Cognitive training reduces the strength of Pavlovian biases

Hugo Fleming[a,b*], Gloria W. Feng[c], Robb B. Rutledge[c,d], Jonathan P. Roiser[b] & Oliver J. Robinson[b]

[a] *MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, UK*

[b] *Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AZ, UK*

[c] *Yale University, Department of Psychology, New Haven, CT, 06511, USA*

[d] *Wellcome Centre for Human Neuroimaging, University College London, London, UK*

* Corresponding author: hugo.fleming@mrc-cbu.cam.ac.uk

Author note:

This study was preregistered on the Open Science Framework (https://doi.org/10.17605/OSF.IO/M3Y8U). All study materials are publicly available at https://app.gorilla.sc/openmaterials/669092. All primary data and analysis scripts are publicly available at https://doi.org/10.17605/OSF.IO/7MSVW.

Abstract

Pavlovian biases are patterns of behaviour that involve approaching stimuli associated with reward and avoiding those associated with punishment (regardless of whether this is actually optimal behaviour). They are an ubiquitous feature of everyday decision-making, and are also believed to play an important role in the symptoms of anxiety and depression. Although Pavlovian biases have classically been described as fixed and automatic, some studies have indicated that their influence on behaviour can actually vary over time and with task demands. While these results hint that people may have some control over their Pavlovian biases, direct behavioural evidence for this control is still lacking. In a preregistered, double-blind, sham-controlled study (N = 800), we tested whether a week-long cognitive training intervention could reduce Pavlovian biases on the Orthogonalised Go/No-Go task, a well-established paradigm for isolating Pavlovian-instrumental conflict. Participants were trained on either high-conflict or no-conflict conditions of the task across five days. Using reinforcement learning models to dissociate components of decision-making, we found that high-conflict training led to a significant reduction in Pavlovian bias—particularly avoidance bias—at follow-up. This result is incompatible with the view that Pavlovian biases are fixed and automatic, and instead implies much greater flexibility in the way that they influence cognition than has previously been understood. The training was kept deliberately simple (i.e. one stimulus per condition, with the correct responses kept constant over sessions) so as to provide a minimal proof of concept of whether Pavlovian biases can be reduced through training, but as a result we did not observe transfer to other tasks or self-reported mood. Nonetheless, these findings demonstrate that targeted cognitive training can modulate Pavlovian biases, which may be beneficial both in everyday life and especially in the context of affective disorders like anxiety and depression.

Keywords

Pavlovian bias; Cognitive control; Reinforcement Learning; Anxiety; Depression;

Introduction

Multiple systems govern how humans and other animals select actions. The instrumental system learns the associations between actions and outcomes, and can thereby select an appropriate response that will maximise reward or minimise punishment (Dickinson & Balleine, 2002). A second system is the Pavlovian system, which learns the associations among different stimuli, and promotes fixed responses: the Pavlovian system invigorates action whenever rewards are expected (often described as an approach bias) and inhibits action when punishment is anticipated (avoidance bias; Dayan & Balleine, 2002; Dayan et al., 2006). While these Pavlovian biases are generally advantageous, sometimes they conflict with the responses produced by the more flexible instrumental system, such as when one needs to resist approaching an immediate reward, or take action (e.g. escape) in a potentially dangerous environment (Boureau & Dayan, 2011; Guitart-Masip et al., 2014). Consider, for example, the ambush predator that starts its chase too early, allowing its prey to escape; or the proverbial rabbit in the headlights, that sees a car advancing towards it at speed, but freezes instead of fleeing.

