Supporting Information for
Neural responses to social rejection reflect dissociable learning about relational value and reward

Begüm G. Babür[1], Yuan Chang Leong[2], Chelsey X. Pan[1], and Leor M. Hackel[1]

University of Southern California[1], University of Chicago[2]

*Corresponding author: Leor M. Hackel

Email: lhackel@usc.edu

This PDF file includes:

Supporting text
Tables S1 to S7
Figures S1 to S7
SI References

Supplemental Methods

Stimuli. After recruitment, participants completed six open-ended questions about themselves, meant to allow other participants to supposedly form impressions of them, with an emphasis on trustworthiness:

1) What are 3 words a close friend would use to describe you?
2) For what in your life do you feel most grateful?
3) When was the last time you asked for forgiveness and what was it for?
4) What obstacles are you currently trying to overcome?
5) When was a time when you were honest, even though you didn't have to be?
6) What would you do if you were given credit for something that someone else actually did?

Task. Participants were greeted by the experimenters at the Dornsife Neuroimaging Center and, after giving informed consent, completed the task instructions outside of the scanner. Participants saw a refresher of the instructions inside the scanner and played practice rounds before starting the task.

Participants were told that 8 others in a "Decider" role had seen their responses to the getting-to-know-you questions they filled out in Part 1, along with the responses of 11 other players who were in a "Responder" role. Participants were told that the Deciders repeatedly chose who to send money in trust games. To ensure participants understood the task structure, instructions emphasized that matches were beyond the control of the Deciders, e.g.:

> On every round, they were shown a random 8 out of 12 Responders and ranked who they wanted to send points to, from 1 (most desired) to 8 (least desired). Deciders' rankings may change from round to round depending on which 8 of the 12 Responders they get to see. Some Deciders might have ranked you higher, and some people might have ranked you lower. You will have to choose Decider partners without knowing much about them at first.
>
> After that, the computer decided how many Responders they could match with on that round from 1 player to 8 players. For example, if the computer decided they could match with 4 Responders, then they would send points to their top 4 ranked Responders and not the bottom 4 ranked Responders. This means that whether you get matched on each round depends on (1) how a Decider ranks you (1st~8th) and (2) how many matches the computer allows (1~8 players).

Participants also saw sample feedback screens explaining what each part of the feedback screen meant, completed 5 practice trials (featuring avatars that were not used in the actual experiment), and had an opportunity to ask the Experimenter any questions before the task began.

Instructions also noted that, because Deciders would see 8 Responders at random out of 12 total, the rank given to the participant could vary from trial to trial, as the set of Responders available to the Decider varied.

As Responders, on each trial, participants saw 2 Decider avatars out of the 8, matched to their gender. Each avatar was equally likely to appear on either side of the screen throughout the trials. Participants chose who they would try to match within 2 seconds, by pressing 1 (left avatar) or 2 (right avatar) on the button box inside the scanner. The avatars were created from avatarmaker.net.

A 3 second feedback screen revealed whether they matched, as indicated by a green box around their blue icon; the number of other Responders who had been allowed to match with that Decider, as indicated by other green boxes around gray icons; and how highly the Decider had ranked them on that round, as indicated by a ranking underneath their avatar. If they did not choose a Decider within 2 seconds, participants saw a "No Response" screen for the same duration as the feedback screen.

For half the trials, participants explicitly saw their rank underneath their avatar (e.g: "Rank: 3rd"), whereas for the other half of trials, their rank was covered by a white square. Four Deciders always provided explicit ranks and four always had hidden ranks. Participants always chose between two rank-explicit Deciders or two rank-hidden Deciders. Within both groups of four Deciders, feedback was generated such that Deciders varied independently in the rank they gave participants (high or low) and the probability of matching they provided (high or low) (see Fig. 1B).

If participants matched with the Decider they selected, they played a trust game. Participants were told that the Decider gave them a number of points which got tripled and they had to choose whether to keep all the points to themselves or return half to the Decider in 2 seconds. Participants made their choices by pressing 1 (return half) or 2 (keep all) on the button box. If participants failed to match, they saw a "No Game" screen and waited 2 seconds without playing the game. The trust game was preceded by a jittered inter-stimulus interval and followed by a jittered inter-trial interval (1-8 seconds).

To generate feedback on each trial, Gaussian noise was added to the chosen Decider's average ranking and average number of matches (*SD* = 1 for each); a truncated distribution was used, such that both quantities had to be between one and eight. If the participant's ranking was within the number of matches allowed, they matched with that Decider for the trial. At the end of the experiment, subjects received cash compensation for their time and a bonus payment depending on the number of points they earned.

Model Comparison. The computational model was compared to two alternative models in which w was constrained to 0 or 1, yielding choices rooted entirely in outcomes or ranks, respectively. Models were compared using Bayesian Model Selection (1) as implemented in the m-fit package for Matlab (2). This procedure was performed over log model evidence, derived as the Laplace approximation to the log model evidence; for further discussion of this derivation, see ref (3).

