Supplementary methods

Secondary tasks: further details

Visual Affective Bias Task (Daniel-Watanabe et al., 2022)

This task comprised two phases. First, participants learnt through trial and error whether to press the 'Z' or 'M' keys in response to small (∅ 200px) and large (∅ 400px) black discs. These discs were associated with small (£1) and large (£4) virtual rewards for correct responses (the association between disc size and reward magnitude was counterbalanced across participants). This acquisition phase comprised 20 trials, 10 of each trial type, presented in a random order.

In the second phase, participants were told that they would now see discs of intermediate sizes as well, and had to categorise these as small or large too, depending on which group they seemed closest in size to. In fact, the intermediate, ambiguous discs were all exactly halfway between the original stimuli (i.e. ∅ 300px. On half of the trials, the ambiguous stimuli were rewarded as if they were small, and in the other half, they were rewarded as if they were large. There were 120 trials in this phase, 40 trials each of high, low and intermediate stimuli, presented in a random order. The main outcome was the proportion of trials on which participants treated the intermediate discs as if they belonged to the more highly rewarded group, quantifying a positive affective bias.

Risk Taking Task (Rutledge et al., 2016)

In this economic decision-making task, participants had to make a series of choices between a guaranteed outcome or a risky 50-50 gamble. On gain trials, the certain outcome was a gain of between 20 and 60 points, and the gamble returned either 0 points or a larger gain (between 1.7 and 5 times the guaranteed outcome); loss trials were equivalent but with negative values for both the guaranteed outcome and the gamble; and in mixed trials, the guaranteed outcome was zero, and the gamble outcomes were either a gain of between 30 and 150 points, or a loss of 0.2 to 2 times the gain. There were 150 trials in total, 50 gain trials, 50 loss trials and 50 mixed trials, presented in random order.

Computational modelling of the Orthogonal Go/No-Go Task

The modelling approach we took was based on reinforcement learning, a standard approach within computational psychiatry. This in turn builds on earlier work by Rescorla and Wagner (1972) who had proposed the use of an algorithm to update the agent's estimate of the expected value of a stimulus by computing the prediction error and multiplying this by a learning rate, i.e. V(t+1) = V(t) + LR * (R(t) – V(t)). An elaboration on this idea is that one can learn the correct instrumental response to a stimulus by sampling an action and then updating the stored value of the action-stimulus combination (the Q-value; Sutton, 1988). Dayan et al. (2006) were amongst the first to consider how this approach might be applied to the problem of Pavlovian-instrumental conflict and proposed a model in which the Q values for action are augmented by adding a further term representing the Pavlovian value of the stimulus (weighted by a Pavlovian influence parameter). Subsequently, the problem was considered by Guitart-Masip et al. (2011), who compared a number of competing models for the Orthogonal Go/No-Go task, all with a similar structure to that proposed by Dayan et al. (2006); through a process of model comparison, they found that the best model out of those they considered was one in which participants had, in addition to the Q-learning machinery, a Pavlovian bias and an overall go bias. This model was the starting point for our own modelling analyses, and is termed the Base model in this paper.

Specification of the Base model

In the Base model, each trial was modelled as a two-step process, beginning with response generation and then, after the outcome of that action was observed, a learning step.

During response generation, the tendency to make a go response depended on the difference between the values assigned to the go and no-go options (q_{go} and q_{nogo}); on the first trial these were initialised at 0. To this, the participant's go bias and a Pavlovian component were both added to give the decision weight for making a go response on that trial. Specifically, the Pavlovian component contained the associative value of the stimulus shown, scaled by a Pavlovian bias parameter. The stimulus value was coded such that negative values indicated expected punishment, and positive values expected reward – therefore the model generated the classic Pavlovian pattern of responses (go to reward, no-go to punishment) whenever the Pavlovian bias parameter was positive. These components are set out in Equation 1 below.

Decision weight

$$w(s_t) = q_{go}(s_t) - q_{nogo}(s_t) + GoBias_{subject} + Pavbias_{valence,subject} \times value(s_t) \quad (1)$$

Subsequently the decision weight was put through a logistic function to give the probability of making a go response on each trial. Finally a response (go or no-go) for each trial was generated by sampling from a Bernoulli distribution with this probability. This is summarised in Equation 2 below.

Response generation

$$pGo_t = logistic\big(w(s_t)\big) \quad (2a)$$

$$response_t = Bernoulli(pGo_t) \quad (2b)$$

The second set of steps involved updating the learned values of the response that had been chosen (q_{go} or q_{nogo}), as well as the associative value of the stimulus shown ($value(s_t)$). These updates were implemented by Rescorla-Wagner update rules, with separate sensitivities to reward and punishment outcomes. This is summarised in Equations 3 and 4 below.

Instrumental Learning

$$\begin{aligned} q_{response,t+1}(s_t) &= q_{response,t}(s_t) \\ &+ LearningRate_{subject} \times (Sensitivity_{valence,subject} \times outcome \\ &- q_{response,t}(s_t)) \end{aligned} \quad (3)$$

Pavlovian Learning

$$value_{t+1}(s_t) = value_t(s_t) + LearningRate_{subject} \times (Sensitivity_{outcome,subject} \times outcome - value_t(s_t)) \quad (4)$$

Choice of priors

The participant-level parameters were passed through appropriate link functions and then given hierarchical (population-level) priors which were determined through a process of prior predictive checking. These are set out in Equations 5 and 6 and plotted in Figures S1–S4.

Link Functions

$$LearningRate_{subject} = logistic\left(raw\_LearningRate_{subject}\right)$$
$$Sensitivity_{Reward,subject} = e^{raw\_Sensitivity_{Reward,subject}}$$
$$Sensitivity_{Punishment,subject} = e^{raw\_Sensitivity_{Punishment,subject}} \quad (5)$$

Priors

$$GoBias_{subject} \sim Normal(\mu_{GoBias}, \sigma_{GoBias})$$
$$PavBias_{subject} \sim Normal(\mu_{PavBias}, \sigma_{PavBias})$$
$$raw\_LearningRate_{subject} \sim Normal(\mu_{LR}, \sigma_{LR})$$
$$raw\_Sensitivity_{Reward,subject} \sim Normal\left(\mu_{Reward\_sens}, \sigma_{Reward\_sens}\right)$$
$$raw\_Sensitivity_{Punishment,subject} \sim Normal(\mu_{Punishment\_sens}, \sigma_{Punishment\_sens})$$

$$\mu_{GoBias} \sim Normal(0,2) \qquad \sigma_{GoBias} \sim Exponential(1)$$
$$\mu_{PavBias} \sim Normal(0,2) \qquad \sigma_{PavBias} \sim Exponential(1)$$
$$\mu_{LR} \sim Normal(0,1) \qquad \sigma_{LR} \sim Exponential(1)$$

$$\mu_{Reward\_Sens} \sim Normal(0,1.5) \qquad \sigma_{Reward\_Sens} \sim Exponential(1)$$
$$\mu_{Punishment\_Sens} \sim Normal(0,1.5) \qquad \sigma_{Punishment\_Sens} \sim Exponential(1) \tag{6}$$

Other models

We gradually extended the Base model in three stages: we included distinct Pavlovian approach and avoidance biases ('Base + 2PavBias'); we included separate learning rates for reward and punishment (but not approach/avoidance biases; 'Base + 2LR'); finally we included both reward/punishment learning rates and approach/avoidance Pavlovian biases in the same model ('Base + 2PavBias + 2LR').

Specifically, the altered parts of the model (in the decision weight part of the calculation, Equation 1 above, and the instrumental/Pavlovian learning parts, Equations 3 and 4) were as follows.

