**Supplementary Material**

A TRIAL CONSORT

Figure A.1: Trial profile. The dashed arrows indicate that participants with missing data at certain sessions remained in the trial. GP=general practitioner. GAD-7=Generalised Anxiety Disorder Assessment.

# B REINFORCEMENT LEARNING MODELLING

## B.1 RL MODELS

|  |  |  |
| --- | --- | --- |
| Model | # Parameter | Description |
| random baseline model | 1 | Only bias model. |
| RW | 2 | Basic reinforcement learning model including learning rate and outcome sensitivity. |
| RW + noise | 3 | RW model with irreducible noise. |
| RW + noise + bias | 4 | RW model with irreducible noise and a constant bias towards the ’go’ action. |
| RW + Pav + noise + bias | 5 | RW model with irreducible noise, a constant bias towards the ’go’ action, and a positive Pavlovian bias parameter. |
| RW + Pav(rew/pun) + noise + bias | 6 | RW model with irreducible noise, a constant bias towards the ’go’ action, a joint reward/loss sensitivity, and separate positively constrained Pavlovian bias parameters for rewards and losses. |
| RW(rew/pun) + Pav(rew/pun) + noise + bias | 7 | RW model with irreducible noise, a constant bias towards the ’go’ action, separate reward and loss sensitivities, and separate positively constrained Pavlovian bias parameters for rewards and losses. |
| RW(rew/pun) + Pav(rew/pun) + noise + bias | 8 | RW model with irreducible noise, a constant bias towards the ’go’ action, separate reward and loss learning rates, separate reward and loss sensitivities, and separate positively constrained Pavlovian bias parameters for rewards and losses. |

We build on models from Guitart-Masip *et al.* (2012) to fit participants’ behaviour. All eight models are reinforcement learning models which assign a probability to each action on each trial . The probability is derived by passing an action weight that depends on the stimulus on trial through a softmax and introducing a lapse parameter referring to noisiness of the response:

The action weight is constructed differently for the different models. In the simpler models, the action weight equals the Q-value. The more complex model include a constant bias towards the ’go’ action and a Pavlovian factor . In conditions where participants received a punishment, the Pavlovian parameter decreased the ’go’ tendency in proportion to the stimulus value , whereas it promotes the action in rewarding conditions.

The Q-value follows a Rescorla-Wagner update equation. refers to the learning rate, is the reward sensitivity, and each trial feedback is defined as. The stimulus value is updated in the same way.

In the most complex model the free parameters *,* , and could be different for rewarding and punishing contexts.

## B.2 MODEL FITTING PROCEDURE

We use the same model fitting procedure as described in Huys *et al.* (2011*a*). A vector of parameters *h* can be inferred for each model and each subject by calculating the maximum a posteriori estimate using a Gaussian prior and the likelihood, where are all actions of subject :

We can factorize over trials, as we assume actions to be independent given the stimuli (not included in equation). The job of the prior distribution is to regularize the parameter values. To infer the parameters of the prior distribution (hyperparameters ) we maximize the likelihood given all data by all subjects :

As shown in Equation 6, we have to integrate over individual parameter values. This integral can be solved using Expectation-Maximization algorithm (MacKay, 2003): For the E-step we use the following Laplacian approximation at the iteration:

E-step:

Further, in the M-step the mean and variance of the prior distribution was calculated as follows:

M-step:

Before inference, all parameters were suitably transformed to enforce constraints (log and inverse sigmoid transforms). Huys *et al.* (2011*a*) verified the model fitting procedure on generated data from a known decision process.

## B.3 MODEL COMPARISON

After fitting each model separately to the data, we evaluated which model fits the data best taking the varying flexibility of the models into account. We used the Bayesian Information Criterion (BIC) to compare the models introducing a penalty term for the number of parameters in the model.

It is labeled integrated BIC (iBIC) because to infer the log-likelihood distribution , we had to integrate over the individual parameters of each subject. These integrals were approximated using sampling (times) from the prior distribution (MacKay, 2003).

Comparing integrated group-level BIC values is similar to a likelihood ratio test and can be shown to reduce to classical statistical tests for simple linear models (Kass and Raftery, 1995).