Model Validation. Beyond identifying which of the models tested best fit the data, we conducted simulations to validate the winning model. First, we tested whether the model could, in principle, infer the correct ranks in the "rank hidden" condition. We simulated the model receiving feedback from each of the four Decider types, using the generative feedback process used in the task. We simulated 100 encounters with each Decider across 1,000 iterations. The model recovered the true underlying ranks for the two Deciders who provided average ranks of 3 (M = 3.01 and M = 3.05 across simulations). For the Deciders with ranks of 7, the model closely approximated these ranks; the inferred ranks had a slight downward bias (M = 6.82, M = 6.64), given that the distributions were truncated such that neither ranks nor group sizes could be larger than eight. (Note that no free parameters were used to conduct this simulation, given that updating was Bayesian and free parameters related only to choice in the model.)

Next, we tested whether the winning model could reproduce qualitative patterns observed in behavior at the group level (4). In particular, the key qualitative feature of the data was an effect of both rank and outcome feedback, as observed in the regression analyses reported in the main text and Supplemental Results. We simulated performance in the learning task using the estimated parameters of each subject 100 times. That is, each simulated subject corresponded to one actual subject and was simulated using that subject's best-fitting parameters and the sequence of stimuli presented to the actual subject, generating a data set of 42 pseudo-subjects; this procedure was repeated 100 times. In this manner, we tested whether the fitted parameters would create the patterns of behavior identified. For each simulation, we refit the regression analysis. The regression recovered a significant effect of outcome feedback in 100% of simulations (mean coefficient across all simulations = .16) and recovered a significant effect of rank feedback in 100% of simulations (mean coefficient across all simulations = .21). For visualization of real and simulated choice data, see Fig. 2.

We further asked whether the model could predict subjects' explicit post-scan ratings indicating how they think they were ranked by each Decider. (One subject could not be included in this analysis because they had no variance in these ratings.) For each subject, we extracted from the model the mean of the rank distribution held towards each Decider as of the end of the learning task. We tested whether these means correlated with subject estimates of the average rank they received from each Decider (from 1-8). Model estimates correlated with subject ratings, on average (M = .32, $t(40)$ = 4.53, p < .001). In addition, we extracted from the model the variance in the belief distribution held towards each Decider at the end of learning. We asked whether these values correlated with subject ratings of uncertainty in perceived rank. The variance of the belief distribution correlated with subject ratings, on average (M = .17, $t(40)$ = 2.80, p = .008). Notably, the present model did not incorporate any parameters that would allow individual differences in updating beliefs; the only individual differences modeled concerned how participants combined rank and outcome information in choice. The present model thus reflects an ideal observer model regarding updating. Future work can extend the model by incorporating systematic deviations from an ideal observer, which might further explain individual differences in idiosyncratic perceptions as reported at the end of the task.

Model recovery & parameter recovery. We next performed model recovery (5, 6), asking whether the three models could be accurately recovered from simulated data. We simulated 100 data sets under each of the three models we tested. Each data set had 42 simulated subjects, corresponding to the actual subjects in our behavioral analyses. Each simulated subject was generated using the best-fitting parameters of their corresponding real subject from the winning model (but leaving out parameters that did not apply to more constrained models). These models were applied to the actual sequence of stimuli viewed by the subjects to generate simulated choices. In all, this procedure generated 300 simulated data sets (100 data sets for each of 3 models), reflecting the simulated choices of each model based on the parameters and choice sets presented to each subject.

We then fit each data set to each of the three models, allowing us to generate a confusion matrix indicating the proportion of times each model was correctly identified (on the diagonal) or mis-identified (off-diagonals). The winning model in each iteration was defined as the model with the highest protected exceedance probability, corresponding to the analysis in the main paper. All three models were recoverable in 100% of simulations (Table S7).

Finally, we used these simulated data to conduct parameter recovery (5, 6), ensuring the winning model could accurately recover parameter values generating behavior. To do so, we averaged the parameters across the 100 data sets generated under the winning model and examined the correlation between these average parameters and each participant's true parameters. This analysis indicates whether the model can recover each participant's parameters on average. Across the two parameters in our model, we observed correlations of .91 (w) and .999 (β) (Fig. S6).

Robustness to choice of distribution. The computational model treated ranks as a continuous variable following a (truncated) normal distribution. This feature reflected the generative model of feedback in the task, as well as instructions given to the subjects. In particular, rank feedback could vary from trial to trial in the task. To explain why this could happen, subjects were told that Deciders saw different sets of Recipients each round, such that subjects were ranked in comparison to different individuals in different rounds. A Decider might therefore rank the subject 3rd when compared to some individuals and 4th when compared to others. Accordingly, the model tracks the *average* rank the Decider gives the target across different choice sets–a model of the Decider's preferences.

Alternatively, however, learning about ranks could be modeled using a Dirichlet distribution to update discrete probabilities of receiving each rank–a model of likely rank feedback. To ensure that results were not dependent on this modeling choice, we refit the model using a Dirichlet distribution representing the probability of receiving each possible rank from each Decider. Specifically, the model assumed the probability of receiving each rank from one to eight from a given Decider was distributed as

$(p1,\ldots,p8)\sim\text{Dirichlet}(\alpha1,\ldots,\alpha8)$

An initial prior of αi=1 was used for all ranks. On each trial, the distribution was updated according to

$(p1,\ldots,pk) \mid (x1,\ldots,xk)\sim\text{Dirichlet}(\alpha1+x1,\ldots,\alpha k+xk)$.

where X1 is set to 1 if that rank was received on that trial and is set to 0 otherwise. This rule provides a Bayesian update for the probability of each event (7). On trials in which ranks were hidden, analogous to the averaging in the original model, updates for different possible ranks were averaged by updating each by 1/N, where N is the number of ranks possible on that trial.