Early work on Pavlovian-instrumental interactions often characterised Pavlovian biases as automatic and evolutionarily hardwired (Boureau & Dayan, 2011; Dayan et al., 2006; Guitart-Masip et al., 2011, 2014). This conclusion was initially inspired by classical conditioning results suggesting that Pavlovian responses are persistent and difficult to overcome (e.g. 'punished pecking' experiments by Williams & Williams, 1969; see also Hershberger, 1986). Computational and theoretical perspectives further highlighted that dopamine plays a dual role in behaviour, as both a learning signal (specifically associated with the reward prediction error in temporal-difference models of learning; Schultz et al., 1997) and a potent driver of motivation and action. In this framework (e.g. Boureau & Dayan, 2011), when a cue is presented which predicts reward, dopamine is released into the dorsal striatum which promotes approach towards and acquisition of the subsequent reward (via the 'Go' pathway); conversely, when a punishment cue is presented, a dip in dopamine causes inhibition of behaviour (via the 'no-go' pathway), facilitating avoidance. Because both action and reward predictions are implemented by the same neurochemical substrate, the link between the two—resulting in the Pavlovian biases—was thought to be fixed.

In the years since, however, emerging evidence has suggested that the expression of Pavlovian biases may be more flexible than initially thought. For example, the balance between Pavlovian and instrumental systems has been shown to vary as a function of task parameters, such as whether outcomes are stochastic vs. deterministic (Dorfman & Gershman, 2019), and to correlate with neural signals associated with top-down cognitive control (e.g., frontal theta; Cavanagh et al., 2013; see also Gershman et al., 2021). These findings suggest that, in specific contexts, people can overcome their Pavlovian biases through cognitive control. Furthermore, reductions in Pavlovian bias have been observed both within a single session (Dorfman & Gershman, 2019) and across extended practice (Schurr et al., 2024), suggesting that these biases are modifiable and can diminish with experience. It remains unclear, however, whether people can learn to strengthen their control over these biases through training, rather than merely adapting incidentally to changes in task structure, stochasticity or repeated exposure. In particular, neither study included a control condition that would allow one to isolate the specific effect of training (i.e. improved engagement of top-down control). Our study addresses this gap by including a randomized control condition, which enables a direct causal test: by comparing high-conflict training with a control group, we can attribute reductions in Pavlovian bias specifically to the targeted training intervention, rather than to general task exposure or repeated practice.

Previous attempts to train Pavlovian biases have so far been unsuccessful. For example, Ereira et al. (2021) conducted four experiments with variants of the Orthogonal Go/No-Go task (Guitart-Masip et al., 2011), looking at the effects of variables like gamification and timing, but found no evidence of training effects (a fifth study showed improvements on a ‘semantic’ Pavlovian bias, which is conceptually quite different from the motor biases on which the literature is based). These findings suggest that direct training effects are difficult to achieve, if indeed this is possible at all. Thus, the question remains as to whether cognitive control over Pavlovian biases, specifically in the motoric domain, can be learned.

It is important to note that the modulation of Pavlovian biases being discussed here is not equivalent to extinction learning. In extinction, the Pavlovian association is weakened or eliminated by withholding the outcome. In contrast, in Pavlovian-instrumental conflict tasks, the Pavlovian contingencies continue to be reinforced, and the challenge lies in overcoming the prepotent behavioural bias they induce. Thus, the question is not whether the association itself can be unlearned, but whether its influence on behaviour can be suppressed.

In addition to its theoretical relevance, the ability to train control over Pavlovian biases has important clinical implications. These biases—particularly the Pavlovian avoidance bias—are enhanced in patients with depression or anxiety, and are thought to contribute to the development and maintenance of symptoms (Mkrtchian et al., 2017; Nord et al., 2018). For instance, a heightened avoidance bias may lead individuals with social anxiety to withdraw in social situations, which can result in more awkward or strained interactions and reinforce negative expectations. In such cases, the expression of a prepotent avoidance tendency prevents effective instrumental engagement – contributing to a self-reinforcing cycle of anxiety and social difficulty. This mirrors the structure of our task, where failure to act in the face of conflict increases the likelihood of punishment. Demonstrating that such biases can be overcome through training suggests a potential route for intervention: by strengthening top-down control over these prepotent tendencies, it may be possible to mitigate symptoms rooted in Pavlovian-instrumental conflict. While our study focused on a healthy population, it provides a foundational proof of concept for targeted cognitive training in clinical contexts.