## B.4 SPECIFY INFORMATIVE DATA

To assess the extent to which the observed behavioral data was informative, we conducted a comparison between the integrated likelihood of the most parsimonious model and the integrated likelihood of a random baseline model for each dataset, individual, and session. The integrated likelihood, which incorporates individual parameters, represents the likelihood of the data given the hyperparameters at the group level. If the integrated likelihood of the random baseline model exceeded the integrated likelihood of the most parsimonious model by more than threefold, the corresponding task run was considered as missing.

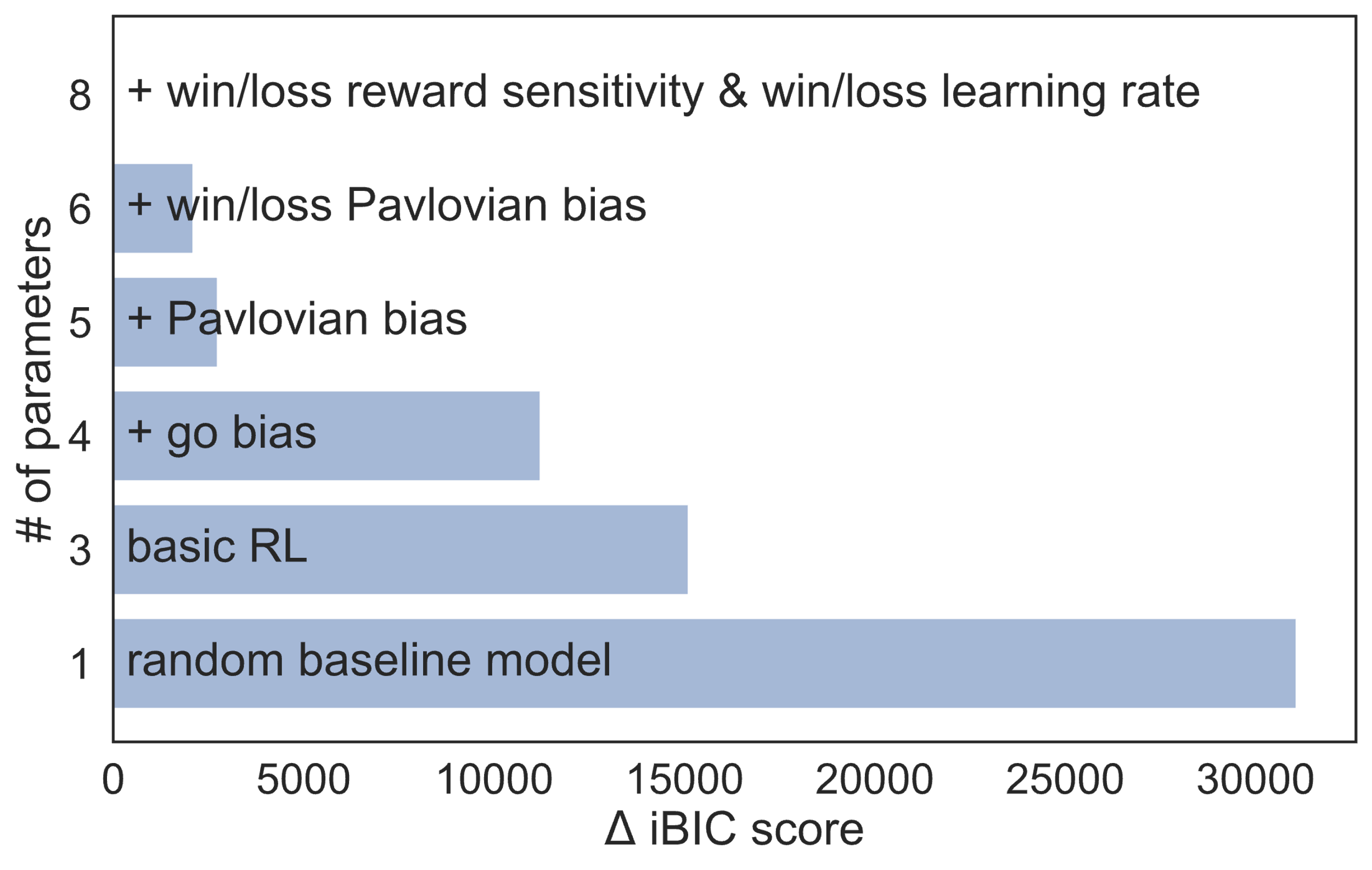


Figure B.2: The model selection process conducted on the reduced dataset using iBIC yielded congruent outcomes when compared to the results obtained from the complete dataset. This figure shows the differences in integrated iBIC scores for all models tested compared to the most parsimonious model using only informative task runs. A smaller iBIC score indicates a more parsimonious model. The y-axis shows the number of free parameters for each model. The most parsimonious model includes separate learning rates for rewards and punishments, win and loss sensitivities, appetitive and aversive Pavlovian biases, irreducible noise, and a constant bias factor added to the action-value for ’go’.