For instance, if one was accepted along with two others, the possible ranks are 1, 2, or 3, and these three ranks would each be updated by 1/3. To generate a rank-based estimate of the likelihood of acceptance on each trial, the probability of receiving each rank was multiplied by the likelihood of acceptance if one were to receive that rank, assuming a uniform distribution over possible number of matches allowed. These probabilities were summed to generate an overall probability of acceptance.

This model yielded similar results to the version with the normal distribution. In model comparison, the hybrid model incorporating rank and outcome again proved the best fit to the data (protected exceedance probability = .97). Next, final model-estimated ranks showed similar correlation with subjects' self-reported perceived ranks, $M = .37$, $SD = .42$, $t(40)=5.74$, $p < .001$. Finally, update values derived from this model also showed a similar relationship with neural signal (see Supplemental Results below).

RSA preprocessing. Before RSA was performed, beta weights were pre-whitened, following approaches found to reduce bias and increase reliability in past work (8, 9). Specifically, residuals in each voxel within the region of interest were concatenated across runs, generating a $t \times n$ matrix, where t refers to the number of timepoints across all runs and n refers to the number of voxels in the ROI. This was used to estimate a spatial covariance matrix of noise, using optimal shrinkage (10). Beta weights were multiplied by the matrix square root of the inverse of the covariance matrix (9).

Supplemental Results.

Regression analyses of choice. To analyze learning in a manner independent of the computational model, choices during the learning task were fit to a supplementary mixed effects logistic regression. This model predicted the probability of staying with the most recently chosen player of the two shown on screen (1 = stay, 0 = switch) as a function of rank feedback on the last trial featuring that player (continuous, standardized to z-scores within subjects), outcome feedback on the last trial featuring that player (1 = match, -1 = no match), the rank visibility condition of the players shown (1 = rank visible, -1 = rank hidden), and interactions of these predictors. For trials on which ranks were not visible, the rank regressor contained the rank used in the generative model of feedback on that trial. This model approximates the full reinforcement learning model by asking whether participants stay with their previous choice as a function of each feedback type (11, 12). A random intercept was included, and all random slopes were initially included; the smallest random slope was iteratively removed until the model converged (13), which led to removing the random slope for rank visibility.

Supporting the computational modeling results, this analysis revealed main effects of outcome and rank: participants were more likely to stay with players who had previously ranked them highly and with players who had previously provided matching outcomes (Table S1). No significant effects of rank visibility were observed.

We conducted two analyses of choice data to ensure these results were robust to other variables related to each trial. First, we refit the regression model predicting participant's choices based on rank and outcome feedback on a prior trial while controlling for (i) the total number of matches shown onscreen during that trial (z-scored) and (ii) prior trust game choices. Specifically, previous trust game choices toward the relevant Decider could reflect three types: kept, repaid, or hadn't yet played a trust game. That is, for any given trial, it is possible that subjects had never received a matching outcome from the relevant Decider and thus had never played the Trust Game; had chosen to repay in their most recent Trust Game with that Decider; or had chosen to keep in their most recent Trust Game with that Decider. These three categories were accounted for by two dummy code regressors. The first regressor accounted for whether subjects had never completed a trust game choice with the relevant Decider (0 = no, 1 = yes). The second regressor indicated whether the most recent trust game choice with the relevant Decider was to repay (0 = no, 1 = yes). The reference category was trials in which the most recent trust game choice with the relevant Decider was to keep. This coding scheme thus accounts for all three possible prior trust decisions toward a Decider while avoiding confusing effects of trust game choice with effects of outcome history (i.e., whether or not a subject had matched before with a Decider). This analysis again revealed main effects of rank and outcome feedback from the previous trial (Table S2); consistent with the computational model, it also revealed an effect of whether the subject had ever matched with the Decider, reflecting longer-term outcome histories.

Next, we aimed to adjust for uncertainty in rank and outcome estimation. To do so, we ran the model forward for each subject, estimating, for each trial, the mean and variance of the belief distributions for rank and outcome held towards each target shown onscreen at the time of choice. We used mixed effects logistic regression to predict whether subjects chose the target on the right side of the screen (arbitrarily chosen) as a function of the difference between targets (right – left) in (i) mean of rank beliefs, (ii) mean of outcome beliefs, (iii) variance of rank beliefs, (iv) variance of outcome beliefs. We found that mean estimates of rank and outcome continued to predict trial by trial choices (Table S3). This analysis indicates that the choice findings were robust to the presence of uncertainty in rank and outcome estimates.

