Supplementary Content for:

NEUROCOMPUTATIONAL MECHANISMS OF ADAPTIVE LEARNING IN SOCIAL EXCHANGES.

Vanyukov, P.M.,* Hallquist, M., Delgado, M., Szanto, K., & Dombrovski, A.Y.

*vanyukovpATupmc.edu

Contents:

Supplementary results: Computational modeling.

Parameter recovery and model identifiability.

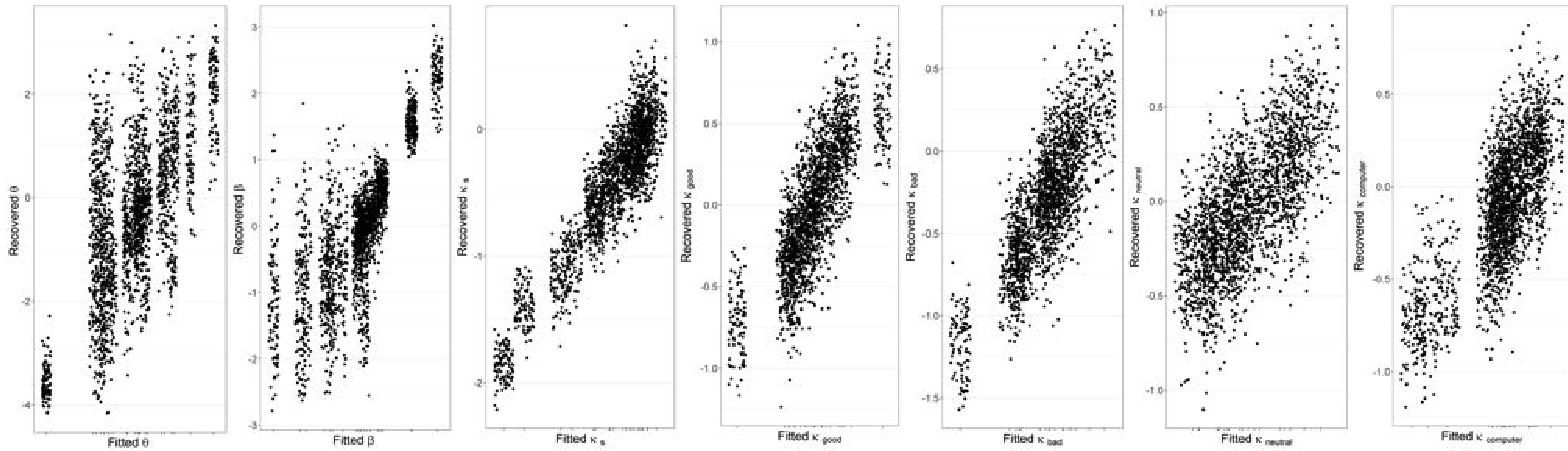
Each single-subject model fit provided a model estimate of value per each trial, which we then used as the probability of choice, in a MATLAB function *binornd*() to generate one hundred behaviors. This dataset of 100 × subjects (one generative dataset per each of the five models we considered in the manuscript) was then used to perform parameter recovery and model identifiability analyses. For parameter recovery, models were fit to their corresponding generative dataset of simulated subjects and correlations were computed between the model-generated posterior and the model-simulated parameter values. All models were successful in recovering their parameters (Table S1 and Figure S1).

Table S1. Correlations between posterior and recovered parameter values (all *p*s < .01).

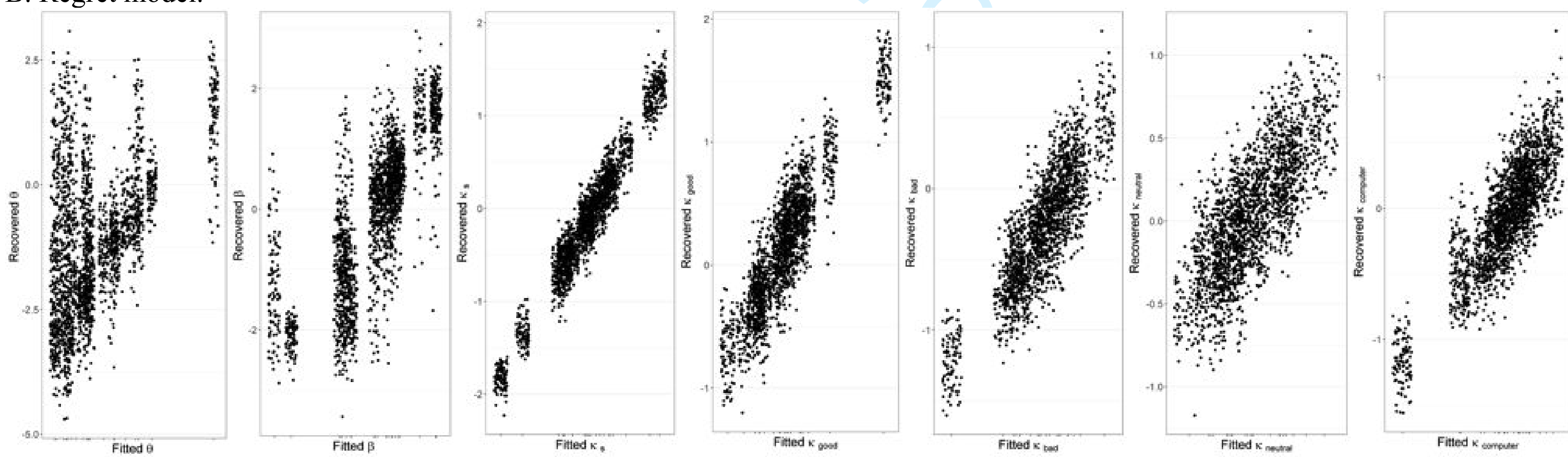
	θ	β	κ_s	κ_{good}	κ_{bad}	$\kappa_{neutral}$	$\kappa_{computer}$
Actual reward	0.66	0.78	0.93	0.80	0.83	0.72	0.76
Regret	0.49	0.77	0.98	0.91	0.86	0.79	0.85
Trustee-count.	0.49	0.63	0.98	0.90	0.85	0.73	0.85
Policy	0.62	0.73	0.95	0.77	0.78	0.66	0.74
	θ		*SVM parameter*		β		
Social Value Model	0.67		0.67		0.86		

Figure S1. Fitted and recovered parameter values.