<u>Base + 2PavBias:</u>

If valence == reward {

$$w(s_t) = q_{go}(s_t) - q_{nogo}(s_t) + GoBias_{subject} + Pavbias_{approach,subject} \times value(s_t)$$

} else {

$$w(s_t) = q_{go}(s_t) - q_{nogo}(s_t) + GoBias_{subject} + Pavbias_{avoid,subject} \times value(s_t)$$

}

<u>Base + 2LR:</u>

If outcome >= 0 {

$$q_{response,t+1}(s_t) = q_{response,t}(s_t) + RewardLearningRate_{subject} \times (Sensitivity_{valence,subject} \times outcome - q_{response,t}(s_t))$$

$$value_{t+1}(s_t) = value_t(s_t) + RewardLearningRate_{subject} \times (Sensitivity_{outcome,subject} \times outcome - value_t(s_t))$$

} else {

$$q_{response,t+1}(s_t) = q_{response,t}(s_t) + PunishmentLearningRate_{subject} \times (Sensitivity_{valence,subject} \times outcome - q_{response,t}(s_t))$$

$$value_{t+1}(s_t) = value_t(s_t) + PunishmentLearningRate_{subject} \times (Sensitivity_{outcome,subject} \times outcome - value_t(s_t))$$

}

Base + 2PavBias + 2LR:

If valence == reward {

$$w(s_t) = q_{go}(s_t) - q_{nogo}(s_t) + GoBias_{subject} + Pavbias_{approach,subject} \times value(s_t)$$

} else {

$$w(s_t) = q_{go}(s_t) - q_{nogo}(s_t) + GoBias_{subject} + Pavbias_{avoid,subject} \times value(s_t)$$

}

If outcome >= 0 {

$$q_{response,t+1}(s_t) = q_{response,t}(s_t) + RewardLearningRate_{subject} \times (Sensitivity_{valence,subject} \times outcome - q_{response,t}(s_t))$$

$$value_{t+1}(s_t) = value_t(s_t) + RewardLearningRate_{subject} \times (Sensitivity_{outcome,subject} \times outcome - value_t(s_t))$$

} else {

$$q_{response,t+1}(s_t) = q_{response,t}(s_t) + PunishmentLearningRate_{subject} \times (Sensitivity_{valence,subject} \times outcome - q_{response,t}(s_t))$$

$$value_{t+1}(s_t) = value_t(s_t) + PunishmentLearningRate_{subject} \times (Sensitivity_{outcome,subject} \times outcome - value_t(s_t))$$

}

Modelling of experimental conditions

The baseline data was fitted with a single model regardless of participants' group membership, because we know *a priori* that the high-conflict and no-conflict groups were identical at baseline since participants were allocated at random (avoiding the so-called Table 1 fallacy). Subsequently, when analysing the data from the follow-up sessions, the model was fitted separately for the two training groups, since at that point there could be a difference between groups. Fitting the groups separately leads to more accurate and less biased parameter estimates (according to parameter-recovery simulations; Valton et al., 2020).

Assessing the models

We compared these models using the Widely-Applicable Information Criterion (WAIC; Watanabe, 2010), which estimates the leave-one-out predictive accuracy of a model; in so doing, WAIC provides both a point estimate and standard error, allowing us to quantify our uncertainty. We selected the best performing model according to their WAIC values, and then examined the posterior estimates of participants' Pavlovian biases according to this model. Our preregistered analysis was to compute the change in Pavlovian bias between baseline and follow-up for each posterior sample, calculate the mean change for each participant and then test (using an independent samples *t*-test) whether there was any difference in this change between the high-conflict and no-conflict training groups.

Prior distributions for the Go/No-Go Models

Figure S1. Prior predictions for the go bias parameters, with distributions $\mu_{GoBias} \sim Normal(0,2)$, $\sigma_{GoBias} \sim Exponential(1)$ and $GoBias_{subject} \sim Normal(\mu_{GoBias}, \sigma_{GoBias})$. (a) and (b) show the analytical distributions of the population mean and standard deviation; (c) shows the prior prediction for the participant-level go bias (log-odds), and (d) the implied go bias probability.

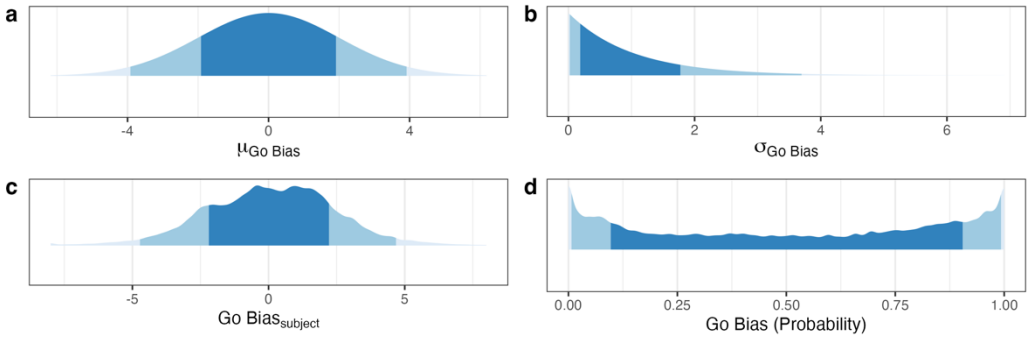


Figure S2. Prior predictions for the Pavlovian bias parameters, with distributions $\mu_{PavBias} \sim Normal(0,2)$, $\sigma_{PavBias} \sim Exponential(1)$ and $PavBias_{subject} \sim Normal(\mu_{PavBias}, \sigma_{PavBias})$. (a) and (b) show the analytical distributions of the population mean and standard deviation; (c) shows the prior prediction for the participant-level Pavlovian bias (dimensionless).

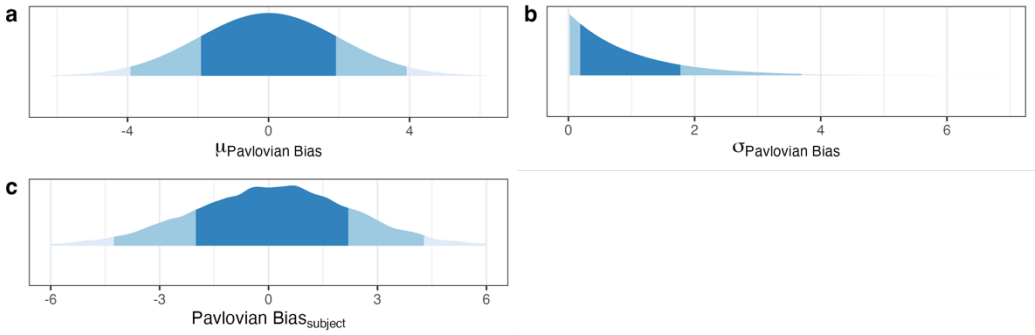


Figure S3. Prior predictions for the outcome sensitivity parameters, with distributions $\mu_{sensitivity} \sim Normal(0,1.5)$, $\sigma_{sensitivity} \sim Exponential(1)$ and $sensitivity_{subject} \sim Normal(\mu_{sensitivity}, \sigma_{sensitivity})$. (a) and (b) show the analytical distributions of the population mean and standard deviation; (c) shows the prior prediction for the participant-level outcome sensitivity parameters, and (d) the implied maximum possible instrumental accuracy (asymptote of learning).

