**Supplementary Material**

# Supplementary Methods

## Participant Characteristics

The final sample consisted of participants living in the US (8%), Canada (7%) and the UK (85%). Of the HD subjects, 3% were Black, 20% Asian, and 78% Caucasian; 15% were employed, 30% unemployed, and 55% students in higher education. Of the LD participants, 8% were Black, 12% Asian, and 80% Caucasian; 31% were employed, 6% unemployed, and 63% students in higher education.

## Procedure

After the screening, eligible participants completed several online questionnaires at home. Besides the questions mentioned in the main paper, subjects were also asked how close they feel to their friends, how much time they spend interacting with these friends, and how pleasant they find these interactions, as well as how difficult they find it to make new friends.

Once participants had completed the online questionnaires, a testing session was scheduled. At the beginning of the session, all participants were asked to choose an avatar that would represent them during the task. Additionally, they were introduced to two other people with whom they were going to perform the task. These introductions included the other people’s names, as well as some personal details, and took place online for the online participants and in person for those tested at the university. All participants then completed the task alone on an online platform.

Subjects were told that, as part of the task, they would be planning a party by making choices between party decoration items on which they would receive feedback. They were further instructed that, in the social condition, the feedback (thumbs up, horizontal, or down) was provided by the other people in real time through the online platform, whereas, in the non-social condition, monetary feedback (wins, no change, or losses) was given by the computer. In reality, the feedback in both conditions was computer-generated and, apart from the stimuli and feedback type, the social and non-social conditions were identical.

## Learning Task

During the task, participants’ aim was to maximize rewards and minimize ‘punishments’. At the beginning of each trial, subjects were shown two party decoration items side-by-side and were asked to select one item by pressing the ‘C’ or ‘M’ key. The images were displayed until a selection was made, and after each choice the words ‘waiting for feedback from others’ or ‘waiting for computer-generated feedback’ were presented for between 3300ms and 7500ms. This relatively long and variable interval was chosen to give the impression that, in the social condition, the other participants were truly selecting the feedback. Subsequently, participants were given positive, neutral or negative feedback, determined probabilistically as described below. Moreover, subjects were asked to rate how they felt about the outcome, using a visual analogue scale ranging from ‘very bad’ to ‘very good’ (0 to 100). The feedback stayed on the screen until the rating was submitted.

In the social condition, the feedback consisted of ‘like’ and ‘dislike’ signs as used on social media (thumb up or down), as well as ‘neutral’ feedback in the form of a horizontal thumb. Participants were told that the feedback came from the other two people with whom they were completing the task, with ‘likes’ or ‘dislikes’ indicating that both of the others approved or disapproved of their selection, respectively. Moreover, horizontal thumbs were said to reflect that one of the other two people liked and the other disliked the participant’s choice. This outcome was chosen, rather than e.g. stating that no rating was available on neutral trials, because the outcomes would have seemed unrealistic if the other two people had always agreed on their feedback. Note that, in reality, all feedback was computer-generated. In the non-social condition, monetary feedback was provided in the form of winning 5 pence, no outcome, or losing 5 pence, represented by an image of a golden coin, a circle, or a crossed-out coin, respectively. Subjects were informed that the money they won during the task would be added to their reimbursement.

It may be argued that the horizontal thumb in the social condition (as a representation of one ‘like’ and one ‘dislike’) is not an entirely neutral outcome, if the perceived pleasantness of ‘likes’ is stronger than the perceived unpleasantness of ‘dislikes’, or vice versa. However, it should be noted that, depending on a given participant’s focus, receiving no monetary outcome could similarly have been be regarded as a slightly positive (‘not a loss’) or a slightly negative (‘not a win’) event. Thus, the two conditions were matched in this regard, and the fact that ‘neutral’ outcomes were potentially not perceived as entirely neutral was taken into account in the computational modelling of the task data (see below).