## B.5 PARAMETER RECOVERY

For each parameter of the most parsimonious model, we sampled parameter values from a normal distribution defined by the mean and standard deviation of the hyperprior over parameter estimates from empirical data, which we used to simulate behaviour for N=200. For each simulated dataset, we fitted the model and analysed the correlation between the simulated and the re-estimated parameters. Here we report the average confusion matrix over 100 times this procedure was performed.

A screenshot of a computer screen

Description automatically generated

Figure B.3: Confusion matrix of parameter recovery.

## B.6 TEST-RETEST RELIABILITY

To evaluate test-retest reliability, we calculated Pearson correlation of individuals’ parameters between the different time points and employed intra-class correlation coefficients (ICCs; McGraw and Wong 1996) only using the informative data (cf. Fig. B.4 and Table B.1) and specifically in the placebo group (cf. Fig. B.5 and Table B.2). ICCs assess the ratio of intra-individual to inter-individual variability. We used a two-way mixed effects model based on single measures ICC. The fixed effect in the model was the testing time-interval, while the random effect was the subject. In this report, we present the consistency ICCs, which assess the relative ranking of participants over time. ICC values range from 0 to 1, with zero indicating low or no reliability and 1 representing perfect reliability. In general, ICCs below 0.4 are considered indicative of poor reliability, while ICCs between 0.4 and 0.75 suggest moderate to good reliability. ICCs above 0.75 are indicative of excellent reliability. Only the aversive Pavlovian parameters showed moderate reliability and all other parameter estimates showed poor to low or no reliability.

A screenshot of a computer screen

Description automatically generated

Figure B.4: Pearson correlation of parameter estimates at different measurement points only based on informative data.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| parameter estimate | ICC | F-Test | | | | 95% Confidence Interval | |
|  |  | F | df1 | df2 | p-value | lower bound | upper bound |
| reward sensitivity | -0.02 | 0.93 | 163 | 326 | 0.69 | -0.10 | 0.07 |
| loss sensitivity | 0.23 | 1.89 | 163 | 326 | *<*0.01 | 0.13 | 0.33 |
| reward learning rate | 0.09 | 1.30 | 163 | 326 | 0.02 | 0.00 | 0.19 |
| loss learning rate | 0.19 | 1.71 | 163 | 326 | *<*0.01 | 0.10 | 0.29 |
| appetitive Pavlovian bias | 0.31 | 2.34 | 163 | 326 | *<*0.01 | 0.21 | 0.41 |
| aversive Pavlovian bias | 0.47 | 3.69 | 163 | 326 | *<*0.01 | 0.38 | 0.56 |
| noise | 0.11 | 1.38 | 163 | 326 | 0.01 | 0.02 | 0.21 |
| go bias | 0.33 | 2.51 | 163 | 326 | *<*0.01 | 0.24 | 0.43 |

Table B.1: Test-retest reliability of informative parameter estimates based on ICC(3,1).

A screenshot of a computer screen

Description automatically generated

Figure B.5: Pearson correlation of parameter estimates at different measurement points only based on informative data from the placebo group.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| parameter estimate | ICC | F-Test | | | | 95% Confidence Interval | |
|  |  | F | df1 | df2 | p-value | lower bound | upper bound |
| reward sensitivity | -0.13 | 0.64 | 94 | 188 | 0.99 | -0.22 | -0.03 |
| loss sensitivity | 0.19 | 1.71 | 94 | 188 | *<*0.01 | 0.07 | 0.33 |
| reward learning rate | -0.01 | 0.98 | 94 | 188 | 0.54 | -0.11 | 0.12 |
| loss learning rate | 0.22 | 1.84 | 94 | 188 | *<*0.01 | 0.09 | 0.35 |
| appetitive Pavlovian bias | 0.35 | 2.64 | 94 | 188 | *<*0.01 | 0.23 | 0.48 |
| aversive Pavlovian bias | 0.53 | 4.38 | 94 | 188 | *<*0.01 | 0.41 | 0.64 |
| noise | 0.1 | 1.32 | 94 | 188 | 0.06 | -0.02 | 0.23 |
| go bias | 0.27 | 2.09 | 94 | 188 | *<*0.01 | 0.14 | 0.40 |