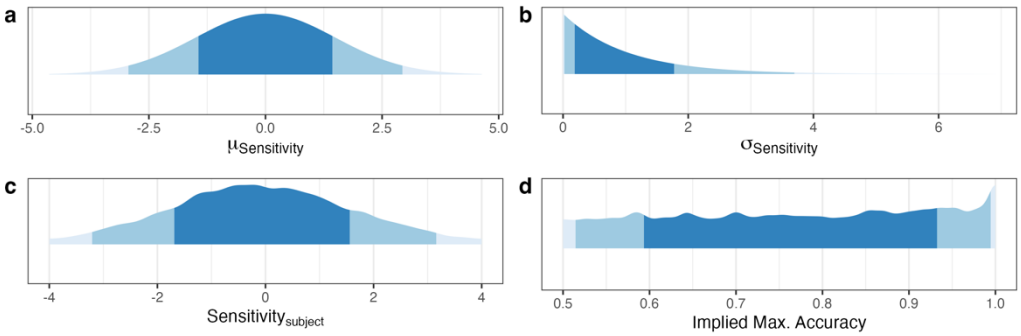
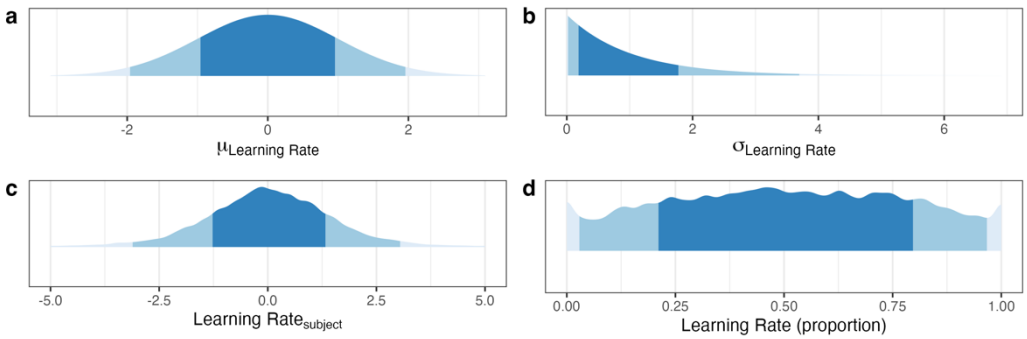


Figure S4. Prior predictions for the learning rate parameters, with distributions $\mu_{LR} \sim Normal(0,1)$, $\sigma_{LR} \sim Exponential(1)$ and $LR_{subject} \sim Normal(\mu_{LR}, \sigma_{LR})$. (a) and (b) show the analytical distributions of the population mean and standard deviation; (c) shows the prior prediction for the participant-level learning rate parameters (in log-odds), and (d) the implied maximum possible learning rate (relative to the outcome sensitivity).



Supplementary results

Descriptive statistics for the study sample

Table S1a. Descriptive statistics for the study sample – age

Age	**N in Active Group**	**N in Sham group**
18-21	65	74
22-25	74	87
26-30	60	69
31-35	50	41
36-40	48	31
41-45	19	13
46-50	11	13
51-55	10	9
56-60	8	8

Table S1b. Descriptive statistics for the study sample – highest level of education achieved

Highest level of education achieved	**N in Active Group**	**N in Sham group**
Secondary Education/High School	103	127
Bachelor's Degree/Degree Apprenticeship	129	122
Foundation Degree	15	9
Higher Apprenticeship	11	15
Masters Degree	75	67
PhD	12	5

Pavlovian Bias Training

Regarding the sequential differences between training sessions, we again observed a difference between groups. The high-conflict training group improved significantly between each of the first, second, third and fourth sessions ($p < .001$, $d = 0.61$; $p < .001$, $d = 0.36$; $p < .001$, $d = 0.27$ respectively), but not between the fourth and fifth sessions ($p = .08$; tests were Bonferroni-corrected for multiple comparisons). By contrast the no-conflict group improved significantly only between the first, second and third sessions ($p < .001$, $d = 0.50$; $p = .009$, $d = 0.17$ respectively) and not between the third, fourth and fifth sessions ($p = .46$ and .55) (albeit the potential for improvement was limited by very accurate performance at baseline in the no-conflict group).

Table S1. Mean accuracy across each of the five training sessions. Both groups improved significantly over the course of the training.

Training condition	Session number	Mean (*SD*) accuracy
No-conflict	1	0.93 (0.08)
	2	0.96 (0.04)
	3	0.97 (0.04)
	4	0.98 (0.03)
	5	0.98 (0.03)
High-conflict	1	0.69 (0.21)
	2	0.78 (0.21)
	3	0.82 (0.20)
	4	0.85 (0.20)
	5	0.86 (0.20)

Affective Bias Task

The affective bias in each of the conditions is plotted in Figure S6. An exploratory 2 X 2 ANOVA (training group X timepoint) revealed a significant main effect of timepoint, $F(1, 688) = 9.90$, $p = .002$, $\eta^2_{partial} = 0.01$. In the baseline session, participants on average showed a negative affective bias (the proportion of responses that equated the ambiguous stimulus to the high-reward exemplar was 0.39, $SD = 0.16$) but after training this bias became less negative (the proportion of 'high' responses increased to 0.41, $SD = 0.19$). The other

effects—the main effect of training condition and the condition X timepoint interaction—were both non-significant, $p = 0.74$ and 0.91 respectively.

Finally we also examined the associations between affective bias (averaged across the two sessions) and scores on each of the mental health symptom scales. None of these correlations were significant (BDI: $p = .35$; STAI-state: $p = .33$; STAI-trait: $p = .70$).

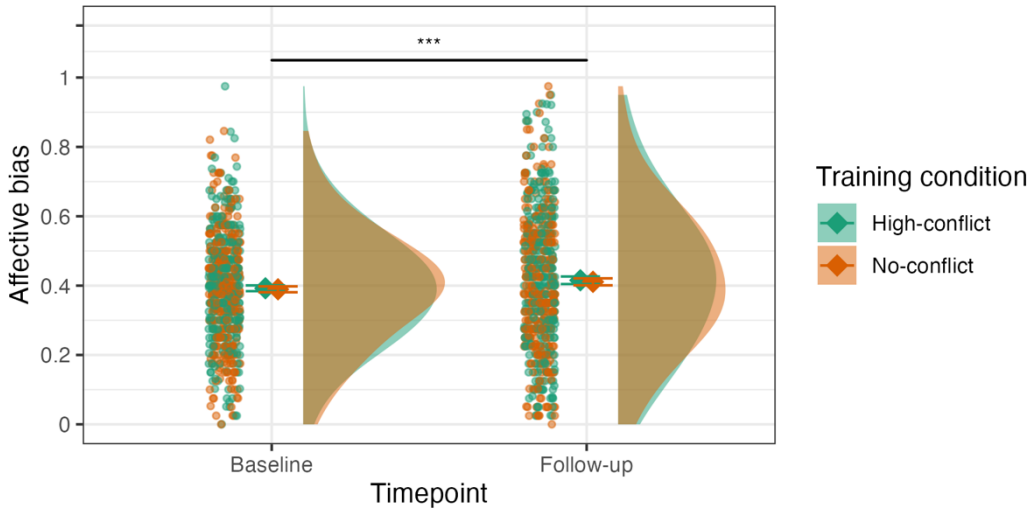


Figure S6. Affective bias before and after training. Affective bias is measured by the proportion of responses matching the ambiguous stimuli to the high-reward exemplar (a value of 0.5 is neutral, <0.5 is a negative bias and >0.5 is a positive bias). There was a significant main effect of timepoint only – the negative affective bias decreased significantly (moved closer to 0.5) after training. Plot shows individual data points (left), mean±SE (centre) and distributions (right).