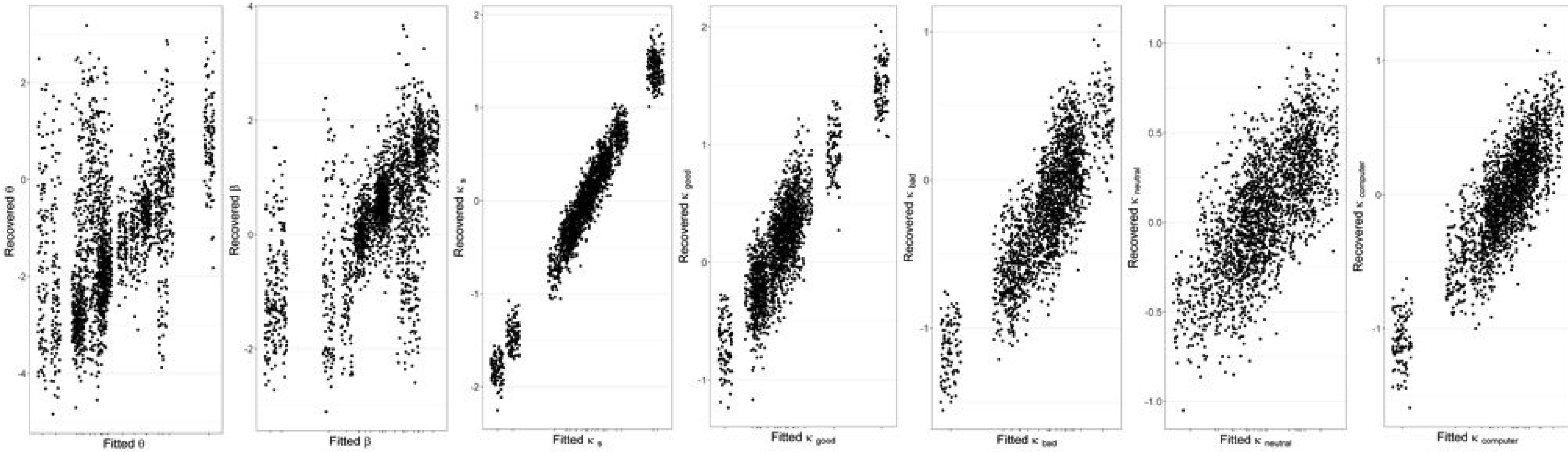
A. Actual rewards model.



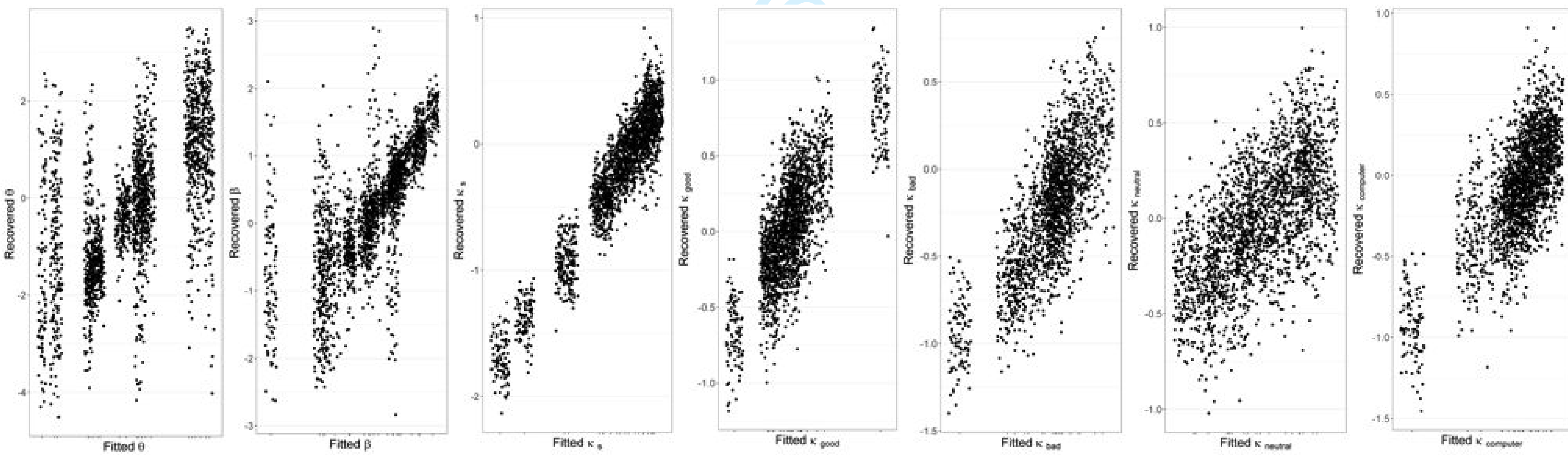
B. Regret model.