Table B.2: Test-retest reliability of informative parameter estimates based on ICC(3,1) only in the placebo group.

## B.7 CORRELATION BETWEEN PARAMETERS

Finally, the RL model parameters exhibited a trade-off among themselves, as evidenced by the significant correlations between them. Notably, the parameters associated with aversive conditions seemed to encapsulate similar aspects of the data. Due to the inherent entanglement of cognitive parameters, it proved challenging to ascertain their precise significance within the available GoNogo task data.

A screenshot of a graph

Description automatically generated

Figure B.6: Pearson correlation between different parameter estimates at the three measurement points including only informative data.

# C FINDINGS IN THE WHOLE SAMPLE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coef. | Std.Err. | z | P*>*|*z*| | [0.025 | 0.975] |
| drug effect on | 0.32 | 0.15 | 2.18 | 0.03 | 0.032 | 0.607 |
| drug effect on | 0.32 | 0.19 | 1.70 | 0.09 | -0.048 | 0.692 |
| drug effect on | -0.35 | 0.20 | -1.75 | 0.08 | -0.741 | 0.042 |
| relation between GAD and | 0.01 | 0.00 | 2.26 | 0.02 | 0.001 | 0.016 |
| effect of on BDI*t4* | 0.05 | 0.02 | 2.15 | 0.03 | 0.004 | 0.087 |
| interaction effect of  on BDI*t4* | 0.01 | 0.03 | 0.47 | 0.64 | -0.044 | 0.071 |

Table C.3: Re-performing analyses on whole sample including non-informative task runs to examine whether main findings can be confirmed using the whole dataset. The effect of sertraline on aversive learning rate at week 2 appeared to be robust, whereas we found only trends towards the effects of sertraline on change of aversive learning rate. The relation between anxiety score and aversive learning rate was significant in the whole sample. Finally, we found evidence for a positive correlation between the early change in aversive Pavlovian bias and depression score at week 12, however, there was no evidence for the interaction effect with group in the whole sample.

# D PREREGISTRATION

Table D.4: Design Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Power Analysis** | **Analysis Plan** | **Interpretation given different outcome** |
| **1)** Is the SSRI sertraline related to the aversive  Pavlovian bias? | Group allocation (sertraline vs. placebo) is associated with the aversive Pavlovian bias. | A standardized effect size of 0.23 is required to be detected with 95% power (*p <* 0*.*05) using a sample size of N = 439 including missing data in the *dependent* variable at different sessions (30% missing at baseline, 32% at follow-up 1 (2 weeks), and 36% at follow-up 2 (6 weeks)). The required effect size was calculated using 1000 simulated datasets from a mixedeffects model for different standardized effect sizes (*d* ∈ [0*,*1]). Standard error and standard deviation of the random intercepts were set to 1. | Mixed-effects  model using group allocation as independent variable and aversive Pavlovian bias as dependent variable allowing for random intercepts and controlling for stratification variables (baseline total CIS-R score in three categories, duration of depressive episode in two categories, and site). | 1. No effect: no evidence that sertraline and aversive Pavlovian bias are related. 2. Significant fixed effect *>* 0: Sertraline increases the aversive Pavlovian bias. 3. **c)** Significant fixed effect *<* 0: Sertraline decreases the aversive Pavlovian bias. |