Risk Taking Task

The proportion of gambles chosen in each condition is plotted in Figure S7. There was a significant interaction between gamble frame and timepoint, $F(2, 1376) = 10.6$, $p < .001$, $\eta^2_{partial} = 0.02$, with participants in both groups choosing to gamble less frequently after training, in the mixed and loss gamble frames only, $t(689) = 2.55$, $p = .01$, $d = 0.1$, and $t(689) = 5.84$, $p < .001$, $d = 0.22$, respectively; in the gain frame there was no change in gambling rates between the baseline and follow-up sessions, $t(689) = 0.12$, $p = .91$. Full descriptive statistics are given in Table S2.

In addition there was a significant main effect of framing, $F(2, 1376) = 1030$, $p < .001$, $\eta^2_{partial} = 0.60$; participants chose to gamble significantly more often during the gain ($M = 0.69$, $SD = 0.24$) versus the mixed frame trials ($M = 0.48$, $SD = 0.24$), $t(1379) = 31.2$, $p < .001$, $d = 0.84$, which in turn was significantly more often than in the loss frame trials ($M = 0.25$, $SD = 0.23$), $t(1379) = 32.9$, $p < .001$, $d = 0.89$. Finally, there was also a significant main effect of timepoint, $F(1, 688) = 16.2$, $p < .001$, $\eta^2_{partial} = 0.02$, with the overall proportion of gambles chosen decreasing from 0.48 ($SD = 0.28$) at Baseline to 0.46 ($SD = 0.32$) at follow up.

The remaining effects – the main effect of training group, the interactions between training group and timepoint, training group and framing, and between training group, timepoint and framing – were all non-significant ($p = .82$, .87, .52 and .70 respectively).

Finally, we also fitted an established computational model to the Risk Taking Task which, amongst other parameters, includes two value-independent bias terms (Rutledge et al., 2016) that have previously been interpreted as representing Pavlovian approach and avoidance. Interestingly, however, we found no correlation between these parameters and the respective approach and avoidance parameters from the Go/No-Go task, $r = 0.02$, $p = .5$, and $r = 0.04$, $p = .25$ (both models fitted to the baseline datasets). This perhaps suggests that the approach and avoidance parameters in the models of the two tasks are in fact measuring different constructs. This would perhaps be an important topic for further investigation.

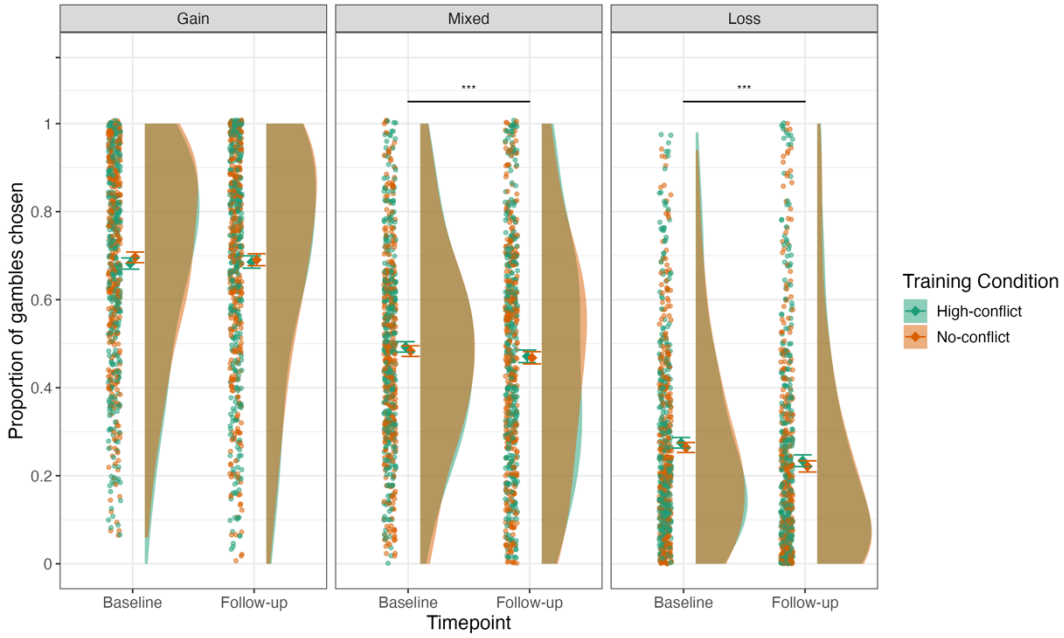


Figure S7. Risk Taking task: Proportions of gambles chosen. Participants chose to gamble less often in the Follow-up session, but only in the mixed and loss frames. In addition, overall the rates of gambling were higher in the gain frame compared with the mixed frame, which in turn was higher than in the loss frame. Plots show (left to right) individual data points, mean±SE and distributions. *** $p < .001$

Table S2. Risk Taking task: Proportion of gambles chosen in each combination of framing and timepoint.

Gamble framing	**Timepoint**	**Proportion of gambles chosen Mean (*SD*)**
Gain	Baseline	0.69 (0.23)
	Follow-Up	0.69 (0.26)
Mixed	Baseline	0.49 (0.22)
	Follow-Up	0.47 (0.26)
Loss	Baseline	0.27 (0.22)
	Follow-Up	0.23 (0.24)

BDI

There was a significant main effect of timepoint, $F(1, 688) = 5.29$, $p = .02$, $\eta^2_{partial} = 0.01$; average depression score decreased from 9.08 ($SD = 8.19$) to 8.69 ($SD = 8.44$) between the baseline and follow-up sessions. The main effect of training group and the training group x timepoint interaction were both non-significant, however, with $p = .79$ and .41 respectively. The BDI scores in each condition are plotted in Figure S8.

We also investigated whether there was any correlation between the change in BDI score and the change in either of the model-derived Pavlovian approach and avoidance bias parameters. These were both non-significant, however: $r = 0.03$, $t(688) = 0.92$, $p = .36$, and $r = –0.03$, $t(688) = 0.77$, $p = .44$ respectively.

Finally, there was no correlation between baseline BDI score and either baseline Pavlovian approach/avoidance biases, or performance on the Go/No-Go task: $r = 0.04$, $t(688) = 1.07$, $p = .29$; $r = 0.02$, $t(688) = 0.50$, $p = .62$; and $r = 0.03$, $t(688) = 0.74$, $p = .46$, respectively.

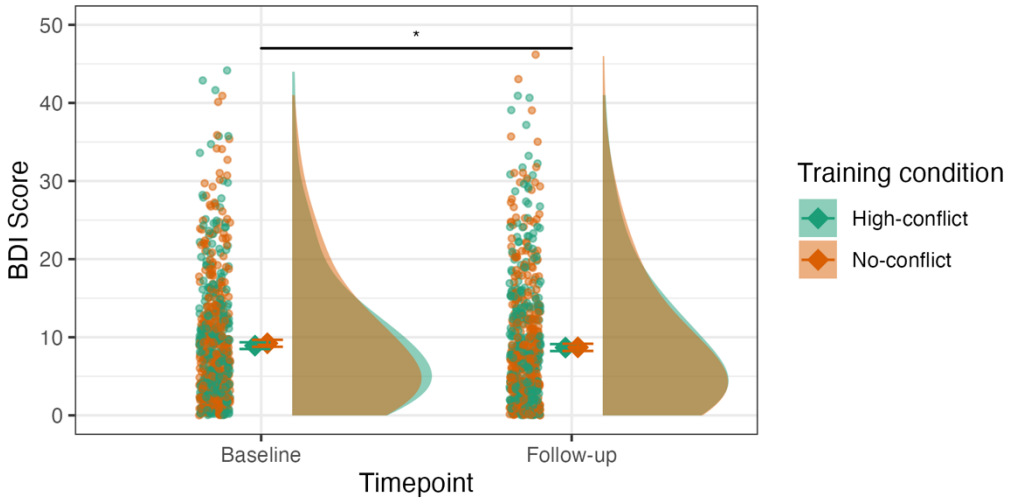


Figure S8. Beck Depression Inventory scores. Scores decreased significantly from Baseline to Follow-up, for both training groups. Plot shows (left to right) individual data points, mean±SE and distributions. * $p < .05$