Eight party decoration items were used as stimuli during the task (balloons, garlands, lanterns, pinwheels, party hats, party horns, candles, and confetti). For each participant, four items were randomly allocated to the social condition and the other four items to the non-social condition. Moreover, for each condition, each item was randomly assigned to one of the following outcome contingencies: 75% (item 1) or 25% (item 2) chance of yielding positive rather than neutral feedback, or 75% (item 3) or 25% (item 4) chance of yielding negative rather than neutral feedback. Note that items 1 and 2 could never yield negative feedback, while items 3 and 4 could never result in positive feedback.

Given that there were four items per condition, there were six possible pairings. Each of the pairings was repeated twelve times (with the two items displayed on counterbalanced sides of the screen), yielding a total of 72 trials per condition.

During the task, participants’ choices, their reaction times, and their ratings in response to the feedback were recorded. Additionally, explicit outcome expectancies were assessed by asking participants to rate each item twice: once on how likely they thought selecting this item would result in positive feedback, and once on how likely they thought choosing this item would result in negative feedback. Ratings were made on a visual analogue scale ranging from ‘very unlikely - 0%’ to ‘very likely - 100%’ and were collected in the middle and at the end of each condition.

## Analysis

### Negative Bias Scores

Negative bias scores were calculated as follows: first, we averaged negative outcome expectancy ratings from the middle and end of the task for those items which never yielded negative feedback (items 1 and 2) and, separately, averaged positive outcome expectancies for those items which never yielded positive feedback (items 3 and 4). We then subtracted the mean positive expectancy from the mean negative expectancy to obtain a negative bias score. This score indicates how much more negative than positive feedback participants expected to receive independently of the actually experienced outcomes (i.e. for choices that never yielded negative or positive outcomes, respectively). To account for the fact that participants’ feedback expectancy ratings for the individual items may be influenced by the *overall* amount of positive and negative outcomes they experienced throughout the task, we also ran relevant analyses with the difference between the actual overall positive and negative feedback counts as a control variable.

### Computational Modelling

Q-learning models were fit separately to the social and non-social data. The Q-values, which indicate the predicted outcome value associated with choosing a given item, were initialised at 0 and updated on each trial (t) for the selected item (A) as follows:

where r(t) is the outcome value and αG and αL are the learning rates for positive and negative prediction errors, respectively. The outcome value was fit individually for each participant with the use of the free parameter *d* (as in Gold *et al.* 2012). Specifically, r(t) was set to 1-d for rewards, to -d for ‘punishments’, and to the midpoint between these values [i.e. ] for ‘neutral’ outcomes (which provided a better fit than using d = 0). Note that, from a theoretical perspective, setting the value of ‘neutral’ outcomes to the midpoint between reward and punishment values is appropriate, particularly in the social condition in which ‘neutral’ outcomes represented receiving a ‘like’ from one person and a ‘dislike’ from the other person (thus likely leading to a slightly positive or negative perception of ‘neutral’ outcomes, depending on whether ‘likes’ or ‘dislikes’ are valued more highly). It should further be noted that d values need to be interpreted relative to the initial Q-value of 0. That is to say, the d parameter determines how large the impact of rewards and ‘punishments’ is in relation to the initial outcome expectation.

To account for potential forgetting of the implicitly learned Q-values while making ratings in the middle of the task (after trial 36), all Q-values were decayed towards 0 for trial 37, with a free parameter (ω) determining the strength of this decay as follows (similar to Collins & Frank 2012)

On every trial, the probability of a given participant’s choice (of item A over B) under the model was computed using a softmax function:

where τ is the explore/ exploit temperature parameter, is an indicator variable which is set to 1 if item A was chosen the last time it was shown and to otherwise (where *γ* is a decay parameter), and φ is the choice bias parameter representing how likely participants are to repeat an item choice *independently of the outcome it yields* (i.e., “sticky choice”; (Schonberg *et al.* 2007)).