In the present study, a large-scale (N=800), preregistered, double-blind trial, we assessed Pavlovian biases using the Orthogonal Go/No-Go task (Guitart-Masip et al., 2011; see also Guitart-Masip, Duzel et al. 2014). In this task, the required response and valence of each trial are varied, such that on half of the trials there is Pavlovian-instrumental conflict, requiring cognitive control over Pavlovian biases to respond correctly (see Table 1). A consistent finding from this task is that, although participants do gradually learn the contingencies up to a point, accuracy for the two high-conflict trial types reaches a plateau which is below that of the no-conflict trial types, and well below 100% (see e.g. Figure 2 of Guitart-Masip et al., 2012). After a baseline testing session with the full task, participants doing the high-conflict training practiced only the ‘hard’, control-demanding conflict trials once a day for five days, while those in the control intervention practiced the ‘easy’, no-control trials. Finally, both groups then repeated the full task at a follow-up assessment session (using the same stimuli as at baseline and that they had trained on). Our primary interest here was in testing whether Pavlovian biases can in principle be controlled, so we simplified the task so that there was just one stimulus to learn per trial type, and we recruited a large sample of 800 participants, enabling us to draw strong conclusions about the efficacy of the training. We hypothesised that the group that practiced the control-demanding, Pavlovian-instrumental conflict trials would show a greater improvement in accuracy at the follow-up session than the control group. We also included depression and anxiety self-report scales and secondary tasks to assess transfer effects to other domains.

Table 1. The four trial types of the Orthogonal Go/No-Go task (Guitart-Masip et al., 2011). Across four different trial types, participants have to make either a go or no-go response, for which they would either receive a reward for a correct response (and a neutral outcome otherwise) or a punishment for an incorrect response (and a neutral outcome otherwise). This produces two 'easy' trial types for which the Pavlovian and instrumental systems are aligned (light grey) and two 'hard' trial types for which they are in conflict (dark grey).

	Reward	**Punishment**
Go	Go to win reward	Go to avoid punishment
No-Go	No-go to win reward	No-go to avoid punishment

Methods

Preregistration

This study was preregistered on the Open Science Framework (DOI:10.17605/OSF.IO/M3Y8U).

Participants

In total, 800 adults participated in this study, recruited through the online platform Prolific. All were fluent in English and reported no history of psychiatric or neurological disorders. Participants were informed that they would be doing a week of practice on the task, but they were not told anything else about the nature of the training or indeed the existence of active and sham versions of the intervention.

After examining the data, we excluded 110 participants (a schedule is provided in Table 2). The exclusion criteria were all preregistered except one, which excluded participants who, during the Go/No-Go task, responded with keys outside the response set ('S' or 'L' keys) on more than 15% of trials. This criterion led to the exclusion of three participants; no other participants made nearly so many wrong-key responses (the maximum among the other participants was just 3%).

This left us with 690 participants whose data was included in the final analysis, exceeding our preregistered minimum sample size of 676 (which was determined by a priori power analysis, using $d = 0.25$, $\alpha = 5\%$, power = 90%).

The study was approved by the UCL Research Ethics Committee (6198/001).

Table 2. Schedule of exclusions. GNG=Go/No-Go task, Aff. Bias=Affective Bias task, STAI=State-Trait Anxiety Inventory. *This was not a preregistered criterion – see *Participants* for details.

Timepoint	**Reason**	**N excluded**	**N remaining**
			800
Baseline Testing	Did not complete baseline session	11	789
	GNG: go to win reward accuracy < 65%	9	780
	GNG: left/right accuracy < 65%	1	779
	Aff. Bias: accuracy on unambiguous trials < 60%	29	750
	Aff. Bias: no response on > 15% of trials	2	748
	STAI: failed attention check	2	746
Training	Did not complete 5 training sessions	46	700
Follow-Up Testing	Did not complete follow up session	1	699
	GNG: go to win reward accuracy < 65%	2	697
	GNG: wrong-key responses >15% (*)	3	694
	Aff. Bias: accuracy on unambiguous trials < 60%	1	693
	Aff. Bias: no response on > 15% of trials	2	691
	STAI: failed attention check	1	690

Procedure

The study comprised three phases, which took place over eight days, as shown in Figure 1.