STAI

Neither the state nor trait subscales of the STAI showed significant effects. For the state subscale, the results were: training group, $F(1, 688) = 0.04$, $p = .84$; timepoint,

$F(1, 688) = 0.35$, $p = .56$; timepoint x group interaction, $F(1, 688) = 1.25$, $p = .26$. For the trait subscale, the results were: training group, $F(1, 688) = 0.40$, $p = .53$; timepoint, $F(1, 688) = 1.85$, $p = .17$; timepoint x group interaction, $F(1, 688) = 0.56$, $p = .46$. The STAI scores in each condition are plotted in Figure S9.

We also investigated the correlations between the change in the STAI scores and the change in the model-derived Pavlovian bias parameters. However, these were all non-significant: for state anxiety, the correlation with approach and avoidance biases respectively was $r = 0.05$, $t(688) = 1.37$, $p = .17$, and $r = 0.01$, $t(688) = 0.38$, $p = .71$; and for trait anxiety, $r = –0.04$, $t(688) = 1.05$, $p = .30$, and $r = –0.01$, $t(688) = 0.46$, $p = .65$.

Finally, there was no correlation between baseline STAI state or trait score and either baseline Pavlovian approach/avoidance biases, or performance on the Go/No-Go task. For state score: $r = 0.06$, $t(688) = 1.46$, $p = .15$; $r = 0.02$, $t(688) = 0.41$, $p = .68$; and $r = 0.03$, $t(688) = 0.79$, $p = .43$. For trait score: $r = 0.03$, $t(688) = 1.07$, $p = .29$; $r = 0.03$, $t(688) = 0.80$, $p = .42$; and $r = –0.01$, $t(688) = 0.15$, $p = .88$.

1

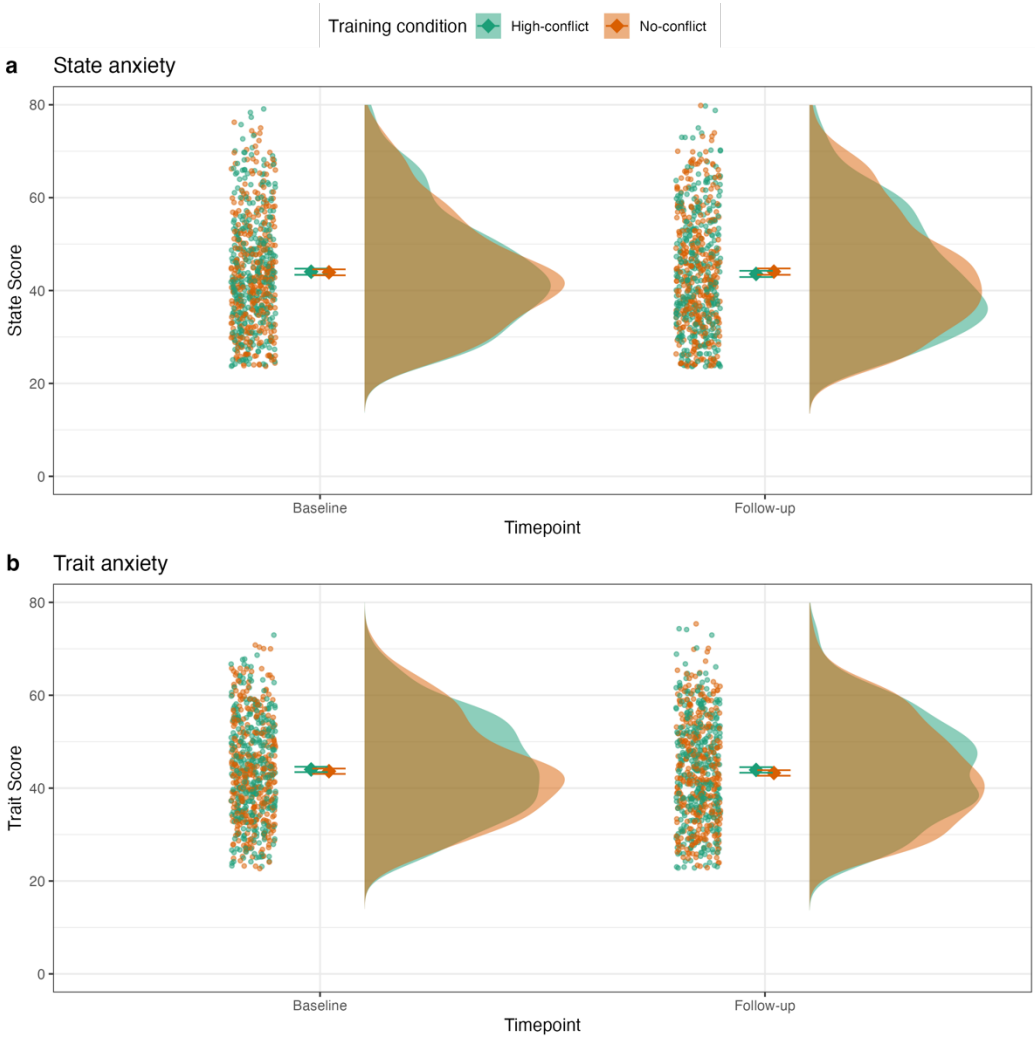


Figure S9. State-Trait Anxiety Inventory Scores. There were no differences between either timepoints or training groups. Plots show (left to right) individual data points, mean±SE and distributions.

Supplementary Results: Comparing participants who dropped out with those who remained

Because the training intervention was relatively intensive, requiring participation every day for a week, one possibility is that some of our results may have been affected by who was retained in the study and who dropped out. To test this, we compared performance from the baseline session between those who subsequently dropped out and those who stayed in the study.

On the Go/No-Go task, we found that participants who dropped out tended to have slightly worse accuracy (M = 53%, SD = 0.09) than those who remained (M = 58%, 0.08), $t(49.6) = 3.44$, $p = .001$, $d = .55$, making it less likely that the effects of training we saw simply reflect reversion to the mean. Crucially, there was no difference in the (model-agnostic) measure of Pavlovian bias, $t(49.3) = 0.68$, $p = .50$.

There were no differences between those who dropped out and those who remained on the Affective Bias task, $t(51.2) = 0.92$, $p = .36$; the Risk Taking task, $t(50.8) = 1.47$, $p = .15$; the BDI, $t(50.2) = .28$, $p = .78$; or the state subscale of the STAI, $t(52.8) = 1.42$, $p = .16$. There was, however, a difference on the trait subscale of the STAI, $t(52.7) = 2.32$, $p = .02$, $d = 0.34$, with those who dropped out having a lower score (M = 40.4, SD = 9.68) than those who remained (M = 43.8, SD = 10.7).