Models containing different combinations of the free parameters (αG, αL, d, φ, *γ*, ω, τ; see Table 2) were fit to each participant’s data by maximising the log likelihood estimate (LLE) of the participant’s choices under the model across all trials, thus maximising:

The model fitting was conducted in two hierarchical steps. In step 1, the maximum likelihood estimation (MLE) was run without a prior, as described above. In step 2, the MLE was re-run using a multivariate Gaussian prior on the parameter values. The prior was parameterised with the mean and covariance (joint across *all* participants) of the parameter estimates from step 1. That is to say, each parameter value was evaluated on the abovementioned prior, and the log of this value was added to the LLE, thereby causing a higher increase in the LLE for parameter estimates that are more likely under the prior. This “shrinkage” procedure reduces the variance in the parameter estimates by bringing extreme values closer to the overall mean (Daw 2011) and approximates hierarchical Bayesian estimation (although, unlike in Bayesian estimation, the same prior was applied to the data of *all* subjects, rather than using group-specific priors). Although this approach does not implement the full Bayesian solution (i.e., it does not yield posterior distributions over parameters), we confirmed that leveraging the statistical structure across the group in this way nevertheless improves parameter recovery from simulated data based on the same task and trial number.

Akaike's Information Criterion weights were computed (as outlined in Wagenmakers & Farrell, 2004) and utilised to compare the relative fit of the different models. Two models (Q16 and Q4 in Table 2) were among the best fitting models for both the social and non-social condition. For one of these models (Q16) the mean values of two of the parameters (learning rate and memory decay) were numerically similar for the two conditions. Thus, three further models were fit in which one or both parameters were shared for the social and non-social data fitting, while the other parameters could vary. AIC weights were compared between the models with shared parameters and the model in which all parameters differed between conditions.

Additionally, the fit of the best model compared to chance was assessed with the use of *pseudo-R2* values, which (as in (Frank *et al.* 2007)) were calculated as follows: *pseudo-R2* = (LLE learning model - LLEnull model) / LLEnull model, where LLEnull model is the log likelihood estimate of the data under a model that assumes random choices [i.e. LLEnull model = number of trials \* log(0.5)]. The data of one LD participant in the social condition and of five LD and five HD subjects in the non-social condition demonstrated a better fit for the null model than for the learning model. When data from these participants were excluded, the same pattern of results arose. Therefore, these data were included in the analyses reported in the main paper.

For model validation, parameters from the best fitting model were used to simulate data. Subsequently, the generating model was fit back to the simulated data (using the two-step procedure described above) to assess if the parameters used in the simulation could be recovered (as assessed by Spearman correlations between the initial and recovered parameters).

Parameter values from the best fitting model were compared between groups using Mann-Whitney U tests. Additionally, to examine whether social learning deficits relate to the experience of positive or negative social outcomes in real life, a multiple regression analysis was performed, using social learning rate, outcome valuation, and temperature parameters (from model Q16; see Table 2 and below) to predict the reported amount of time spent in pleasant and unpleasant social situations. Given that anhedonia, anxiety and negative biases are likely to additionally influence the reported amount of time spent in these situations, RSAS, SAQ and negative bias scores were also entered into the analysis, and BDI scores were added as a control variable. Again, the assumption of normally distributed residuals was violated for the raw data, which is why the regression was performed on rank transformed data using L-statistics as described above. Some of the predictor variables were moderately correlated; however, collinearity assumptions were not violated (all VIF < 5).

To confirm the robustness of our findings, the above analyses were also performed on the estimated parameters of a different model (Q4 in Table 2) which had an AIC weight close to that of the best fitting model. Given that the rank-transformed learning rate and temperature parameter values were highly correlated in this model, the collinearity assumption of the regression analyses was violated (VIF > 10). Thus, one of the parameters which did not show a significant predictive effect when model Q16 parameters were used (namely the temperature and learning rate parameters for the prediction of time spent in unpleasant and pleasant social situations, respectively) was removed from the regression analyses.