Robustness analyses in univariate data. To ensure the univariate findings regarding updating of relational value were robust when accounting for additional task variables, we performed additional analyses of univariate data. First, we aimed to test whether the updating findings held across trial types (rank visible or hidden); the primary whole-brain analysis reported in the main text included trials from both conditions, maximizing the number of trials and increasing power for the whole-brain analysis. However, to test whether this finding truly holds across each condition, we examined responses in the regions identified. Specifically, we extracted trial-by-trial beta weights corresponding to the feedback epoch of each trial, providing an estimate of neural activity in response to feedback on each trial. This was done using a GLM identical to GLM 2 in the main text, in which each feedback epoch was modeled as its own event, except that 8mm smoothing was used as in other univariate analyses. Beta weights were extracted from the regions identified in the univariate analysis of unsigned rank updating (Fig. 3), providing an estimate of neural response within these regions on each trial. To ensure that region of interest (ROI) creation was independent of further statistical tests, we used a Leave-One-Subject-Out procedure, in which the ROI for each subject was defined using a group-level analysis for all other subjects (but excluding the subject under consideration) (14). In this manner, each subject's data was independent of the data used for ROI definition.

We fit a mixed effects regression predicting the average trial-by-trial beta weight from this ROI, using the same computational signals that served as parametric modulators in the whole-brain analysis (unsigned model updating, signed updating, and surprise, for both rank and outcome). (The mean correlations between regressors for rank update and reward update, as computed by averaging across subjects, were $M = .49$ for signed updates and $M = .60$ for unsigned updates.) Next, we added an indicator for trial type (1 = rank explicit, -1 = rank inferred). Trial type was interacted with the rank regressors to allow for the possibility that the effects of rank depend on condition. Finally, we adjusted for additional task variables, including (1) the total number of matches shown onscreen, (2) uncertainty in the posterior distribution of rank (i.e., variance in the belief distribution at that timepoint), (3) uncertainty in the posterior distribution of outcome, and (4) previous trust game choices (using the same coding scheme as in the behavioral regression analysis described above). The analysis was performed using the lme4 and lmerTest packages for R (15–17). Random intercepts were included, and all random slopes were initially included; the smallest random slope was iteratively removed until the model converged (13). Continuous regressors were z-scored.

We continued to observe a main effect of unsigned rank update in predicting neural responses, $b = .28$, $SE = .08$, $t(5502.61) = 3.41$, $p < .001$, and observed no interaction with trial type (Table S4). When testing simple effects in each trial type, the effect of rank update was significant both when rank was explicit ($b = .32$, $SE = .11$, $t(5503.66) = 2.82$, $p = .005$) and when rank was inferred ($b = .23$, $SE = .09$, $t(5479.49) = 2.61$, $p = .009$). This finding indicates that these regions tracked model updating about ranks both when ranks were explicit and when ranks were inferred, above and beyond other task variables.

Finally, as described above, we ensured that the results were robust to modeling choices when using a Dirichlet distribution, rather than a normal distribution, to model rank updates. The effect of rank update remained when using updates defined using this alternative model ($b = .75$, $SE = .30$, $t(5496.62) = 2.45$, $p = .01$).

Decider identity RSA. As described in the main text, RSA was performed to test for voxel patterns encoding Decider identity. Specifically, trial-by-trial neural dissimilarity was compared (with Spearman correlations) to an indicator for whether each pair of trials shared the same Decider or featured different Deciders. This analysis yielded a significant average relationship between neural dissimilarity and Decider identity across participants, $M = .005$, $SD = .01$, $t(39) = 2.23$, $p = .03$.

Robustness checks in RSA. To ensure analysis of trial-by-trial voxel patterns was robust when accounting for additional task variables, we refit multiple regression RSA while adjusting for the same variables described above.

Neural similarity was examined, again restricted to cross-run trials (i.e., trials from different runs). Predictors included similarity across trials in (i) rank distribution (z-scored), (ii) outcome distribution (z-scored), (iii) the number of total matches displayed onscreen (z-scored), (iv) prior trust game choices toward the Decider (0 = same choice, 1 = different choice), (v) uncertainty in rank estimation (i.e., variance of the belief distribution), (vi) uncertainty in outcome estimation (i.e., variance of the belief distribution), (vii) trial number, and (viii) trial type (rank visible or rank hidden, effect coded as -1 or 1). The average coefficients were computed across subjects. We continued to observe effects of rank similarity, $M = .01$, SD = .02, $t(39) = 2.65$, $p = .01$, and outcome similarity, $M = .03$, SD = .03, $t(39) = 5.03$, $p < .001$. When reconducting the permutation feature importance analyses described in the main text, similarity between encoding of Deciders and rank, as opposed to Deciders and outcome, continued to correlate with analogous difference scores in post-task ratings of being liked, $r = .59$, $p < .001$.

Whole brain contrast of explicit vs. hidden feedback: The robustness analyses described above found that the effect of rank updating held when rank feedback was explicit or hidden. However, to further understand potential effects of receiving explicit or inferred feedback, we tested for differences in neural activity across these two trial types. To do so, we fit a GLM identical to GLM 1 described in the main text, except that different trial types were entered in different conditions. We contrasted overall activation in “rank explicit” versus “rank hidden” trials, regardless of updating; this analysis therefore tested overall effects of trial type, regardless of the feedback presented or learning signals involved. When ranks were explicit, as opposed to inferred, greater activation was observed in posterior cingulate cortex, precuneus, and inferior parietal lobule (Fig. S7).

Table S1. Regression coefficients for analysis of choice, with odds ratios and Wald confidence intervals.