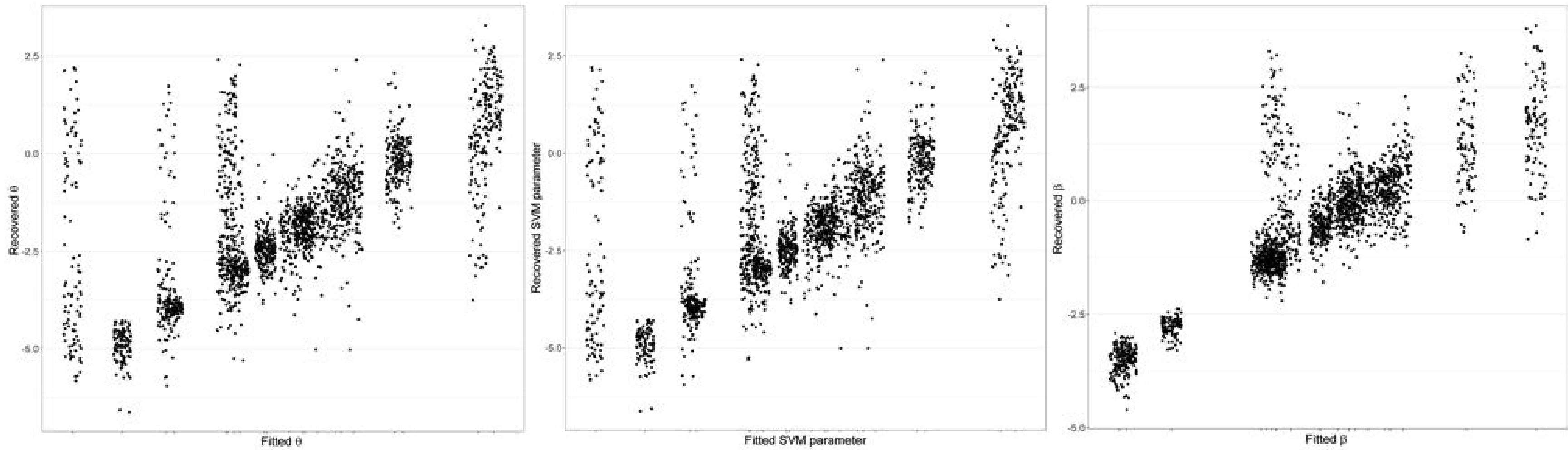
C. Trustee-counterfactual model.



D. Policy model.



E. Social Value model (SV).



For model identifiability analyses, each model was used to fit each generated dataset. Bayesian model comparisons provided evidence indicating that each model was uniquely identifiable, i.e., did not characterize data simulated from other models.

Single session and simple choice rule.

Initially we implemented computational models with a simple choice rule that employed only subject-level parameter, κ_s, to reflect a bias of the participant to keep or to share regardless of reinforcement. These models also attempted to fit all data at once, rather than assuming that certain parameters may be different between blocks and, in this way, account for block-level differences. Model comparison of all four alternatives indicated that the policy model dominated the alternatives, although the effect did not withstand the correction for Bayesian omnibus risk in the smaller, self-paced Study 1 (BOR; Study 1: Bayesian Omnibus Risk or BOR = 0.057; Study 2: BOR < 0.001; Study 3: BOR = 0.001).

Multisession and augmented choice rule. Overall differences in investment rates between trustee conditions, which paralleled participants' pre-experimental ratings of trustees strongly suggested that participants treated the blocked trustee conditions as separate sessions of the same task. Meaning, in order to describe condition-level differences, the models could be made to fit the data in a way that allows for certain parameters or initial states to vary between conditions. We implemented and contrasted the following alternatives for the choice rule: 1) single varying or condition-level bias parameter in the choice rule; 2) fixed subject-level with a varying or condition-level bias parameters in the choice rule. Results of the comparison between the two models (pooled over the three datasets) suggested that model augmented with subject-level and condition-level parameters was preferred (BOR = .470; ep = 1).

Other implementations of trustee bias.

The primary focus of this paper was to determine which representations of rewards likely drove the learning in the trust task, rather than to refine the specification of trustee type effects. However, it may still be important to note that still other implementations of trustee type on the investment rates are possible and would reflect different hypotheses regarding the mechanism of the condition-level effects. One can argue that an interesting and important distinction should be made between models that implement condition-level effects as part of instrumental vs. non-instrumental processes. For example, these effects could be thought of as direct influences on the calculation of action values, e.g., different initial values of the invest action or different learning rates for each trustee condition (e.g., Fouragnan et al., 2013), or on the action selection, e.g., different temperature or bias parameters in the choice function (as described above). Of course, some mixture of two or more condition-level parameters is also possible. Furthermore, the structure of the task, e.g., blocked conditions vs. mixed trials or a match between condition and reinforcement (Fouragnan et al., 2013) vs. lack thereof (present study) or fast vs. slow decision-making, as well as the degree to which condition information is perceived by participant as informative to reinforcement, may also influence whether condition-level effects will act on instrumental or non-instrumental processes. Exhaustive testing of all of these possibilities is outside of the paper's scope, however, a comparison of four implementations of trustee-level effects (initial state of action value, learning rate, temperature, bias; pooled over the three datasets) indicated that the model with condition-level bias parameter was preferred over the other three (BOR = 0.0056; ep = 1.00).

Reinforcement learning model comparison with the social value model.

The SV architecture is somewhat different from the models described in the present study. SV does not directly learn the expected value of the "share" decision, updating the probability of partner's reciprocation, $P(t)$, instead. *P(t)* in turn is additively modulated by the perceived trustworthiness or the social value term.

$P_c(t+1) = P_c(t) + \alpha \times [\gamma_c(t) - P_c(t)]$, where $P_c(t)$ is probability of the trustee *c*, sharing on trial *t*, α is the learning rate, and $\gamma_c(t)$ = 1 if the trustee shared on the current trial and 0 otherwise.

$EV_c(t) = P_c(t) \times \left[1.5 + \left(\Phi \times \frac{T_c}{\max(T_c)}\right)\right]$, where *EV* is the expected value of the share decision.

As in the original study, we used participants' pre-task ratings of trustees, T_c, in order to initialize the SV term, a subject-level constant scaled by a free parameter Φ. The model was implemented in the VBA framework. SV was developed for a task with fixed contingencies and no counterfactual feedback. For a fair comparison, we allowed it to learn the trustee's probability of sharing from both actual and counterfactual feedback. Comparison was equivocal (over three datasets pooled), although SV did numerically worse than the policy model (BOR = 0.654; ep = 0.984).