*Continued on next page*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Power Analysis** | **Analysis Plan** | **Interpretation given different outcome** |
| **2)** Is the aversive Pavlovian bias related to anxiety? | The aversive  Pavlovian bias is associated with anxiety. | A standardized effect size of 0.16 is required to be detected with 95% power (*p <* 0*.*05) using a sample size of N = 439 including missing data in the *independent* variable at different sessions (30% missing at baseline, 32% at follow-up 1 (2 weeks), and 36% at follow-up 2 (6 weeks)). The required effect size was calculated using 1000 simulated datasets from mixedeffects model for different effect sizes (*d* ∈ [0*,*1]). Standard error and standard deviations of the random slopes and intercepts were set to 1. | Mixed-effects  model using the aversive Pavlovian bias as independent variables and log-transformed total GAD-7 score as dependent variable allowing for random slopes and random intercepts and controlling for group allocation and various possible confounders (baseline variables). | **a)** No effect: no evidence that aversive Pavlovian bias and anxiety are related.  **b)** Significant fixed effect *>* 0: Aversive Pavlovian bias and anxiety are positively related. **c)** Significant fixed effect  *<* 0: Aversive Pavlovian bias and anxiety are negatively related. |

*Continued on next page*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Power Analysis** | **Analysis Plan** | **Interpretation given different outcome** |
| **3)** Does the aversive Pavlovian bias mediate the effect of sertraline on anxiety? | The aversive Pavlovian bias mediates the effect of sertraline on anxiety. | A mediation effect size of 0.06 is required to be detected with 95% power (*p <* 0*.*05) based on the  Sobel test Sobel (1982). Standard error and standard deviation of the predictor and the mediator were set to 1. | A potential outcome mediation analysis in a multilevel setting will be performed.  First, we will take the fixed effect of sertraline on the aversive Pavlovian bias from Hypothesis 1 (E1). Second, we will adapt Hypothesis 2 by controlling for an interaction between group allocation and the aversive Pavlovian bias and extract the fixed effect of the aversive Pavlovian bias on anxiety (E2) and the fixed effect of the interaction term on anxiety (E3). Two indi-  rect/mediation  effects will be calculated (i) E1 \* E2, and (ii) E1 \* E2 + E1 \* E3.  Their confidence intervals will be inferred using bootstrapping.  The effects are significant if the CIs do not contain zero. | **a)** No effect: no evidence that the aversive Pavlovian bias mediates the effect of sertraline on anxiety.  **b)** Significant mediation effect: As the effect of the aversive Pavlovian bias on anxiety will only be correlational, a significant mediation effect will not provide evidence for a causal mediation. Therefore, a timelagged regression will be performed using the aversive Pavlovian bias at previous timepoints (baseline, at 2 and 6 weeks) to predict anxiety at the next timepoint (12 weeks), controlling for baseline variables. If we find an effect there, this will support the direction from the aversive Pavlovian bias to anxiety and not vice versa. If we will not find an effect there, we cannot make any claims about a causal relationship between the aversive Pavlovian bias and anxiety or the mediation effect. |

*Continued on next page*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Power Analysis** | **Analysis Plan** | **Interpretation given different outcome** |
| **4)** Can the baseline aversive Pavlovian bias predict treatment outcome? | The slopes of the baseline aversive Pavlovian bias predicting anxiety differ between groups (sertaline vs placebo). | A standardized effect size of the difference in the slopes of 0.38 is required to be detected with 95% power  (*p <* 0*.*05) using a sample size of N = 300 (150 in each group). This sample size is based on the fact that 30% of the independent variable is missing at baseline. The required effect size was calculated using G\*Power  Faul *et al.* (2007,  2009) using linear bivariate regression: Two groups, difference between slopes (all standard deviations set to 1). | Linear regression regressing the  log-transformed GAD-7 total sum score measured at 12 weeks on the *baseline* aversive Pavlovian bias including the interaction effect between baseline aversive Pavlovian bias and group allocation to compare the regression coefficient between groups. | **a)** No effect: No evidence for difference in the slopes.  **b)** Significant interaction effect: Influence of the baseline aversive Pavlovian bias differs between groups. The baseline aversive Pavlovian bias can predict treatment outcome. |
| **5)** Is the appetitive Pavlovian bias related to depression? | The appetitive Pavlovian bias is associated with the log-  transformed  PHQ-9 total sum score. | A standardized effect size of 0.16 is required to be detected with 95% power (*p <* 0*.*05) using a sample size of N = 439 based on the same power simulation as for Question 2). | Mixed-effects  model using appetitive Pavlovian bias as independent variable and log-transformed  total PHQ-9 scores as dependent variable allowing for random slopes and random intercepts and controlling for group allocation and various confounders (baseline variables). | **a)** No effect: no evidence that the appetitive Pavlovian bias is related to depression.  **b)** Significant fixed effect *>* 0: The appetitive Pavlovian bias is positively associated with depression.  **c)** Significant fixed effect *<* 0: The appetitive Pavlovian bias is negatively associated with depression. |