Variable	***b***	***SE***	***z***	***p***	***OR***	**95% CI**
Intercept	.22	.06	3.44	<.001	1.24	[1.10, 1.41]
Outcome	.22	.05	4.23	<.001	1.24	[1.12, 1.37]
Rank	.21	.06	3.56	<.001	1.23	[1.10, 1.38]
Rank Visibility	.04	.03	1.24	0.22	1.04	[0.98, 1.10]
Outcome × Rank Visibility	-.01	.04	-.39	0.70	0.99	[0.92, 1.06]
Rank × Rank Visibility	.04	.05	.78	0.43	1.04	[0.94, 1.16]

Table S2. Regression coefficients for secondary analysis of choice adjusting for total number of matches and trust game choices, with odds ratios and Wald confidence intervals.

Variable	***b***	***SE***	***z***	***p***	***OR***	**95% CI**
Intercept	.18	.07	2.47	0.01	1.19	[1.04, 1.37]
Outcome	.16	.07	2.48	.01	1.18	[1.04, 1.34]
Rank Visibility	.04	.03	1.25	.21	1.04	[0.98, 1.10]
Rank	.21	.08	2.61	0.009	1.24	[1.05, 1.45]
Total number of matches onscreen	.03	.07	.48	.63	1.04	[0.90, 1.19]
No Trust Choices (never matched)	-.34	.14	-2.39	.02	.71	[0.54, .94]
Repaid Trust	.15	.09	1.69	.09	1.17	[.98, 1.39]
Outcome × Rank Visibility	-.01	.04	-.24	0.81	0.99	[0.93, 1.06]
Rank × Rank Visibility	.03	.05	.60	0.55	1.03	[0.93, 1.14]

Table S3. Regression coefficients for secondary analysis of choice based on the mean and variance of the belief distributions of rank and outcome, with odds ratios and Wald confidence intervals.

Variable	***b***	***SE***	***z***	***p***	***OR***	**95% CI**
Intercept	-.16	.10	-1.71	0.09	0.85	[0.71, 1.03]
Rank Mean	.37	.09	4.00	< .001	1.45	[1.21, 1.74]
Outcome Mean	.34	.07	4.79	< .001	1.41	[1.22, 1.62]
Rank Variance	.-.002	.06	-.04	.97	.998	[0.88, 1.13]
Outcome Variance	-.19	.08	-2.35	0.02	0.82	[0.70, 0.97]

Table S4. Regression coefficients for ROI analysis of trial-by-trial univariate data, examining effects across trial types and adjusting for additional task variables.

Variable	***b***	***SE***	***df***	***t***	***p***
Intercept	0.09	0.13	56.52	0.67	0.50
Signed Rank KL	0.01	0.06	5463.95	0.19	0.85
Trial Type	0.06	0.06	5395.56	0.96	0.34
Signed Outcome KL	0.07	0.06	5481.22	1.09	0.28
Unsigned Rank KL	0.28	0.08	5502.61	3.41	< 0.001
Unsigned Outcome KL	-0.26	0.13	5504.18	-1.96	0.05
Rank Surprise	-0.17	0.09	5454.24	-1.81	0.07
Outcome Surprise	0.23	0.10	5497.73	2.20	0.03
Total number of matches	0.10	0.06	55.06	1.66	0.10
Rank Variance	0.06	0.09	97.41	0.64	0.52
Outcome Variance	-0.02	0.07	4854.07	-0.30	0.77
No Trust Choices (never matched)	0.16	0.13	120.30	1.21	0.23
Repaid Trust	0.21	0.15	241.68	1.36	0.18
Signed Rank KL × Trial Type	0.05	0.05	5457.06	1.03	0.30
Trial Type × Unsigned Rank KL	0.05	0.06	5492.99	0.78	0.44
Trial Type × Rank Surprise	-0.09	0.08	5452.27	-1.16	0.25

Table S5. ANOVA results for ratings of being liked.

Variable	**Df_{Num}**	**Df_{Denom}**	***F***	***p***	**η_p^2**
Rank Visibility	1	41	<.001	0.998	<.001
Rank	1	41	20.39	< .001	0.33
Outcome	1	41	13.76	< .001	0.25
Visibility × Rank	1	41	13.02	<.001	0.24
Visibility × Outcome	1	41	1.45	0.24	0.03
Rank × Outcome	1	41	2.35	0.13	0.05
Visibility × Rank × Outcome	1	41	.03	0.86	<.001

Table S6. Best-fitting parameter estimates from the computational model.

Parameter	Median	Mean
w	.52	.48
β	1.51	2.46

Table S7. Results of model recovery. Data were simulated based on each of the three possible models and fit to all three models, allowing model comparison when the generative model was known. This procedure was used to generate a confusion matrix indicating the proportion of times each model was correctly identified (on the diagonal) or mis-identified (off-diagonals). The winning model in each iteration was defined as the model with the highest protected exceedance probability. All three models were recoverable in 100% of simulations.