There were several experimental and analytical reasons for SV model may have been at a disadvantage in fitting the data from our sample.

1) Differences in experimental design between the present study and Fareri et al., 2015.

In the present study, participants received counterfactual feedback, i.e., participants received feedback on the partner's decision, whether participants kept/shared. This was not the case in Fareri et al., and, thus, participants would have been unable to learn from their "keep" decisions. In terms of the SV model, this meant that the $P_c(t)$ was only updated on the trials for which participant made a "share" decision. In addition, the *reputation* conditions were not blocked in the Fareri et al, which employed *friend, confederate,* and *computer* as trustees. Close friend condition could have induced higher cooperation rates from participants, whereas the blocked presentation of trustees in the present study allowed the effect of the reputation to decay faster for our participants. The manipulation of the reward schedule in the present study (shifts from 50% share rate to 25% or to 88%) may have made our participants more aware of the inconsistency between the reputation and the actual trustee behavior.

It should also be noted that the reputation effects may be expected to decay over the course of the interaction, but the SV model does not appear to reflect the decay. That is, the contribution of the social value to the expected value of the share decision does not change as a function of time, or the number of exchanges with the trustee.

2) Using different approaches in the estimation of parameters and model comparison.

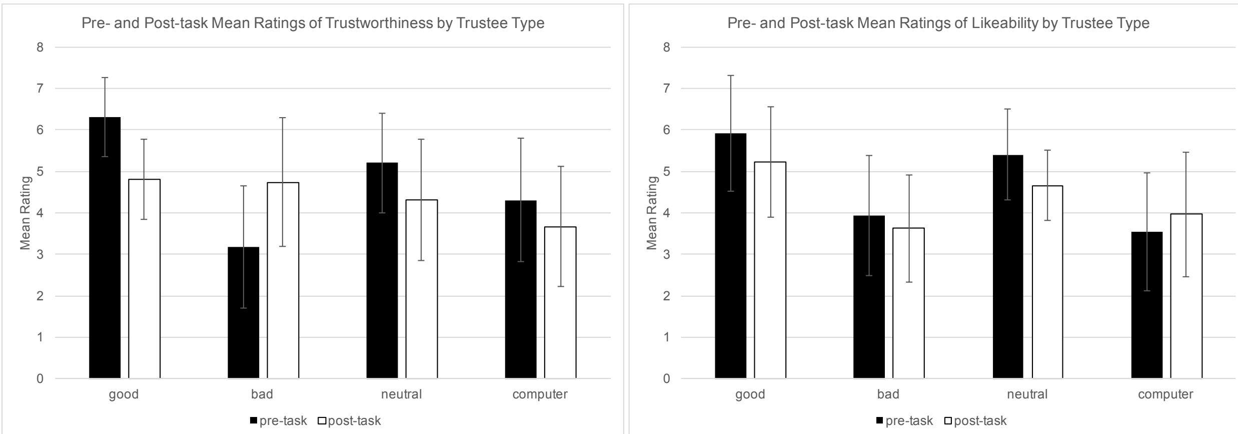
VBA modeling approach requires for parameters to behave in Gaussian fashion, so in describing the ranges of the relevant parameters (e.g., SV parameter), we took advantage of the logarithmic transform to bound the parameter between 0 and 1 and scale it by a factor of 5. Thus, although the ranges of the parameter values were equivalent to Fareri et al., 2015, the model-fitting procedure may have explored the parameter space differently.

Supplementary results: Survey ratings analyses.

Figure S2 (*Figure 3* in the manuscript) illustrates the trustworthiness and likeability ratings (Study 1 and 3 only), which were analyzed as in (Fareri, Chang & Delgado, 2012). Trustee (good, bad, neutral and computer) ratings were entered at both time points (pre-task and post-task) into a 4 × 2 repeated-measures ANOVA. Similarly to previous studies, ratings differed by trustee type [*trustworthiness:* $F(3, 102) = 19.60$, $p < .01$; *likeability:* $F(3, 102) = 26.56$, $p < .01$] and changed from before and after the experiment [*trustworthiness:* $F(3, 34) = 9.70$, $p < .01$; *likeability: :* $F(3, 34) = 7.92$, $p < .01$]. Post-hoc *t*-test comparisons with Bonferroni correction, indicated that on average participants rated "good" trustees as most trustworthy [$t(34)_{bad} = 7.04$, $t(34)_{neutral} = 3.88$, $t(34)_{computer} = 8.36$, all *p*s < .01] and more likeable than bad and computer trustees [$t(34)_{bad} = 6.86$, $t(34)_{computer} = 7.85$, all *p*s < .01], whereas bad trustees were rated least trustworthy and least likeable of human trustees [*trustworthiness:* $t(34)_{neutral} = 2.91$, $p < .05$; *likeability:* $t(34)_{neutral} = 4.90$, $p < .01$].

A significant interaction of trustee type × time for both measures (*trustworthiness:* $F[3, 102] = 20.50$, $p < .01$; *likeability:* $F[3, 102] = 4.69$, $p < .01$) suggested that the pattern of pre- and post-task ratings was different for the different trustees. Qualitatively, although on average all trustees were rated as less trustworthy over time, bad trustees may have been rated as more trustworthy post-task. Similarly, although on average all trustees were rated as less likeable over time, computer was rated as more likeable after the experiment.

Figure S2 (*Figure 3 in manuscript)*. Pre- and post-task mean ratings of trustworthiness and likeability by trustee type.



Supplementary results: Effects of averaged and recent reinforcement history.

In conventional model-free analyses, we attempted to model separately effects of averaged reinforcement history, indexed by the reward schedule (RS), and recent reinforcement history, indexed by the previous decision by trustee (PTD). Recall that all trustees began with a block of 16 trials during which their rate of return was at chance (50%). Thereafter, the order of "poor" (or 25% rate of trustee returning investment) and "rich" (88%) blocks was counterbalanced. In order to take full advantage of the counterbalanced design, the analyses reported below exclude the 50% block.