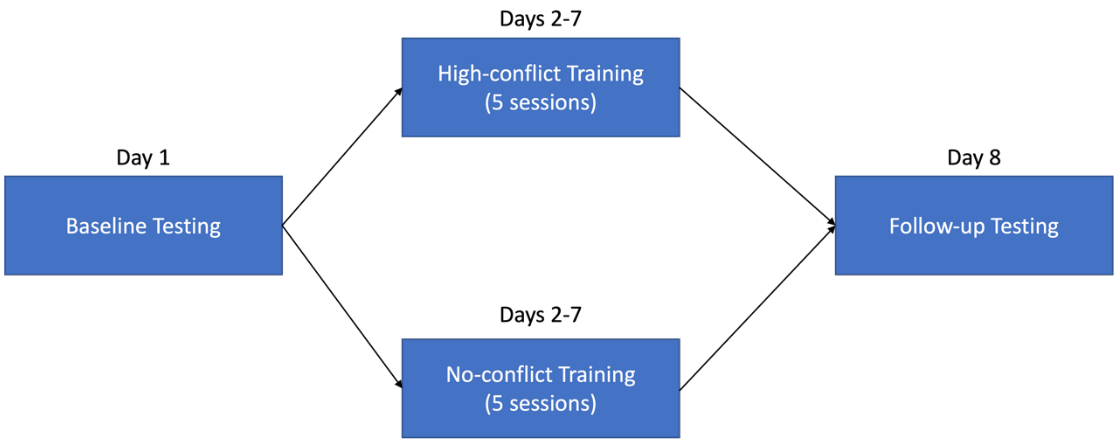


Figure 1. Timeline of the Study. Participants completed the full battery of tasks at baseline, then were randomised to receive either the high-conflict or no-conflict training. They completed five practice sessions over six days, then repeated the full battery of tasks at follow-up.

On the first day, participants completed a baseline testing session in which they completed the Orthogonal Go/No-Go task (Guitart-Masip et al., 2011), the Affective Bias task (Daniel-Watanabe et al., 2022), the Risk Taking task (Rutledge et al., 2016) and two mental health questionnaires (the Beck Depression Inventory, Beck et al., 1996; and the State-Trait Anxiety Inventory, Spielberger et al., 1983). Then, they were randomly allocated to receive either the high-conflict or no-conflict training – those in the high-conflict group were given solely the control-demanding, Pavlovian conflict trials of the Go/No-Go task to practice, while those in the no-conflict group practiced just the no-conflict trials. Participants had to complete five training sessions over six days (with a maximum of one training session per day allowed); participants who had not completed all five sessions by the end of the training period were excluded from the study. Finally, on the eighth day of the study, participants completed a follow-up session containing the same battery of assessments as at baseline.

These sessions were conducted entirely online, using the experiment platform Gorilla (www.gorilla.sc), which also performed the randomisation to the high-conflict or no-conflict training groups automatically.

Measures and Tasks

Orthogonal Go/No-Go Task (Guitart-Masip et al., 2011)

The full procedure for this task is set out in Figure 2a. A trial consisted of three events, each displayed for 1000ms with a 250ms inter-stimulus interval: first an initial fractal cue was shown in the centre of the screen; then a circle target was displayed on one side of the screen, to which participants chose whether or not to respond; finally the outcome of their response was displayed.

Each fractal was associated with both a required response (‘go’ or ‘no-go’) and a valence (correct responses allowed participants either to win points, or avoid losing them). Combining these responses and valences orthogonally produced four unique trial types (see Figure 2b), each represented by a different fractal: go to win reward, go to avoid punishment, no-go to win reward, no-go to avoid punishment. Participants had to learn the correct responses (and maximise their points) through trial and error. Note that the outcomes were probabilistic, such that a correct response only led to reward (or avoided punishment) 80% of the time. In addition, to hold participants’ attention, the exact keypress for a ‘go’ response (‘S’ or ‘L’) had to match whichever side of the screen the target was presented on.

After a number of practice rounds, during which the complexity of the task was built up gradually, the main phase of the task comprised 80 trials (20 per condition) presented in a random order. The fractal allocation was randomised for each participant at the start of the baseline session, and then retained for the training and follow-up sessions.