Proportion best-fitting model	**Generative Model**			
		***w* = free**	***w* = 0**	***w* = 1**
	***w* = free**	1	0	0
	***w* = 0**	0	1	0
	***w* = 1**	0	0	1

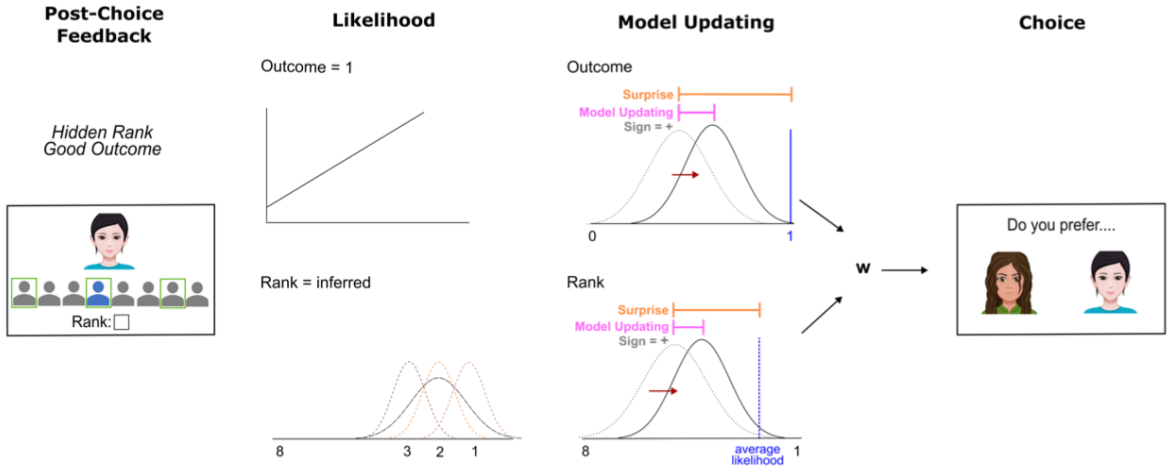

Fig S1. Schematic of computational model when ranks were hidden and had to be inferred. In this example, the participant was accepted along with two others. It is therefore possible that they received a rank of 1, 2, or 3. The three likelihood curves (dashed colorful lines) that would occur for these ranks are averaged to create an overall likelihood from the feedback (black line). This overall likelihood is combined with the prior to generate a posterior. The belief distributions for outcome and rank are then combined to guide choice, as they are when ranks are visible.

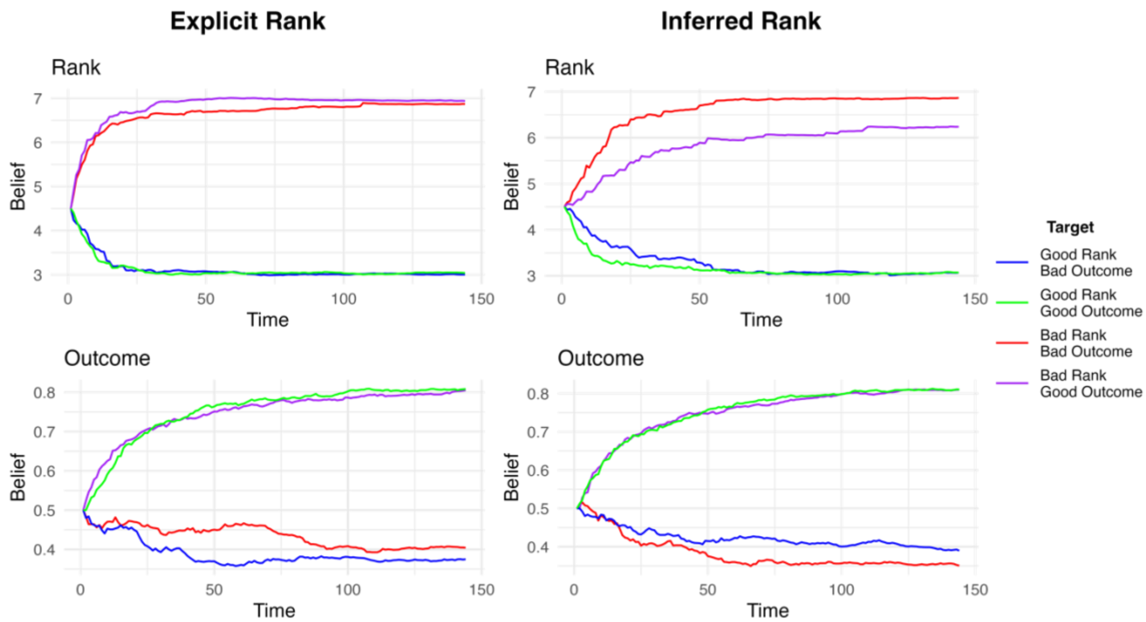

Fig S2. Average belief trajectories as estimated by the computational model. For each subject, the model-estimated mean of the rank distribution and mean of the outcome distribution was computed for each trial; these values were then averaged across subjects at each timepoint. Over time, the model learns whether Deciders tend to give better or worse ranks and outcomes.