# Supplementary Results

## Real-Life Social Interactions

HD subjects reported finding it less pleasant to spend time with friends (U = 1122, p = 0.042), feeling less close to their friends (U = 1169, p = 0.014), and finding it more difficult to form new friendships (U = 531, p = 0.001) than LD subjects. When adjusted for the reported number of friends, the amount of time spent with friends did not differ between groups (U = 773, p = 0.293).

## Learning Task

### Ratings of Arousal to the Task Feedback

A mixed measure ANCOVA (group x condition x valence, controlling for testing location) on participants’ arousal ratings for positive, neutral and negative feedback revealed a significant main effect of condition (*F*(1, 83) = 6.48, *p* = 0.013), with higher reported arousal in the social than in the non-social condition. Additionally, a significant main effect of valence was found (*F*(2, 166) = 33.48, *p* < 0.001) due to higher arousal to positive feedback than to negative (*t*(88) = 4.39, *p* < 0.001) or neutral (*t*(88) = 12.84, *p* < 0.001) outcomes, as well as higher arousal to negative than to neutral feedback (*t*(88) = 6.94, *p* < 0.001). Moreover, a group by condition by valence interaction was observed (*F*(2, 166) = 5.47, *p* = 0.005; see main paper for follow-up analysis).

### Ratings of Emotional Responses to the Task Feedback

A mixed measure ANCOVA (group x valence x condition, controlling for testing location) on participants’ emotional responses to the positive, neutral and negative feedback showed the expected main effect of valence (*F* (1.21, 106.22) = 124.82, *p* < 0.001), with participants feeling better after receiving positive than after getting neutral (*t*(90) = 16.22, *p* < 0.001) or negative (*t*(90) = 17.50, *p* < 0.001) feedback and after receiving neutral than after getting negative feedback (*t*(90) = 14.32, *p* < 0.001) across the social and non-social conditions. No other significant main effects or interactions were observed (all *F* < 0.4).

### Regression – Prediction of Time Spent with Friends

Regarding the regression analysis using social anhedonia and negative bias scores to predict the amount of time spent with friends in real life (see main paper), the following should be noted: it cannot be ruled out that the negative biases observed in the task were the result of a generalization from negative social experiences in real life to the experimental setting (see discussion). It could therefore be the case that, in the regression analysis, the negative bias values act as a ‘proxy’ for a direct effect of negatively perceived interpersonal encounters on social withdrawal. We thus reran the analysis with the reported amount of time spent in unpleasant social situations as an additional control variable. The observed pattern of results was similar as that reported in the main paper, with both RSAS scores (β = -0.61, *p* < 0.001) and negative biases (β = -0.19, *p* = 0.046) contributing significantly to the prediction of the amount of time spent with friends. This indicates that, independent of how many negatively perceived social situations they encounter, individuals who are more anhedonic and who expect more negative social outcomes spend less time with their friends.

### Task Feedback Ratings

A Mann-Whitney U test was performed on participants ratings of how sure they were that the feedback in the social condition came from other people. There were no group differences (*U* = 887, *p* = 0.909), with average ratings of around 5 out of 10 across all participants (MHD = 5.08, SDHD = 2.82; MLD = 5.06, SDLD = 3.03). Although this indicates that participants did not fully believe that the feedback was provided by other people, it should be noted that, as long as subjects thought there was a chance that the feedback came from the others, they are likely to have behaved as if it did. Moreover, the very question itself may have induced participants to be uncertain about the source of the feedback, and in response to a more open question (‘Did you notice anything strange or unexpected during the task? If so, what?’) only two participants expressed doubt over whether the feedback was provided by the others.