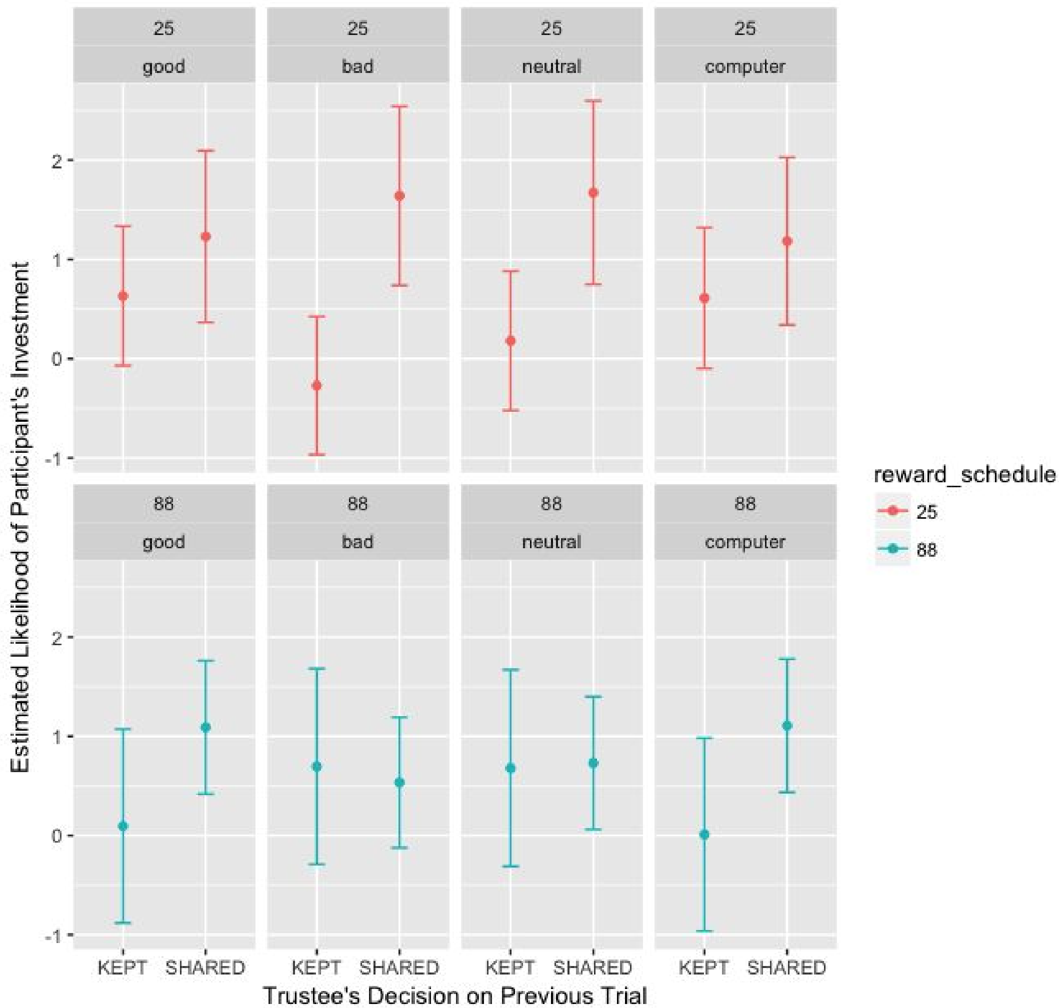
The random effects structure for the model with averaged reinforcement history included a by-subject random slope for previous trustee decision (to accommodate individual differences in learning), by-subject random intercept (to account for individual differences in cooperation rate), and a nested within-subject effect of the block order for a trustee.

A model with averaged reinforcement history explained significantly more variance in Study 1, but not in Studies 2 and 3 separately or combined (Study 1: $AIC_{w/out\ RS} = 2323.7$ vs. $AIC_{w/\ RS} = 2331.4$, χ^2 [1] = 29.69, p < .005), Study 2: $AIC_{w/out\ RS} = 3279.6$ vs. $AIC_{w/\ RS} = 3279.8$, χ^2 [1] = 1.836, *p* = .175; Study 3: $AIC_{w/out\ RS} = 3168.1$ vs. $AIC_{w/\ RS} = 3169.1$, χ^2 [1] = 0.991, *p* = .340; Study 2 and 3 combined: $AIC_{w/out\ RS} = 6439.5$ vs. $AIC_{w/\ RS} = 6438.7$, χ^2 [1] = 2.793, *p* = .095). As noted earlier, unlike the other studies, Study 1 was self-paced, outside of the scanner, and had a smaller sample size, which may account for the difference in findings. Therefore, we report below only the results of the model with averaged reinforcement history included for Study 1.

Once the averaged reinforcement history was added to the model deemed best by previous analyses, the main effects of trustee type and recent reinforcement history were no longer significant (trustee type: χ^2 [3] = 5.55, *p* = .126; PTD: χ^2 [1] = 0.15, p = .698). Nor was there a main effect of the reward schedule (RS: χ^2 [1] = 1.90, p = .169). Participants' likelihood to invest with the trustee decreased with the number of exchanges with trustee (number of exchanges: b = -0.02, SE = 0.010, $p < .05$).

As before, the influence of the trustee's previous action increased with the number of exchanges with a trustee (trustee type × PTD: b = 0.05, SE = 0.013, $p < .001$). There was a significant 2-way interaction of trustee type and reward schedule (χ^2 [3] = 11.45, $p < .01$), which was qualified by a 3-way interaction of trustee type × RS × PTD (χ^2 [3] = 18.26, $p < .001$). This latter interaction suggested that recent reinforcement history (PTD) was most influential when interacting with bad, neutral trustees during poor RS, and good trustee and computer during rich RS. A tentative interpretation for this finding is that these may have been instances of the most surprising and, therefore, salient outcomes for participants, e.g., bad trustee returning the investment during a poor rate of return or good trustee keeping the investment during a rich rate of return.