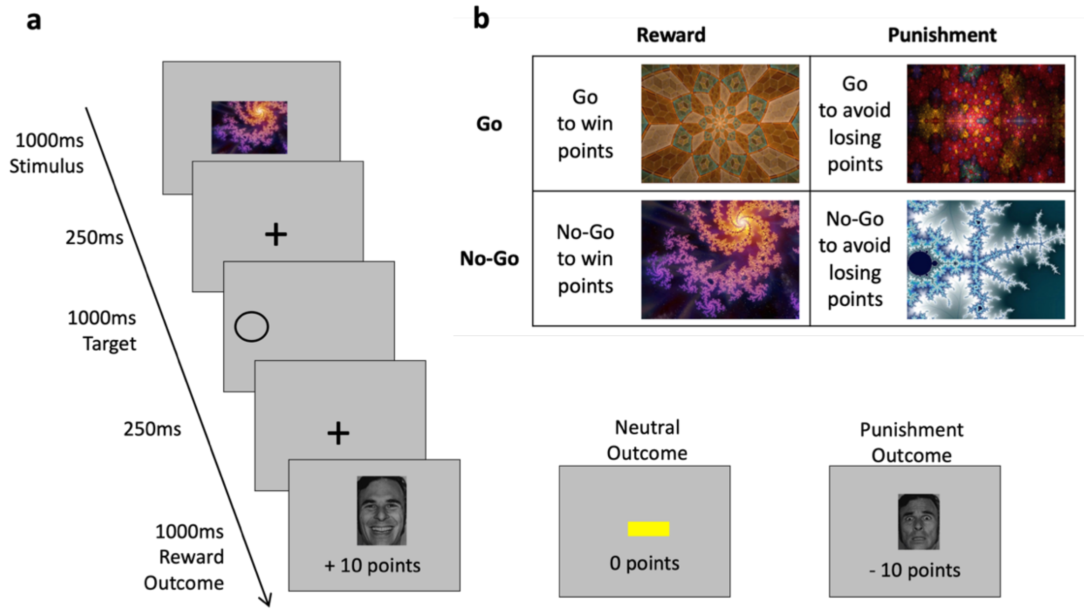


Figure 2. Procedure for the Orthogonal Go/No-Go task. (a) Participants were shown an initial fractal cue, associated with a required response (go or no-go) and a valence (reward or punishment). Participants made or omitted a response when the target appeared, and then received an outcome depending on their response and the fractal valence. (b) The four possible trial types (combining the possible responses and valences orthogonally). This figure builds on Table 1, additionally showing a possible allocation of the fractals to the four trial types. In the study, the fractal allocation was randomised for each participant at Baseline.

Pavlovian Bias Training

Participants were trained on a variant of the same Orthogonal Go/No-Go task described above, but with just a subset of conditions: the high-conflict training group practiced just the trial types which involved Pavlovian-instrumental conflict (‘go to avoid punishment’ and ‘no-go to win reward’), while those in the no-conflict group were trained on just the no-conflict trial types (‘go to win reward’ and ‘no-go to avoid punishment’). The fractal allocation was the same as at baseline and follow-up. Each training session comprised 48 trials (24 trials per condition, in a random order), and participants were required to complete at least five training sessions in six days (one session per day). We did not collect information on the time of day participants completed the training sessions but, as participants were randomised to the two conditions, this would not be expected to vary systematically between the groups.

Secondary outcomes

We included two further tasks to investigate whether the training, if successful on the main task, would also transfer to other domains. The Visual Affective Bias task (Daniel-Watanabe et al., 2022) assessed how participants' perceptual judgements about the size of an ambiguous stimulus (a medium-sized black disc) were affected by receiving asymmetric rewards for choosing large versus small. The Risk Taking Task (Rutledge et al., 2016) measured participants' tendency to gamble versus take a safe, certain outcome, as the expected values of these options were varied. See Supplement for further details.