## Computational Modelling

Model Q16 fit the non-social data similarly well as the best fitting model (Q5), and the numerical values of the mean learning rate and memory decay parameters for Q16 were similar for the social and non-social condition. Thus, three further models were fit in which either or both of the latter parameters were shared between the conditions. AIC weights indicated that the model in which all parameters varied between conditions provided the best fit (AIC weight = 0.293), closely followed by the model in which the memory decay parameter was shared (AIC weight = 0.292). The models in which the learning rate (AIC weight = 0.190) or both learning rate and memory decay parameters (AIC weight = 0.226) were shared fit slightly less well. The analysis in the main paper (and below) thus focuses on model Q16 with no shared parameters.

### Parameter Estimates

Mann-Whitney U tests on parameters form the social condition (for model Q16) found significantly lower learning rates (*U* = 1277, *p* = 0.040) in HD compared to LD subjects, while no group differences were observed for the outcome valuation (*U* = 1095, *p* = 0.549), memory decay (*U* = 1047, *p* = 0.829), or temperature (*U* = 1099, *p* = 0.528) parameters. In the non-social condition (for model Q5), there were no significant group differences in the learning rate (*U* = 1025, *p* = 0.968), choice bias (*U* = 1105, *p* = 0.497), outcome valuation (*U* = 1035, *p* = 0.905), or temperature (*U* = 885, *p* = 0.280) parameters. Although the numerical group difference in the learning rate value went in the same direction for the social and non-social condition, the main effect of group (based on the average learning rate across conditions) did not reach significance (*U* = 1222, *p* = 0.106).

To assess the robustness of our findings, we additionally examined parameter group differences for similarly well-fitting models, namely Q4 for the social condition, and Q4 and Q16 for the non-social condition. For model Q4, Mann-Whitney U tests on parameters form the social condition found significantly lower learning rates (*U* = 1314, *p* = 0.019) in HD subjects compared to LD controls, while no group differences were observed for the outcome valuation (*U* = 1157, *p* = 0.273) or temperature (*U* = 1242, *p* = 0.076) parameters. In the non-social condition, there were no significant group differences in the learning rate (Q4: *U* = 1019, *p* = 0.994; Q16: *U* = 1017, *p* = 0.981 ), outcome valuation (Q4: *U* = 1094, *p* = 0.554; Q16: *U* = 1097, *p* = 0.538), or temperature (Q4: *U* = 876, *p* = 0.250; Q16: *U* = 854, *p* = 0.184) parameters. The consistency of the findings across models provides evidence for the robustness of our results.

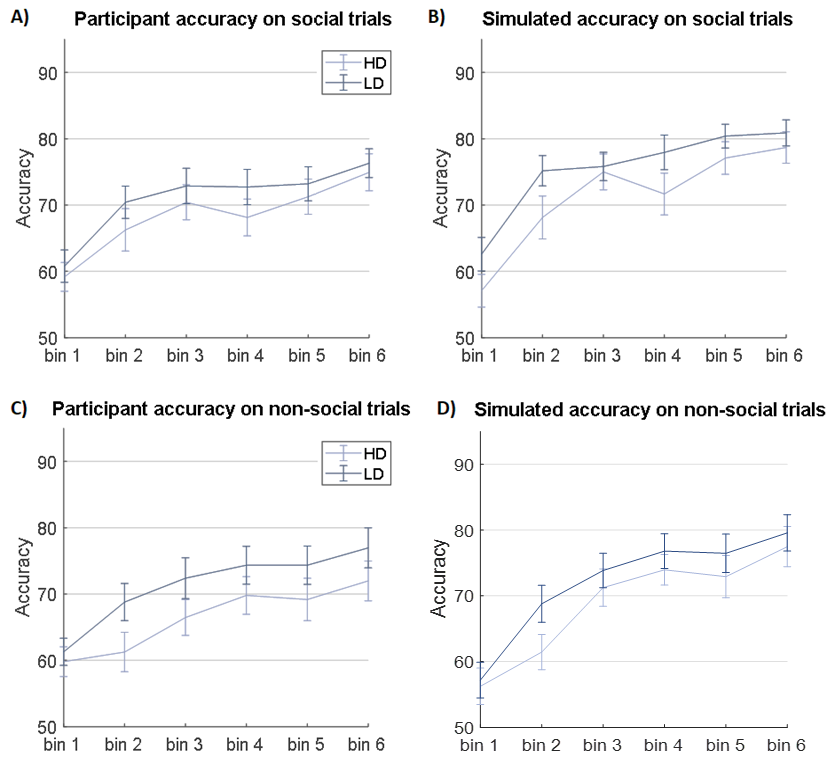
Similarly, we confirmed the robustness of the regression findings reported in the main paper by using the parameters of other well-fitting models as predictors. For model Q4, this analysis revealed a significant association of social model parameter values and questionnaire measures with the reported amount of time spent in unpleasantly perceived social situations (*L*(5) = 16.80, *p* = 0.005, R2 = 0.20). This predictive relation was driven by outcome valuation (Q4: β = 0.29, *p* = 0.025) and learning rate (β = -0.31, *p* = 0.020) parameter values, as well as BDI scores (β = 0.28, *p* = 0.017). By contrast, SAQ social anxiety scores (Q4: β = 0.15, *p* = 0.205), and negative biases (Q4: β = 0.09, *p* = 0.374) had no significant effect. These results were highly similar to those obtained when using parameters from model Q16 (see main paper), thus providing evidence for the robustness of the observed effect.