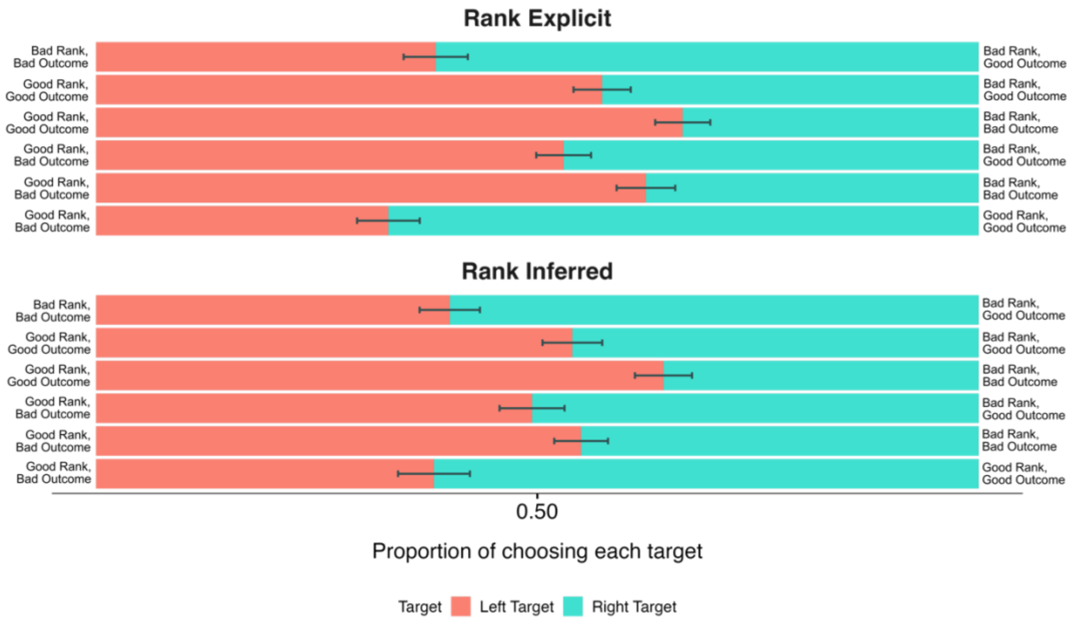

Fig S3. Average choice proportions in choices between each pair of Deciders. Each row represents choices between one of possible pairs of Deciders, split between Rank Explicit and Rank Inferred trials. Bars depict the average proportion of choices in which participants chose the target type displayed here on the left (red) or right (blue). Participant choices were influenced by both rank and outcome.

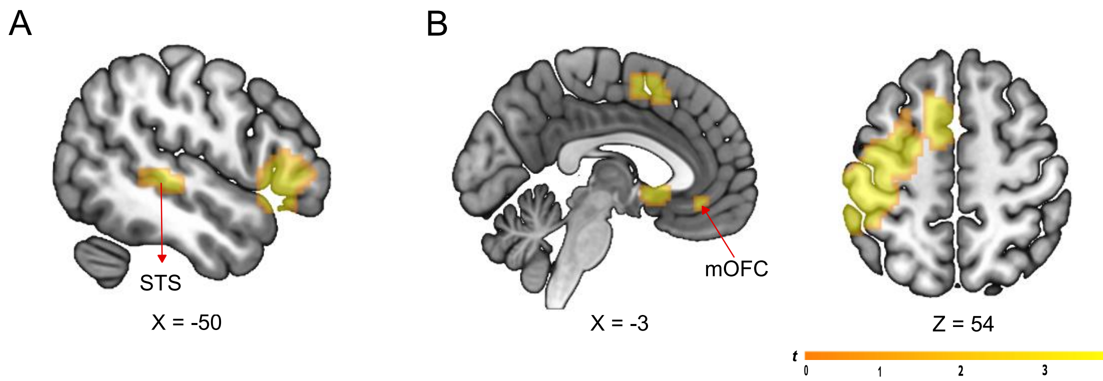

Figure S4. Additional activations for univariate analyses of unsigned rank update and positive outcome update. **(A)** Activation for unsigned rank updating was observed in superior temporal sulcus (STS). **(B)** Activation for signed outcome updating was observed in medial orbitofrontal cortex (mOFC). Additional activation was observed stretching from posterior parietal cortex to superior frontal gyrus.

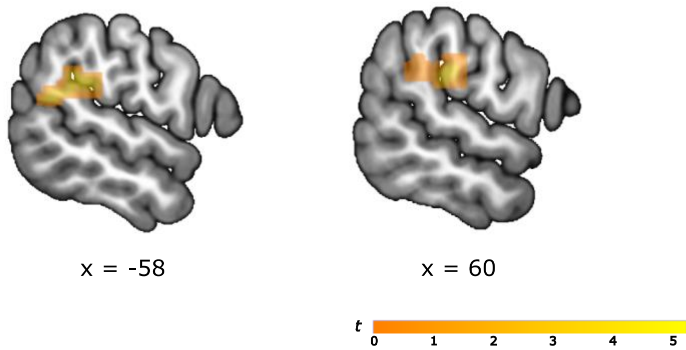

Figure S5. Activations for univariate analysis of negative outcome update. Activation was observed in bilateral temporoparietal junction/supramarginal gyrus.

