ✻ offer amount ✻ AI training others	0.135	0.111	0.226

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: AI training others

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	3.765	0.486	< .001 ***
opponent AI	0.038	0.111	0.733
offer amount	2.531	0.136	<.001 ***
AI training for self	0.326	0.645	0.614
control condition	1.316	0.694	0.058 .
opponent AI ✻ AI training for self	0.027	0.152	0.858
opponent AI ✻ control condition	-0.002	0.170	0.991
opponent AI ✻ offer amount	0.003	0.079	0.970
offer amount ✻ AI training for self	0.214	0.184	0.245
offer amount ✻ control condition	-0.552	0.172	0.001 **
opponent AI ✻ offer amount ✻ AI training for self	-0.131	0.109	0.229
opponent AI ✻ offer amount ✻ control condition	-0.135	0.111	0.226

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S14. Mixed-Effects Regression Results: Experiment 4 Session 2

Reference level: control condition

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	4.414	0.442	< .001 ***
opponent AI	0.104	0.098	0.289
offer amount	1.963	0.100	<.001 ***
AI training for self	-0.772	0.617	0.211
AI training for others	-1.079	0.585	0.065 .
opponent AI ✻ AI training for self	-0.118	0.140	0.398
opponent AI ✻ AI training for others	-0.077	0.130	0.554
opponent AI ✻ offer amount	0.002	0.066	0.975
offer amount ✻ AI training for self	-0.065	0.136	0.631
offer amount ✻ AI training for others	0.063	0.132	0.636
opponent AI ✻ offer amount ✻ AI training for self	-0.014	0.094	0.885
opponent AI ✻ offer amount ✻ AI training others	-0.100	0.089	0.265

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: AI training others

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	3.335	0.407	< .001 ***
opponent AI	0.027	0.086	0.755
offer amount	2.026	0.094	<.001 ***
AI training for self	0.307	0.597	0.607
control condition	1.079	0.585	0.065 .
opponent AI ✻ AI training for self	-0.042	0.132	0.753
opponent AI ✻ control condition	0.077	0.130	0.554
opponent AI ✻ offer amount	-0.098	0.060	0.106
offer amount ✻ AI training for self	-0.128	0.132	0.331
offer amount ✻ control condition	-0.063	0.132	0.636
opponent AI ✻ offer amount ✻ AI training for self	0.086	0.090	0.342
opponent AI ✻ offer amount ✻ control condition	0.100	0.089	0.265

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Table S15. Exploratory Mixed-Effects Regression Results: Experiment 4 Session 2

Reference level: control condition; conditioned on unfair offers

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	0.853	0.582	0.143
opponent AI	0.125	0.109	0.253
offer amount	1.908	0.180	<.001 ***
AI training for self	-0.998	0.854	0.243
AI training for others	-1.725	0.814	0.034 *
opponent AI ✻ AI training for self	-0.084	0.160	0.600
opponent AI ✻ AI training for others	0.065	0.152	0.668
opponent AI ✻ offer amount	-0.134	0.133	0.313
offer amount ✻ AI training for self	0.243	0.264	0.358
offer amount ✻ AI training for others	-0.142	0.241	0.556
opponent AI ✻ offer amount ✻ AI training for self	0.396	0.196	0.043 *
opponent AI ✻ offer amount ✻ AI training others	-0.132	0.185	0.477

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: control condition; conditioned on unfair offers and AI partners

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	0.827	0.625	0.186
offer amount	1.642	0.274	<.001 ***
AI training for self	-1.000	0.896	0.264
AI training for others	-1.765	0.891	0.048 *
offer amount ✻ AI training for self	0.924	0.436	0.034 *
offer amount ✻ AI training for others	-0.150	0.360	0.677

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

Reference level: control condition; conditioned on unfair offers and human partners

Effect	Estimate	SE	*p*
Fixed Effects			
Intercept	0.855	0.691	0.216
offer amount	2.076	0.264	<.001 ***
AI training for self	-1.034	1.008	0.305
AI training for others	-2.217	0.985	0.031 *
offer amount ✻ AI training for self	-0.017	0.369	0.963
offer amount ✻ AI training for others	0.209	0.371	0.573

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*