We also administered two well-validated mental health questionnaires: the Beck Depression Inventory (BDI; Beck et al., 1996) and the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983). Both instruments have robust psychometric properties, including high internal consistency and test–retest reliability, as reported in the original validation studies. We removed one question from the BDI which asks about thoughts of suicide, due to the safeguarding risk. We also added a catch question ("Press the very much so button") at the end of the STAI to detect inattentive participants.

Preregistered Analyses

To test our primary hypothesis that the training would enhance control over Pavlovian biases in the Orthogonal Go/No-Go task, we planned two related analyses, one model-agnostic and another that involved computational modelling.

Model Agnostic Analysis of the Orthogonal Go/No-Go Task

For the model-agnostic analysis, we first calculated a measure of Pavlovian bias for each participant in each session. This was defined as the sum of the accuracies for the two Pavlovian-instrumental conflict trial types (go to avoid punishment and no-go to win reward) minus the sum of the two no-conflict trial types (go to win reward and no-go to avoid punishment). We then computed a training effect, which was the change in this measure between the baseline and follow-up sessions. Finally we tested (using an independent samples *t*-test) our primary hypothesis that there would be a difference in the change in this Pavlovian bias metric between the high-conflict and no-conflict training groups.

We also preregistered a secondary hypothesis that participants in both groups would exhibit Pavlovian biases at a baseline. To test this, we ran a 2 x 2 (required response x valence) repeated-measures ANOVA on the accuracy data from the baseline session, followed by four planned paired-samples t-tests comparing the go to avoid punishment and no-go to win reward conditions with each of the go to win reward and no-go to avoid punishment conditions.

Computational modelling of the Orthogonal Go/No-Go Task

In parallel we also tested our primary hypothesis using computational modelling. Our models build on established reinforcement learning frameworks that combine instrumental Q-learning (Rescorla & Wagner, 1972; Sutton, 1988) with Pavlovian value influences (Dayan et al., 2006). In particular, we use as a starting point the winning model from Guitart-Masip et al. (2012), and call this the Base model. In this model, each trial's action value is computed from the expected instrumental reward (the difference between the q-values for go and no-go), modulated by fixed biases reflecting approach and avoidance tendencies. These biases multiply the Pavlovian value of the present stimulus, and so determine to what extent the Pavlovian estimates can influence the action weights; they promote go responses when reward is anticipated (i.e. value is positive) and no-go responses when punishment is expected (i.e. value is negative). In addition, we also assume that participants have a general go bias, which invigorates action regardless of the instrumental or Pavlovian values on that trial. These factors are summarised in the equation below, which shows how the action weight ($w(s)$, the logodds of making a go response) is calculated on each trial (where s indexes the stimulus shown on that trial):

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

Here, $\mathrm{q_{go}}(\mathrm{s_t})$ and $q_{\mathrm{no}go}(s_t)$ are the instrumental Q-values for the go and no-go actions, respectively; $GoBias_{subject}$ reflects a general tendency to make go responses; $Pavbias_{subject}$ is the Pavlovian bias parameter; and $value(s_t)$ is the Pavlovian value of stimulus s_t.

The other key subject-level parameters in the Base model are the outcome sensitivity, which acts as a multiplier on the outcome received on each trial (following Guitart-Masip et al., we assume that participants may value rewards and punishments differently), and the learning rate, which acts as a multiplier on the prediction error observed on each trial.

We subsequently considered three extensions to this model: one with separate Pavlovian approach and avoidance biases ('Base + 2PavBias'), which applied when the Pavlovian value of the stimulus was rewarding or punishing respectively; another with distinct learning rates for reward and punishment outcomes ('Base + 2LR'); and a third with both approach/avoidance Pavlovian biases and reward/punishment learning rates ('Base + 2PavBias + 2LR').

The full model specifications for all four models, including parameter definitions, formal equations, parameter recovery and model comparison results, are provided in the Supplement (p. 2), where there is also further discussion on the relationship between these models and previous work. All model code and data is also publicly available on OSF (https://doi.org/10.17605/OSF.IO/7MSVW).