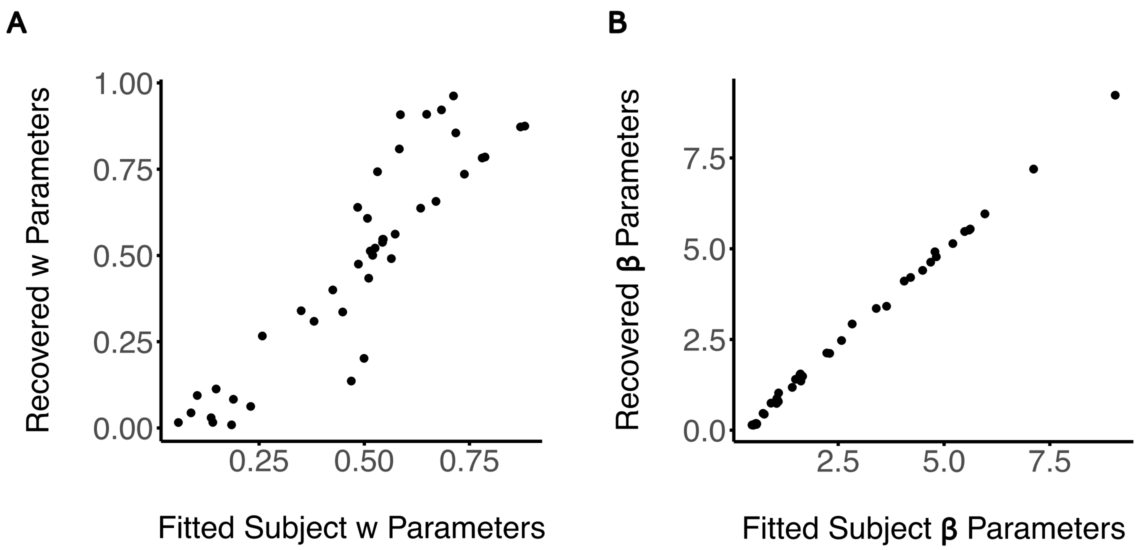

Fig. S6. Results of parameter recovery. The x axes display fitted subject parameters, and the y axes display average recovered parameters in simulations, for (a) *w* and (b) β.

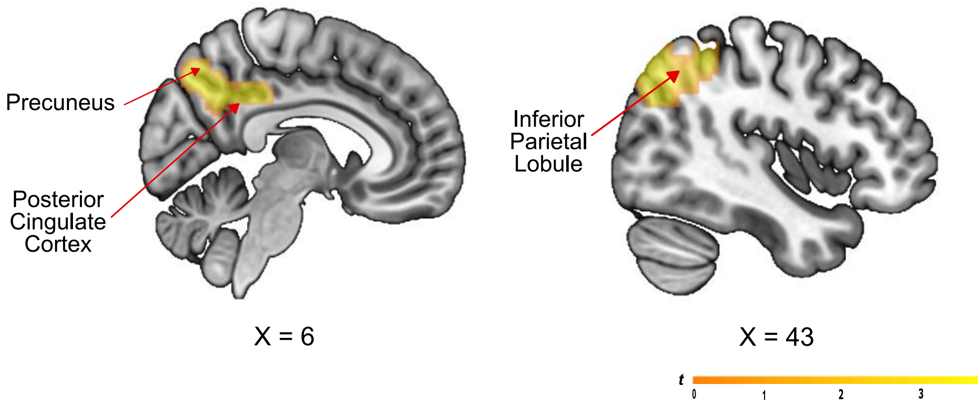

Fig S7. Whole brain contrast of overall responses in rank explicit versus rank inferred trials. This analysis tested for differences in neural activity when rank was displayed explicitly versus inferred, regardless of updating. Greater activation was observed during the rank explicit trials compared to rank inferred trials in posterior cingulate cortex/precuneus and inferior parietal lobule.

Supplemental References

1. K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
2. S. J. Gershman, Empirical priors for reinforcement learning models. *J. Math. Psychol.* **71**, 1–6 (2016).
3. N. D. Daw, Trial-by-trial data analysis using computational models. *Decis. Mak. Affect Learn. Atten. Perform. XXIII* **23** (2011).
4. S. Palminteri, V. Wyart, E. Koechlin, The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).
5. R. C. Wilson, A. G. Collins, Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, e49547 (2019).
6. S. Suzuki, K. Katahira, Applying reinforcement learning to the psychopathology of obsessive-compulsive and gambling disorders: practices and pitfalls in computational model fitting. (2024).
7. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian data analysis* (Chapman and Hall/CRC, 1995).
8. M. B. Cai, N. W. Schuck, J. W. Pillow, Y. Niv, Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLoS Comput. Biol.* **15**, e1006299 (2019).
9. A. Walther, *et al.*, Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).
10. O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).
11. B. B. Doll, K. D. Duncan, D. A. Simon, D. Shohamy, N. D. Daw, Model-based choices involve prospective neural activity. *Nat. Neurosci.* **18**, 767–772 (2015).
12. A. R. Otto, S. J. Gershman, A. B. Markman, N. D. Daw, The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* **24**, 751–761 (2013).
13. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
14. M. Esterman, B. J. Tamber-Rosenau, Y.-C. Chiu, S. Yantis, Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage* **50**, 572–576 (2010).
15. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *ArXiv Prepr. ArXiv14065823* (2014).
16. R Core Team, R: A language and environment for statistical computing. (2016). Deposited 2016.
17. A. Kuznetsova, P. B. Brockhoff, R. H. Christensen, lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).