*Continued on next page*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Power Analysis** | **Analysis Plan** | **Interpretation given different outcome** |
| **6)** Is reward sensitivity related to anhedonia? | Reward sensitivity is associated with the anhedonia score from the PHQ-9. | A standardized effect size of 0.16 is required to be detected with 95% power (*p <* 0*.*05) using a sample size of N = 439 based on the same power simulation as for Question 2). | Mixed-effects  model using reward sensitivity parameters as independent variable and  log-transformed anhedonia scores as dependent variable allowing for random slopes and random intercepts and controlling for group allocation and various confounders  (baseline variables). | **a)** No effect: no evidence that reward sensitivity is related to anhedonia.  **b)** Significant fixed effect *>* 0: Reward sensitivity is positively associated with anhedonia.  **c)** Significant fixed effect *<* 0: Reward sensitivity is negatively associated with anhedonia. |

E ADDITIONAL TABLES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2-week follow-up | | 6-week follow-up | |
|  | **coef** | **p-value** | **coef** | **p-value** |
| Age (years) | -0.01 | 0.064 | -0.01 | 0.218 |
| GAD-7 | 1.37 | 0.003 | 0.99 | 0.010 |
| PHQ-9 | 0.88 | 0.078 | 0.56 | 0.176 |
| BDI | 1.55 | 0.009 | 0.62 | 0.185 |
| Site | 0.32 | 0.001 | 0.37 | *<*0.001 |
| CIS-R total score | 0.35 | 0.023 | 0.18 | 0.150 |
| CIS-R depression duration (years) | 0.07 | 0.758 | 0.27 | 0.182 |
| Highest educational qualification | 0.19 | 0.309 | 0.11 | 0.490 |
| Antidepressants in the past | -0.45 | 0.041 | -0.06 | 0.739 |
| Gender | 0.14 | 0.529 | 0.14 | 0.465 |
| Ethnicity | 0.28 | 0.012 | 0.20 | 0.057 |
| Financial difficulty | 0.42 | 0.005 | 0.24 | 0.067 |
| Employment status | 0.18 | 0.440 | 0.05 | 0.820 |
| Marital status | 0.14 | 0.373 | 0.09 | 0.525 |
| Treatment group allocation | 0.11 | 0.608 | 0.23 | 0.236 |

Table E.5: Univariate logistic regression was used to analyze the relationship between baseline variables and follow-up missing data. Data was missing due to not completing the GoNogo task or GAD-7.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2-week follow-up | | 6-week follow-up | |
|  | **coef** | **p-value** | **coef** | **p-value** |
| Age (years) | 0.06 | *<*0.001 | 0.06 | *<*0.001 |
| GAD-7 | -0.29 | 0.376 | -0.98 | 0.005 |
| PHQ-9 | -0.32 | 0.383 | -0.62 | 0.107 |
| BDI | -0.57 | 0.165 | -0.72 | 0.099 |
| Site | 0.01 | 0.851 | 0.07 | 0.37 |
| CIS-R total score | -0.19 | 0.088 | -0.32 | 0.006 |
| CIS-R depression duration (years) | 0.07 | 0.712 | -0.14 | 0.472 |
| Highest educational qualification | 0.92 | *<*0.001 | 1.15 | *<*0.001 |
| Antidepressants in the past | 0.61 | 0.001 | 0.76 | *<*0.001 |
| Gender | 0.29 | 0.106 | 0.05 | 0.782 |
| Ethnicity | -0.05 | 0.692 | 0 | 0.984 |
| Financial difficulty | 0.05 | 0.686 | -0.13 | 0.325 |
| Employment status | 0.12 | 0.51 | 0.52 | 0.008 |
| Marital status | -0.01 | 0.951 | 0.1 | 0.454 |
| Treatment group allocation | 0.48 | 0.006 | -0.01 | 0.969 |

Table E.6: Univariate logistic regression was used to analyze the relationship between baseline variables and non-informative task runs. Significant predictors of non-informative data were used as covariates in all statistical analyses.