By contrast, a somewhat different pattern of results was obtained when using model Q4 (instead of Q16) parameters as predictors of the amount of time spent in pleasantly perceived social situations (*L*(4) = 19.32, *p* = 0.001, R2 = 0.23). While RSAS social anhedonia scores similarly had a significant predictive effect (β = -0.46, *p* < 0.001), it was the outcome valuation parameters that additionally made a significant contribution (β = 0.24, *p* = 0.030), whereas temperature parameters (β = 0.10, *p* = 0.377), and BDI scores (β = 0.02, *p* = 0.863) had no significant effect. The same pattern of results was observed when learning rate values (β = 0.08, *p* = 0.547) were included as predictors instead of temperature parameters. It should be noted that the fact that the regression results vary depending on which model parameters are used indicates that the findings are not robust and should thus be interpreted with caution.

### Model Validation

In terms of model validation (based on models Q16 and Q5 for the social and non-social conditions, respectively), it can be seen from Figure S1 that, although the overall accuracy was slightly overestimated, the relative accuracy pattern of the two groups in the simulated data closely resembled that of the real data. Additionally, *pseudo-R2* values indicated that in both the social (*pseudo-R2* = 0.34) and non-social (*pseudo-R2* = 0.33) condition the model provided a relatively good fit for the data, and no group differences in *pseudo-R2* values were observed in either condition (*U* = 1116, *p* = 0.443 and *U* = 1229, *p* = 0.095, respectively). Moreover, in both the social and non-social condition, participants’ parameter values and the parameter estimates from the simulated data were significantly correlated (see Table S2), and, for the social condition, group differences in the learning rate could be recovered from the simulated data (*U* = 1349, *p* = 0.009).

***Figure S1:*** *Percent of accurate choices in six bins of twelve trials for participants with high (HD) and low (LD) depression scores on A) social and C) non-social trials, as well as for the data simulated using the parameters from the best fitting models on B) social and D) non-social trials.*