Figure S3. Estimated likelihood of participant's investment by reward schedule, trustee type and trustee's response on the previous trial.


25
good
25
bad
25
neutral
25
computer
88
good
88
bad
88
neutral
88
computer
2
1
0
-1
Estimated Likelihood of Participant's Investment
KEPT
SHARED
Trustee's Decision on Previous Trial
reward_schedule
25
88

Supplementary results: Main effect of task.

Task performance compared to fixations elicited a robust activation of visual regions implicated in recognition of facial stimuli, e.g., fusiform gyrus and middle temporal gyrus, attentional control, e.g., middle frontal gyrus, and learning, e.g., cingulate gyrus. These findings are described in Table S2 and Figure S4.

Figure S4. BOLD response to task performance ($p_{voxelwise}$ < .001, cluster size = 88 voxels, p_{corr} < .05).

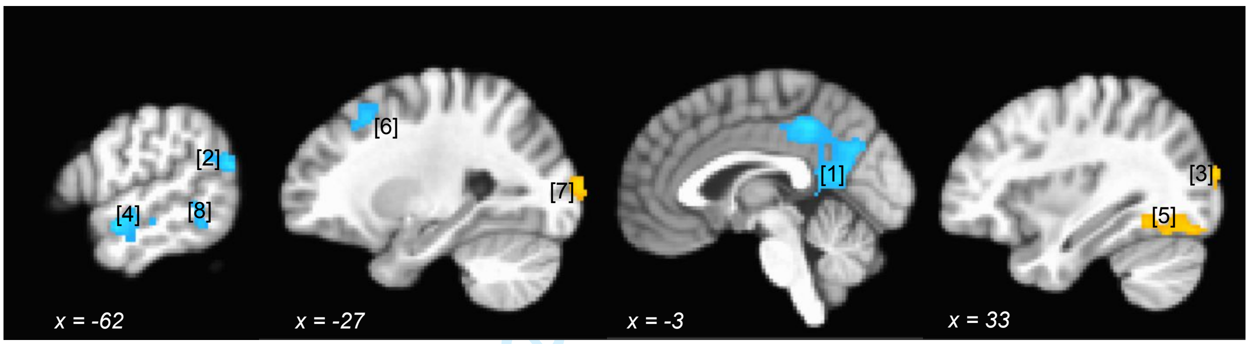


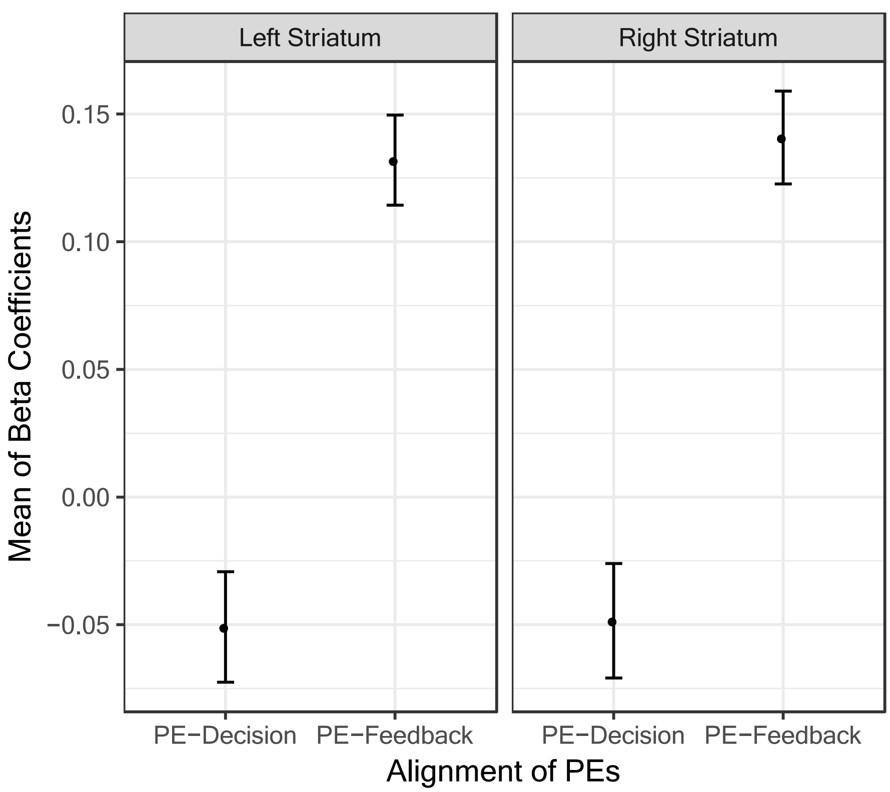
Table S2. BOLD response to task performance.

Region	MNI Coordinates	Peak $t(21)$	Cluster Size (mm^3)
[1] Left Cingulate Gyrus (BA 31)	-3, -50, 29	-5.10	19285
[2] Left Middle Temporal Gyrus (BA 39)	-50, -66, 27	-4.54	4429
[3] Right Middle Occipital Gyrus (BA 19)	22, -98, 12	4.78	3735
[4] Left Middle Temporal Gyrus	-55, -6, -13	-5.08	3711
[5] Right Fusiform Gyrus (BA 19)	33, -68, -13	4.34	3103
[6] Left Middle Frontal Gyrus (BA 8)	-27, 21, 51	-4.15	1436
[7] Left Middle Occipital Gyrus (BA 19)	-22, -102, 10	4.29	1375
[8] Left Middle Temporal Gyrus (BA 21)	-62, -49, -5	-3.93	1241

Supplementary results: Model-estimated prediction errors aligned to decision and feedback phases of the task.