Supplementary Results – Computational Modelling

Model Checking

First the models were compared on the basis of their WAIC values. Figure S10 shows the estimated difference in WAIC (and standard error of this difference) between the best performing model and each model in turn. We found that the best model constituted the Base model plus two Pavlovian biases (approach and avoidance) and two learning rates (for reward and punishment); the distance to the second best model is 3 times its standard error, so we can be reasonably confident that this model has better out-of-sample predictive accuracy than the other models considered.

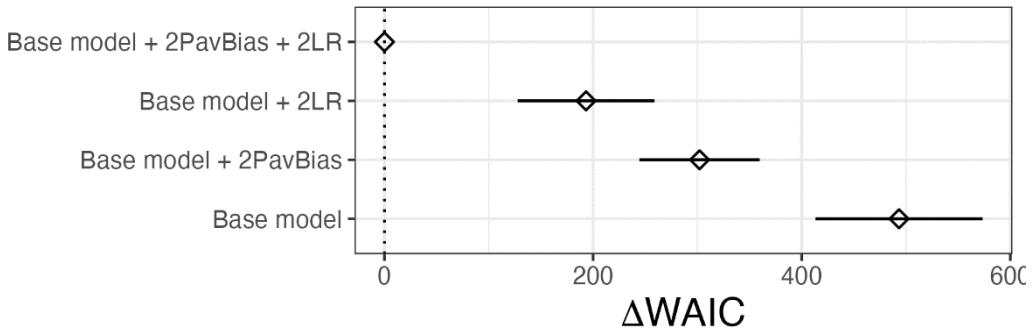


Figure S10. Model comparison results for the Orthogonal Go/No-Go task. Plots show the difference in WAIC (and SE of this difference) between the best performing model, indicated by the vertical dotted line, and each of the models in turn. The best model by nearly three standard errors of difference constituted the Base model plus two Pavlovian biases and two learning rates (for reward and punishment).

Next we examined the trial-wise posterior predictions from the winning model. These are plotted in Figure S11, along with the empirical data for comparison. We see that the model generates predictions that are largely well matched to the empirical data, including the greater variability in participants' performance on the no-go to win reward trials (compare with the distributions of mean accuracy in Figure 3a).

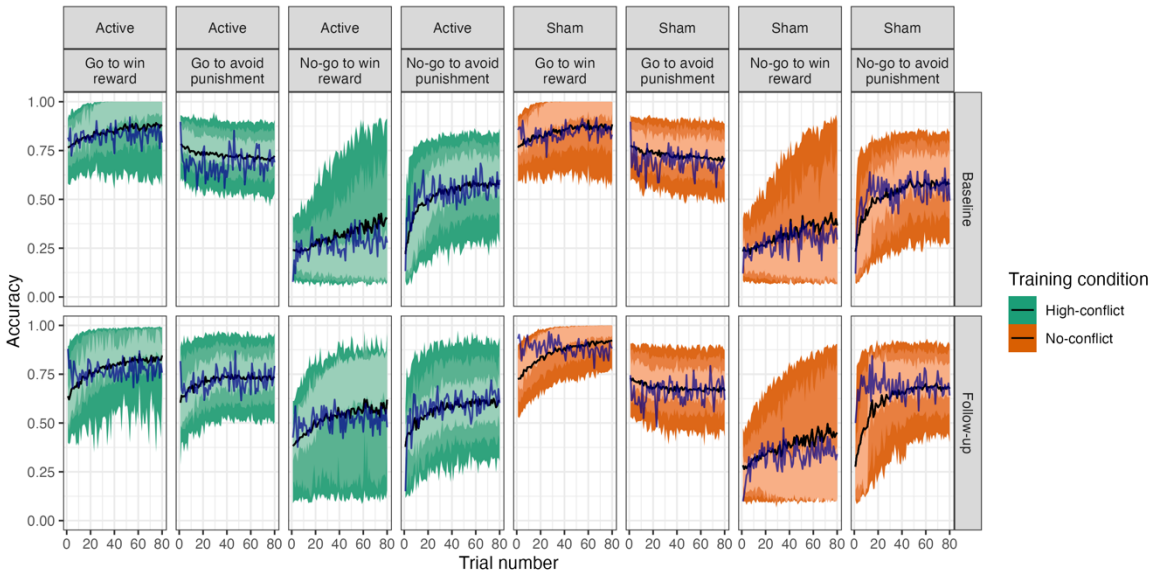


Figure S11. Posterior predictions from the winning model (Base model +2Pavbias + 2LR). Plots show the mean (black line) and 50/80/95% highest density continuous intervals (HDCI) for the posterior predictions, as well as the empirical data (blue line) for comparison.

Exploratory computational modelling

We then examined the posterior distributions of the population-level parameters. These are plotted in Figure S12a, and show training effects in nearly every parameter. To test this more rigorously we computed the changes between sessions for each parameter, and also the differences in these changes between the high-conflict and no-conflict groups, allowing us to infer whether any training effects present differed between groups. These are also plotted in Figures S12b and S12c respectively.

In line with our hypotheses, and supporting the frequentist results reported in the main text, we found that the Pavlovian avoidance bias parameter decreased substantially after training in both groups, but by a much greater amount in the high-conflict training group. From a value of approximately 1.5 at Baseline it was reduced to nearly zero at Follow-up for participants in the high-conflict group, but only to 1.25 in the no-conflict training group. The approach bias was estimated to be almost zero at baseline, and so had very little scope to

change with the training (although it is interesting to note nevertheless that most of the probability density favours a reduction in the approach bias for the high-conflict group too).

The go bias decreased fairly substantially after training in both the no-conflict and high-conflict groups, from log-odds of approximately 1.3 at Baseline to 1 and 0.5 respectively (see Figure S12a; these changes imply a decrease in the probability of making a go response from 79% to 73% and 62% respectively after training). Figure S12b confirms that the go bias decreased by more in the high-conflict group, as the 95% HDCI for the difference in the two changes does not overlap zero.

For the remaining parameters—reward sensitivity, punishment sensitivity, reward learning rate and punishment learning rate—the pattern is more complex. The plots in Figure S12c confirm that the two groups differed in the changes in all four of these parameters. In the case of the reward parameters, the no-conflict group appears not to have shown any change following the training, whereas the high-conflict group show decreased reward sensitivity and increased reward learning rate. Because of the way that sensitivity and learning rate trade-off (in the model, learning rate is expressed as a proportion of the sensitivity), this means that, following the high-conflict training, participants in fact maintain the same absolute learning rate although the asymptote of their learning about rewards is decreased. In the case of the punishment parameters, the high-conflict group similarly exhibit an increase in punishment sensitivity that is compensated by decreased punishment learning rate, meaning that the asymptote for their learning about punishments is increased while the absolute learning rate stays about the same. In contrast, participants in the no-conflict group maintained the same punishment sensitivity but showed increased punishment learning rate.