****

***Table S1:*** *Overview of the models fit to the data and the associated Akaike's Information Criterion (AIC) weights for the social and non-social condition.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Learning rate  (αG, αL) | Outcome valuation (d) | | Choice bias (φ) | | Choice bias decay (*γ*) | | Memory decay (ω) | | | Tempe-rature (τ) | AIC weight  social | | AIC weight  non-social |
| Q1 | 1 | |  | |  | |  | |  | x | | 0.007 | 0.010 | |
| Q2 | 1 | |  | | x | |  | |  | x | | 0.018 | 0.023 | |
| Q3 | 1 | |  | | x | | x | |  | x | | 0.026 | 0.024 | |
| Q4 | 1 | | x | |  | |  | |  | x | | 0.086 | 0.087 | |
| Q5 | 1 | | x | | x | |  | |  | x | | 0.066 | 0.096 | |
| Q6 | 1 | | x | | x | | x | |  | x | | 0.066 | 0.056 | |
| Q7 | 2 | |  | |  | |  | |  | x | | 0.030 | 0.023 | |
| Q8 | 2 | |  | | x | |  | |  | x | | 0.046 | 0.031 | |
| Q9 | 2 | |  | | x | | x | |  | x | | 0.035 | 0.037 | |
| Q10 | 2 | | x | |  | |  | |  | x | | 0.078 | 0.067 | |
| Q11 | 2 | | x | | x | |  | |  | x | | 0.075 | 0.057 | |
| Q12 | 2 | | x | | x | | x | |  | x | | 0.035 | 0.033 | |
| Q13 | 1 | |  | |  | |  | | x | x | | 0.005 | 0.007 | |
| Q14 | 1 | |  | | x | |  | | x | x | | 0.011 | 0.015 | |
| Q15 | 1 | |  | | x | | x | | x | x | | 0.017 | 0.018 | |
| Q16 | 1 | | x | |  | |  | | x | x | | 0.089 | 0.080 | |
| Q17 | 1 | | x | | x | |  | | x | x | | 0.062 | 0.073 | |
| Q18 | 1 | | x | | x | | x | | x | x | | 0.021 | 0.055 | |
| Q19 | 2 | |  | |  | |  | | x | x | | 0.023 | 0.012 | |
| Q20 | 2 | |  | | x | |  | | x | x | | 0.029 | 0.023 | |
| Q21 | 2 | |  | | x | | x | | x | x | | 0.026 | 0.024 | |
| Q22 | 2 | | x | |  | |  | | x | x | | 0.066 | 0.060 | |
| Q23 | 2 | | x | | x | |  | | x | x | | 0.042 | 0.054 | |
| Q24 | 2 | | x | | x | | x | | x | x | | 0.042 | 0.036 | |

For models Q1 to Q6 and Q13 to Q18, the same learning rate was used for positive and negative prediction errors (i.e. αG = αL), while for the remaining models separate learning rates were utilised. An x indicates that the parameter was estimated in the model, while the other parameters were fixed at 0 (for choice bias and decay) or removed (for d, with r(t) being set to 1, 0, and -1 for positive, neutral, and negative outcomes, respectively).

***Table S2:*** *Correlations between participant parameters used for data simulation and parameters recovered from the simulated data for the social and non-social condition*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **social** | | **non-social** | |
|  | *rs value* | *p value*  *(one-tailed)* | *rs value* | *p value*  *(one-tailed)* |
| learning rate (α) | 0.54 | <0.001 | 0.39 | <0.001 |
| choice bias (φ) | N/A | N/A | 0.52 | <0.001 |
| outcome valuation (d) | 0.42 | <0.001 | 0.45 | <0.001 |
| memory decay (ω) | 0.32 | 0.001 | N/A | N/A |
| temperature (τ) | 0.53 | <0.001 | 0.61 | <0.001 |

# References

**Collins AGE, Frank MJ** (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience* **35**, 1024–1035.

**Daw ND** (2011). Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning: Attention and Performance XXIII* Eds MR Delgado, EA Phelps & TW Robbins, pp3–38. Oxford University Press: Oxford.

**Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE** (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences* **104**, 16311–16316.

**Gold JM, Waltz JA, Matveeva TM, Kasanova Z, Strauss GP, Herbener ES, Collins AGE, Frank MJ** (2012). Negative Symptoms and the Failure to Represent the Expected Reward Value of Actions. *Archives of General Psychiatry* **69**, 129.

**Schonberg T, Daw ND, Joel D, O’Doherty JP** (2007). Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making. *Journal of Neuroscience*

**Wagenmakers EJ, Farrell S** (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* **11**, 192–196.