# F SUBSAMPLE ANALYSES

To explore potential factors contributing to the null results concerning our preregistered hypotheses, we conducted analyses on two subsamples.

First, we examined a subsample comprising individuals with mild depression, aiming to determine whether the clinical nature of this sample could potentially account for the null results. This consideration is pertinent because the studies serving as the basis for our preregistered hypotheses predominantly involve healthy individuals. However, our analysis of patients with a baseline PHQ-9 score below 10 did not yield any significant evidence of a sertraline effect on the aversive Pavlovian bias (cf. Table F.7). Additionally, there was no evidence supporting the remaining preregistered hypotheses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **H1) effect of sertraline on aversive Pavlovian bias** | | | | |
|  | **placebo mean (SEM)** | **sertraline mean (SEM)** | **mean difference [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| baseline | -0.53 (0.13) | -0.60 (0.10) |  |  |
| 2 | -0.49 (0.11) | -0.77 (0.09) | -0.23 [-0.51,0.05] | 0.10 |
| 6 | -0.83 (0.12) | -0.86 (0.11) | -0.02 [-0.33,0.30] | 0.91 |
| over time | … | … | -0.09 [-0.30,0.12] | 0.41 |
| group by time interaction | … | … | … | 0.99 |
|  |  |  |  |  |
| **H2) association between aversive Pavlovian bias and log-transformed GAD-7 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.02 [-0.07,0.03] | 0.42 |
| 6 |  |  | -0.03 [-0.07,0.02] | 0.28 |
| over time |  |  | -0.03 [-0.07,0.02] | 0.22 |
|  |  |  |  |  |
| **H4) association between baseline aversive Pavlovian bias and log-transformed GAD-7 total score at week 12** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
|  |  |  | -0.05 [-0.13,0.03] | 0.25 |
|  |  |  |  |  |
| **H5) association between appetitive Pavlovian bias and log-transformed PHQ-9 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.04 [-0.10,0.01] | 0.13 |
| 6 |  |  | -0.06 [-0.12,0.01] | 0.08 |
| over time |  |  | -0.04 [-0.09,0.01] | 0.13 |
|  |  |  |  |  |
| **H6) association between reward sensitivity and log-transformed PHQ-9 anhedonia item score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | 0.02 [-0.05,0.09] | 0.62 |
| 6 |  |  | 0.02 [-0.05,0.09] | 0.62 |
| over time |  |  | 0.03 [-0.03,0.10] | 0.31 |

Table F.7: Prergistered Hypotheses were tested on a subsample of patients (N=161) with mild depression at baseline (baseline PHQ *<* 10).

Second, we attempted to understand the contribution of poor test-retest measurements. We conducted an analysis within the subset characterized by an absolute slope of the aversive Pavlovian response over the three sessions measuring below 0.2, indicating a minor change in the aversive Pavlovian bias. In this context as well, our analysis did not yield substantial evidence to support the notion of sertraline significantly affecting aversive Pavlovian bias (cf. Table F.8) nor did we find support for any of the other hypotheses that were preregistered.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **H1) effect of sertraline on aversive Pavlovian bias** | | | | |
|  | **placebo mean (SEM)** | **sertraline mean (SEM)** | **mean difference [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| baseline | -0.61 (0.12) | -0.70 (0.11) |  |  |
| 2 | -0.60 (0.10) | -0.76 (0.12) | -0.15 [-0.40,0.09] | 0.22 |
| 6 | -0.66 (0.12) | -0.68 (0.10) | -0.03 [-0.14,0.90] | 0.62 |
| over time | … | … | -0.09 [-0.26,0.08] | 0.29 |
| group by time interaction | … | … | … | 0.75 |
|  |  |  |  |  |
| **H2) association between aversive Pavlovian bias and log-transformed GAD-7 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | 0.02 [-0.04,0.07] | 0.54 |
| 6 |  |  | -0.03 [-0.09,0.03] | 0.38 |
| over time |  |  | -0.01 [-0.06,0.04] | 0.75 |
|  |  |  |  |  |
| **H4) association between *baseline* aversive Pavlovian bias and log-transformed GAD-7 total score at week 12** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
|  |  |  | -0.01 [-0.12,0.01] | 0.88 |
|  |  |  |  |  |
| **H5) association between appetitive Pavlovian bias and log-transformed PHQ-9 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | 0.01 [-0.04,0.07] | 0.63 |
| 6 |  |  | 0.00 [-0.07,0.07] | 0.99 |
| over time |  |  | 0.01 [-0.04,0.06] | 0.63 |
|  |  |  |  |  |
| **H6) association between reward sensitivity and log-transformed PHQ-9 anhedonia item score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.01 [-0.11,0.08] | 0.76 |
| 6 |  |  | -0.01 [-0.11,0.09] | 0.85 |
| over time |  |  | -0.02 [-0.11,0.07] | 0.66 |