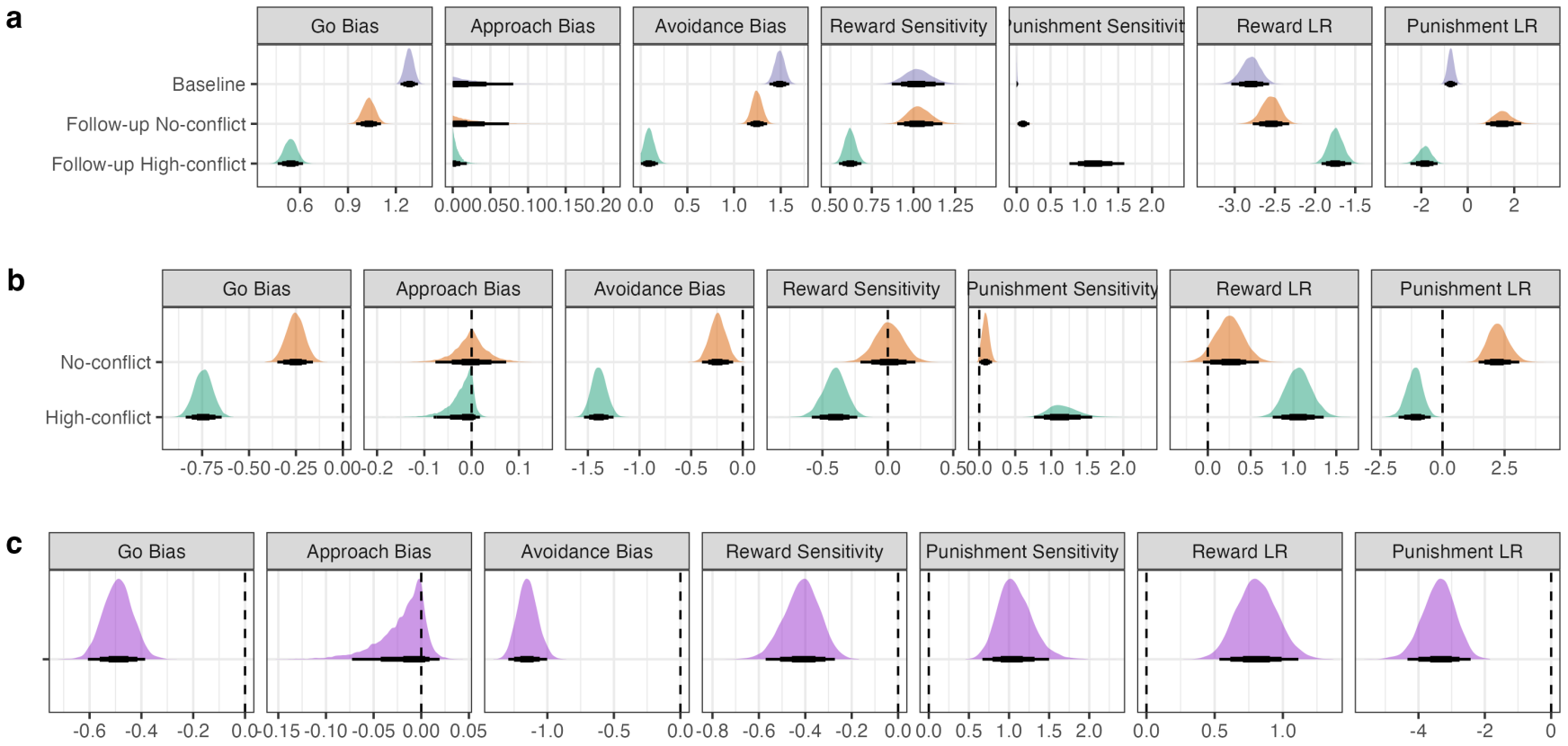


Figure S12. Posterior estimates of the population-level parameters according to the winning model (Base plus two Pavlovian biases plus two learning rates). Each plot shows the posterior distribution and the 50/80/95% highest density continuous intervals (black lines) for the parameter shown. (a) The distributions of the parameters themselves; (b) Distributions of the change (Follow-up minus Baseline) within each group; (c) Distributions of the difference in the change (high-conflict group minus no-conflict).

Parameter recovery for the winning model

The correlations between original and simulated parameters were as follows:

- Go bias, $r = 0.94$, $p < .001$;
- Approach Pavlovian bias, $r = 0.86$, $p < .001$;
- Avoidance Pavlovian bias, $r = 0.83$, $p < .001$;
- Reward learning rate, $r = -0.08$, $p = .38$,
- Punishment learning rate, $r = 0.89$, $p < .001$;
- Reward sensitivity, $r = -0.03$, $p = .81$;
- Punishment sensitivity, $r = 0.37$, $p < .001$

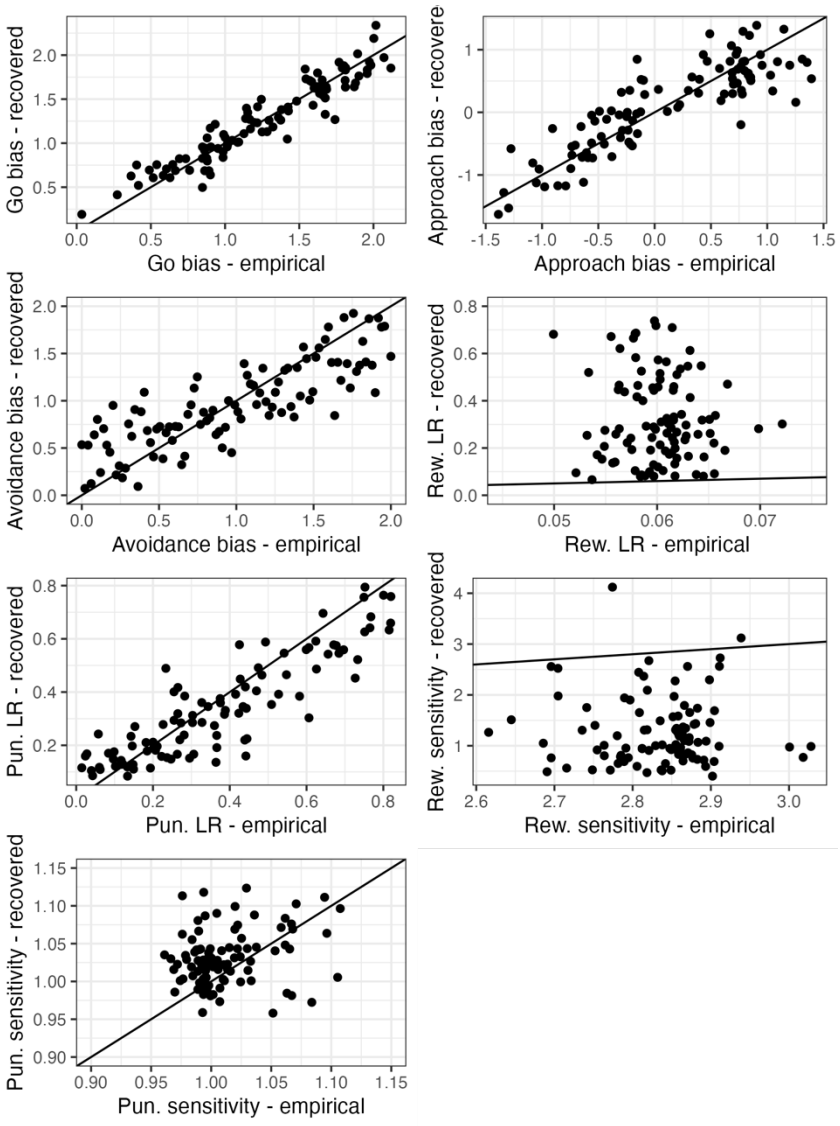


Figure S13. Results of the parameter recovery for the winning model, Base+2Pav+2LR. The black line indicates perfect recovery, y = x.

We also examined the correlations among recovered parameters to better understand the poor recovery of the reward learning rate and sensitivity parameters. The average pairwise correlations across participants are shown in Figure S14. Notably, the strongest correlations were observed between parameters that exhibited good recovery—specifically, between go bias and approach bias ($r = –0.56$), and between go bias and avoidance bias ($r = 0.49$). In contrast, correlations involving the poorly recovered parameters (reward learning rate, reward sensitivity, punishment sensitivity) were generally lower.