In order to corroborate that the striatal response was indeed associated with the computation of prediction errors, i.e., at the feedback phase of the task, we also aligned the model-estimated PEs to the decision phase of task and included them in the neuroimaging model with the feedback-aligned PEs (Figure S5). From the group map generated by ***3dMEMA***, we extracted the β coefficients of the two regressors using the ventral striatal mask from a PE meta-analysis (i.e., same as in the manuscript: Chase et al., 2015) and the meta-analytic estimates of variance (τ^2). A significantly stronger positive striatal signal was associated with the PEs aligned at feedback (z = -26.81, p < .001).

Figure S5. Striatal signal associated with the prediction errors aligned at decision and feedback phases.



Supplementary results: Neuroimaging: Prediction Errors (PE) signals.

Figure S6. BOLD response to PEs derived from the actual rewards model ($p_{voxelwise} < .001$, cluster size = 100 voxels, $p_{corr} < .05$).

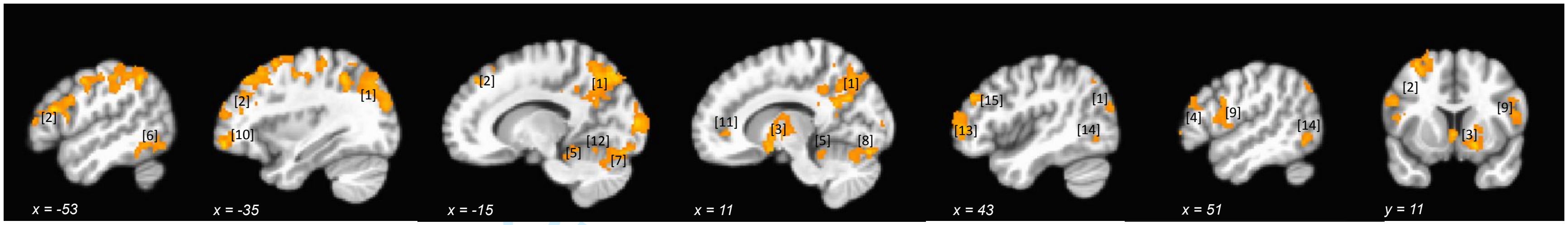


Table S3. BOLD response to prediction errors derived from the actual rewards model.

Region	MNI Coordinates	Peak *t*(21)	Cluster Size (mm^3)
[1] Left/Right Precuneus (BA 7)	-15, -48, 42	8.42	98662
[2] Left Middle Frontal Gyrus (BA 46)	-47, 35, 19	7.16	8712
[3] Right Lentiform Nucleus	14, 0, 0	7.34	7602
[4] Right Middle Frontal Gyrus (BA 10)	39, 52, 12	7.17	4344
[5] Left/Right Culmen	-8, -40, -15	7.16	3528
[6] Left Middle Temporal Gyrus (BA 37)	-54, -56, -13	5.90	2592
[7] Left Declive	-14, -79, -18	6.02	2287

[8] Right Declive	8, -76, -16	6.83	1874
[9] Right Inferior Frontal Gyrus (BA 44)	53, 8, 16	5.83	1789
[10] Left Middle Frontal Gyrus (BA 10)	-33, 56, -8	7.07	1728
[11] Right Medial Frontal Gyrus (BA 10)	1, 52, 4	6.09	1691
[12] Left Culmen	-4, -59, -16	6.29	1606
[13] Right Middle Frontal Gyrus	43, 38, 30	5.98	1606
[14] Right Inferior Temporal Gyrus	52, -64, -5	6.56	1338
[15] Right Superior Frontal Gyrus (BA 8)	25, 28, 50	5.37	1290

Figure S7. BOLD response to PEs derived from the regret model ($p_{voxelwise} < .001$, cluster size = 100 voxels, $p_{corr} < .05$).

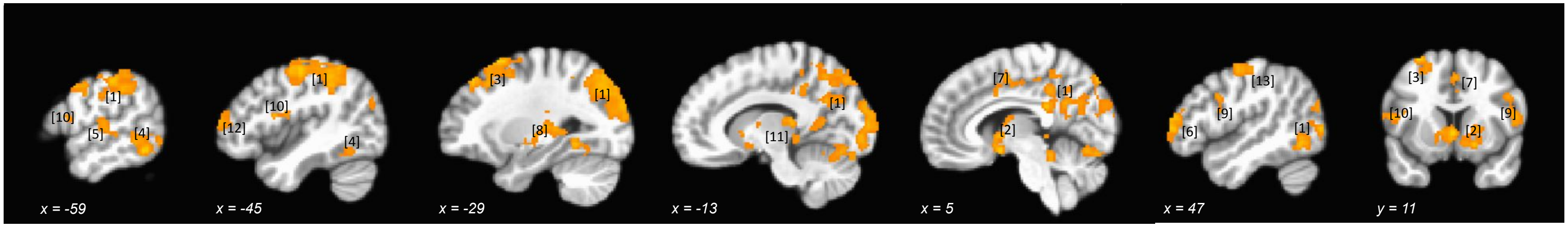


Table S4. BOLD response to prediction errors derived from the regret model.

Region	MNI Coordinates	Peak $t(21)$	Cluster Size (mm^3)
[1] Left/Right Precuneus (BA 7)	-7, -63, 30	8.47	116037
[2] Right/Left Caudate	8, 3, 0	8.06	9223
[3] Left Middle Frontal Gyrus (BA 8)	-30, 23, 52	7.61	8140
[4] Left Inferior Temporal Gyrus (BA 37)	-56, -55, -9	7.10	5998
[5] Left Superior Temporal Gyrus (BA 42)	-63, -18, 8	6.05	3407
[6] Right Middle Frontal Gyrus (BA 10)	44, 49, 9	6.88	3212
[7] Right Cingulate Gyrus (BA 24)	2, 1, 49	5.69	2616
[8] Left Thalamus	-28, -27, 4	7.20	2567
[9] Right Inferior Frontal Gyrus (BA 44)	52, 10, 20	5.42	2056
[10] Left Inferior Frontal Gyrus (BA 44)	-51, 5, 17	6.13	1874

[11] Right/Left Culmen	2, -37, -8	5.74	1728
[12] Left Middle Frontal Gyrus	-42, 54, 14	5.67	1728
[13] Right Precentral Gyrus (BA 4/BA 6)	48, -11, 56	6.29	1484
[14] Right Superior Frontal Gyrus (BA 8)	28, 27, 55	4.94	1290

Figure S8. BOLD response to PEs derived from the trustee-counterfactual model ($p_{voxelwise} < .001$, cluster size = 100 voxels, $p_{corr} < .05$).