Table F.8: Prergistered Hypotheses were tested on a subsample of patients (N=86) with high test-retest stability of the aversive Pavlovian.

# G POST-HOC ANALYSES

As suggested by reviewers, it is plausible that some of the observed negative results may be influenced by medication changes or other interventions, such as psychotherapy. These variables were highly heterogeneous, making it challenging to derive meaningful statistics. Nevertheless, we performed supplementary analyses, accounting for the use of other antidepressants and/or psychotherapy during the trial, adherence score (measured using a five-item self-report adherence scale developed for the CoBalt study (Wiles *et al.*, 2013)), and the number of tablets (sertraline or placebo). The inclusion of those variables did not alter the reported results (cf. Table G.9 and G.10).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **H1) effect of sertraline on aversive Pavlovian bias** | | | | |
|  | **placebo mean (SEM)** | **sertraline mean (SEM)** | **mean difference [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| baseline | -0.50 (0.07) | -0.55 (0.07) |  |  |
| 2 | -0.55 (0.06) | -0.71 (0.07) | -0.09 [-0.27,0.10] | 0.35 |
| 6 | -0.77 (0.07) | -0.70 (0.07) | 0.13 [-0.06,0.32] | 0.19 |
| over time | … | … | 0.02 [-0.12,0.16] | 0.82 |
| group by time interaction | … | … | … | 0.24 |
|  |  |  |  |  |
| **H2) association between aversive Pavlovian bias and log-transformed GAD-7 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.01 [-0.04,0.01] | 0.35 |
| 6 |  |  | -0.02 [-0.05,0.01] | 0.17 |
| over time |  |  | -0.02 [-0.04,0.00] | 0.06 |
|  |  |  |  |  |
| **H4) association between *baseline* aversive Pavlovian bias and log-transformed GAD-7 total score at week 12** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
|  |  |  | -0.02 [-0.07,0.03] | 0.44 |
|  |  |  |  |  |
| **H5) association between appetitive Pavlovian bias and log-transformed PHQ-9 total score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.01 [-0.04,0.02] | 0.54 |
| 6 |  |  | -0.03 [-0.06,0.01] | 0.10 |
| over time |  |  | -0.02 [-0.05,0.01] | 0.13 |
|  |  |  |  |  |
| **H6) association between reward sensitivity and log-transformed PHQ-9 anhedonia item score** | | | | |
|  |  |  | **regression coefficient [95% CI]** | **p-value** |
| Follow-up assessments (weeks) | | | | |
| 2 |  |  | -0.01 [-0.06,0.03] | 0.57 |
| 6 |  |  | 0.04 [-0.01,0.09] | 0.15 |
| over time |  |  | 0.01 [-0.04,0.05] | 0.81 |

Table G.9: Preregistered analyses were re-run adjusting for additional covariates, such as the use of other antidepressants and/or psychotherapy during the trial, adherence score, and the number of tablets.

|  |  |  |
| --- | --- | --- |
|  | regression coefficient [95% CI] | p-value |
| 1) Drug effect on loss learning rate at week 2 | | |
|  | 0.48 [0.09,0.87] | 0.02 |
| 2) Loss learning rate related to anxiety over time | | |
|  | 0.02 [0.01,0.03] | 0.004 |
| 3) Early change in aversive Pavlovian bias related to treatment outcome | | |
|  | 0.08 [0.02,0.14] | 0.01 |
| ∆avPav:group | 0.11 [-0.02,0.23] | 0.08 |

Table G.10: Exploratory analyses were re-executed, accounting for additional variables, including the use of other antidepressants and/or psychotherapy during the trial, adherence score, and the number of tablets.