This pattern suggests that poor recovery in the reward-related parameters is not primarily due to collinearity or trade-offs with other parameters, but instead reflects reduced power to estimate these parameters precisely. As noted in the main text, we therefore exclude these parameters from inference. To test whether their inclusion might have distorted estimation of the other parameters, we conducted a sensitivity analysis using the simpler Base model (which omits these parameters). This yielded results consistent with our main findings, providing additional reassurance about the robustness of the training effect on Pavlovian bias.

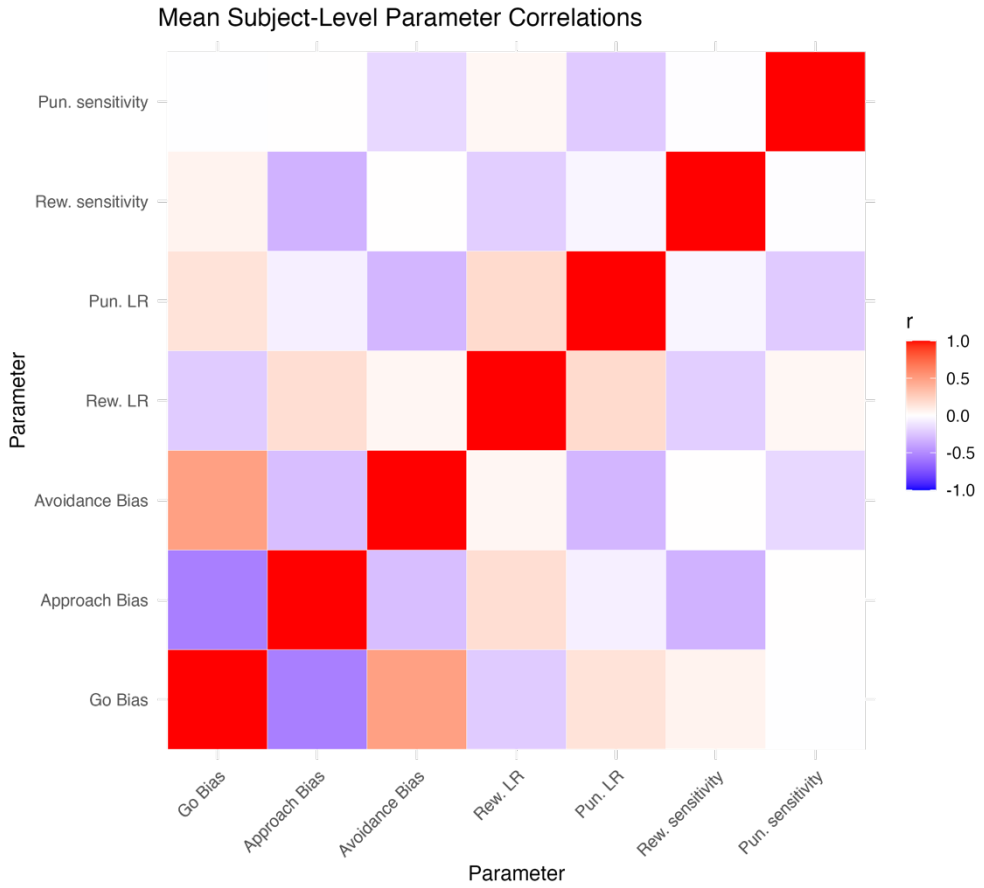


Figure S14. Average within-subject correlations for each pair of parameters. Although the reward sensitivity and learning rate parameters had been the hardest to recover, they showed relatively low correlations with other parameters, suggesting that they were not trading off against each other or other parameters in the model.

Sensitivity analysis with the Base model

As a sensitivity analysis, we also fitted the simplest 'Base' model to our data and compared the results to those obtained from the winning model. The results agreed closely: we observed a substantial difference between the two training groups in the change in Pavlovian bias before and after the training, $t(688) = 13.2$, $p < .001$, $d = 1$. Specifically, the Pavlovian bias parameter for the high-conflict training group decreased from 0.55 to 0.02 after the training, whereas for the no-conflict training group the Pavlovian bias in fact increased slightly, from 0.54 to 0.63. See Figure S15.

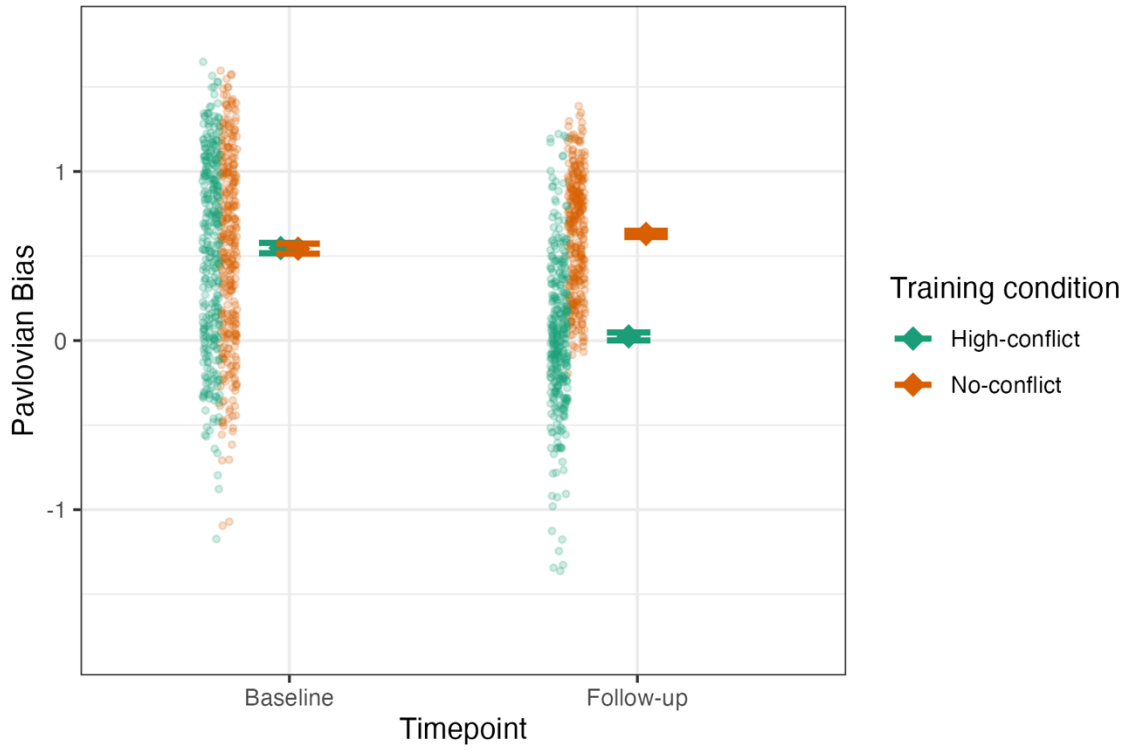


Figure S15. Subjects' Pavlovian bias values according to the simplest 'Base' model. Results are equivalent to those from the winning model; compare with Figure 4 in the main text.

Adaptive Pavlovian-instrumental control: Model simulations

As a complement to our main modelling analysis, we also investigated whether the adaptive model described by Dorfman and Gershman (2019) could explain the change in Pavlovian bias seen with training in this study. To do this, we simulated data from 690 subjects, each doing the full baseline Go/No-Go task (80 trials), five sessions of training on either the conflict or no-conflict trial types (N=345 in each condition, 100 trials), and then the full follow-up task (80 trials again). In order to implement the training effect, we assumed that the Pavlovian weight parameter was 'remembered' from one session to the next, allowing participants to progressively improve over the course of training.

1
2