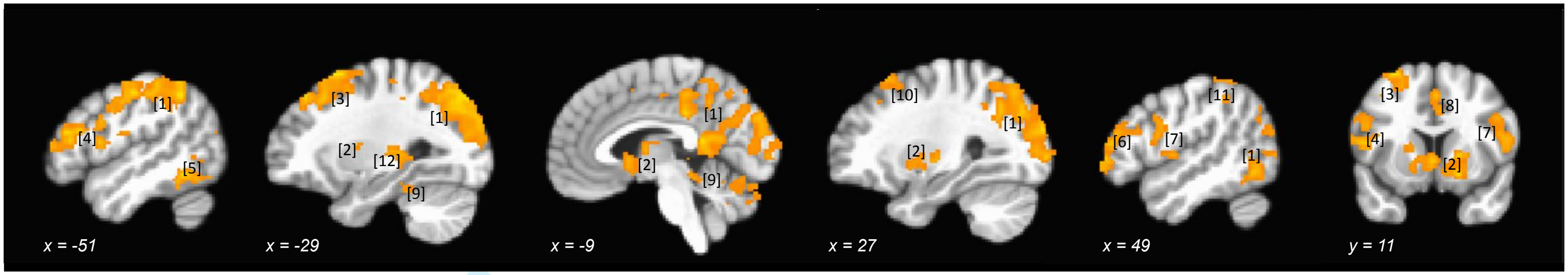


Table S5. BOLD response to prediction errors derived from the trustee-counterfactual model.

Region	MNI Coordinates	Peak $t(21)$	Cluster Size (mm^3)
[1] Left/Right Precuneus (BA 7)	-10, -61, 30	8.17	141746
[2] Right/Left Caudate	8, 2, 1	7.40	12350
[3] Left Middle Frontal Gyrus (BA 8)	-30, 24, 51	7.75	11583
[4] Left Inferior Frontal Gyrus (BA 45)	-48, 29, 18	7.74	9855
[5] Left Inferior Temporal Gyrus (BA 37)	-56, -54, -9	7.49	7142
[6] Right Middle Frontal Gyrus	44, 47, 11	7.27	5913
[7] Right Inferior Frontal Gyrus (BA 44)	52, 10, 17	5.80	4015
[8] Right Cingulate Gyrus (BA 32)	3, 9, 44	5.73	3954

[9] Left/Right Culmen	-5, -38, -12	6.24	3516
[10] Right Superior Frontal Gyrus	28, 26, 55	5.34	2178
[11] Right Inferior Parietal Lobule (BA 40)	51, -40, 53	5.46	2117
[12] Left Thalamus	-28, -33, 4	6.71	1557

Figure S9. BOLD response to non-model based contrast of congruent (invest-return and keep-keep) vs. incongruent (keep-return and invest-keep) trials ($p_{voxelwise} < .001$, cluster size = 60 voxels, $p_{corr} < .05$).

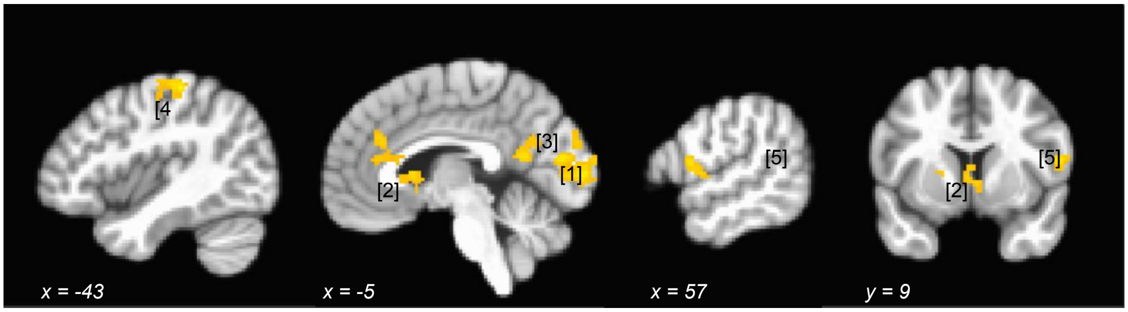


Table S6. BOLD response to non-model based contrast of congruent (invest-return and keep-keep) vs. incongruent (keep-return and invest-keep) trials.

Region	MNI Coordinates	Peak $t(21)$	Cluster Size (mm^3)
[1] Left Cuneus (BA 18)	-4, -90, 14	4.76	7544
[2] Left Caudate	-3, 16, 11	4.58	4806
[3] Right Cingulate Gyrus	2, -61, 25	4.48	3005
[4] Left Precentral Gyrus	-38, -23, 55	4.74	1922
[5] Right Precentral Gyrus (BA 44)	56, 7, 13	4.21	900

Supplementary results: Trustee effects on neural signals.

Reputation condition was included as a trial-level regressor and was also allowed to interact with the feedback-aligned prediction error. To be comprehensive, we evaluated all of the following contrasts and corresponding interactions: bad vs. good conditions, bad vs. all other conditions, good vs. all other conditions, human vs. computer conditions. None of these contrasts yielded activations above the significance threshold. It is important to note that, detracting from our ability to capture the reputation effects, trustee was a block variable which varied by run. Therefore unique condition-wise intercepts for reputation could not be recovered. This is in contrast to the previous study with a mixed trial design (Delgado et al., 2005). We chose this experimental design because we focused on the effects of reinforcement within each trustee block, and the reputation effects have been documented in previous studies. Importantly, we do replicate reputation effects on choice in behavioral data.