For the Baseline session, we fitted the models to the entire sample of participants all together, and then for the follow-up session, we fitted the models to the high-conflict and no-conflict groups separately (thus assuming that the two groups were identical before the intervention, but might differ afterwards).

All models were fitted using MCMC in Stan (Stan Development Team, 2023). Sampling was run for four chains each with 2000 iterations. Subsequent to fitting we carried out diagnostics (visual inspection of the chains; divergences or treedepth warnings; E-BFMI < 0.3, effective sample size > 400, split-$\hat{R}$ < 1.01; Betancourt, 2018). Besides a negligible number of divergences (<1% for all models) there were no issues. We also inspected the posterior predictions for the winning model and observed a good overall fit to the data (Figure S12).

We compared these models using WAIC (Watanabe, 2010), which provides an estimate of out-of-sample predictive accuracy, then examined the posterior parameter values from the winning model. As with the model-agnostic approach described above, our preregistered analysis used an independent samples *t*-test to assess whether the mean change in the

Pavlovian bias term differed between the high-conflict and no-conflict training groups. Further exploratory analyses are reported in the supplement.

Affective Bias Task

Our preregistered analysis for this task involved assessing (using an independent samples *t*-test) whether the two training groups differed in their change in affective bias between baseline and follow-up.

Risk Taking Task

Our preregistered analysis for this task involved assessing, using a 3 x 2 x 2 ANOVA (framing x timepoint x training group), whether there was an interaction between timepoint and group on gambling choices.

Materials, data and code availability

All study materials are publicly available at https://app.gorilla.sc/openmaterials/669092. All primary data and analysis scripts are publicly available at DOI: 10.17605/OSF.IO/7MSVW.

Results

As expected, both groups exhibited significant Pavlovian biases at baseline (action-by-valence interaction, $F[1, 689] = 709$, $p < .001$, $\eta^2_{partial} = 0.51$; see Figure 3a, left panel). Accuracy was worse when participants had to go to avoid punishment compared with go to win reward, $t(689) = 18.4$, $p < .001$, $d = 0.70$, and likewise when participants had to no-go to win reward compared with no-go to avoid punishment, $t(689) = 23.9$, $p < .001$, $d = 0.91$ (both tests remained significant using Bonferroni correction of $\alpha = .025$). We also found significant action and valence biases: participants found it easier to learn go than no-go responses ($M = 0.76$, $SD = 0.17$, versus $M = 0.40$, $SD = 0.25$; $F[1, 689] = 1680$, $p < .001$, $\eta^2_{partial} = 0.71$); and they were better at avoiding punishment than winning reward ($M = 0.61$, $SD = 0.15$, versus $M = 0.55$, $SD = 0.36$; $F[1, 689] = 97.3$, $p < .001$, $\eta^2_{partial} = 0.12$).

Our second planned analysis looked at the model-agnostic measure of Pavlovian bias. At follow-up, we saw a substantial difference between the training groups in the change in their Pavlovian biases, $t(688) = 11.9$, $p < .001$, $d = 0.91$. Specifically, the high-conflict training led to a large reduction in Pavlovian bias, $t(344) = 9.90$, $p < .001$, $d = 0.53$, while this bias in fact became stronger following the no-conflict training, $t(344) = 6.86$, $p < .001$, $d = 0.37$ (both post hoc tests again remained significant after Bonferroni correction). These results are plotted in Figure 3b, and descriptive statistics are given in Table 3.

In an exploratory analysis, we also used ANCOVA to control for baseline differences in Pavlovian bias. In agreement with our preregistered analysis, we found there was a significant effect of training condition, $F(1,687) = 210$, $p < .001$, $\eta^2_{partial} = 0.23$. Separately, we also verified that the high-conflict group showed a significantly greater improvement in overall accuracy, $t(688) = 2.29$, $p = .022$, $d = 0.17$.

Table 3. Model-agnostic Pavlovian bias measure in each condition.

Training condition	Timepoint	Pavlovian Bias Mean (*SD*)
High-conflict	Baseline	0.40 (0.39)
	Follow-Up	0.12 (0.48)
No-conflict	Baseline	0.40 (0.40)
	Follow-Up	0.58 (0.38)

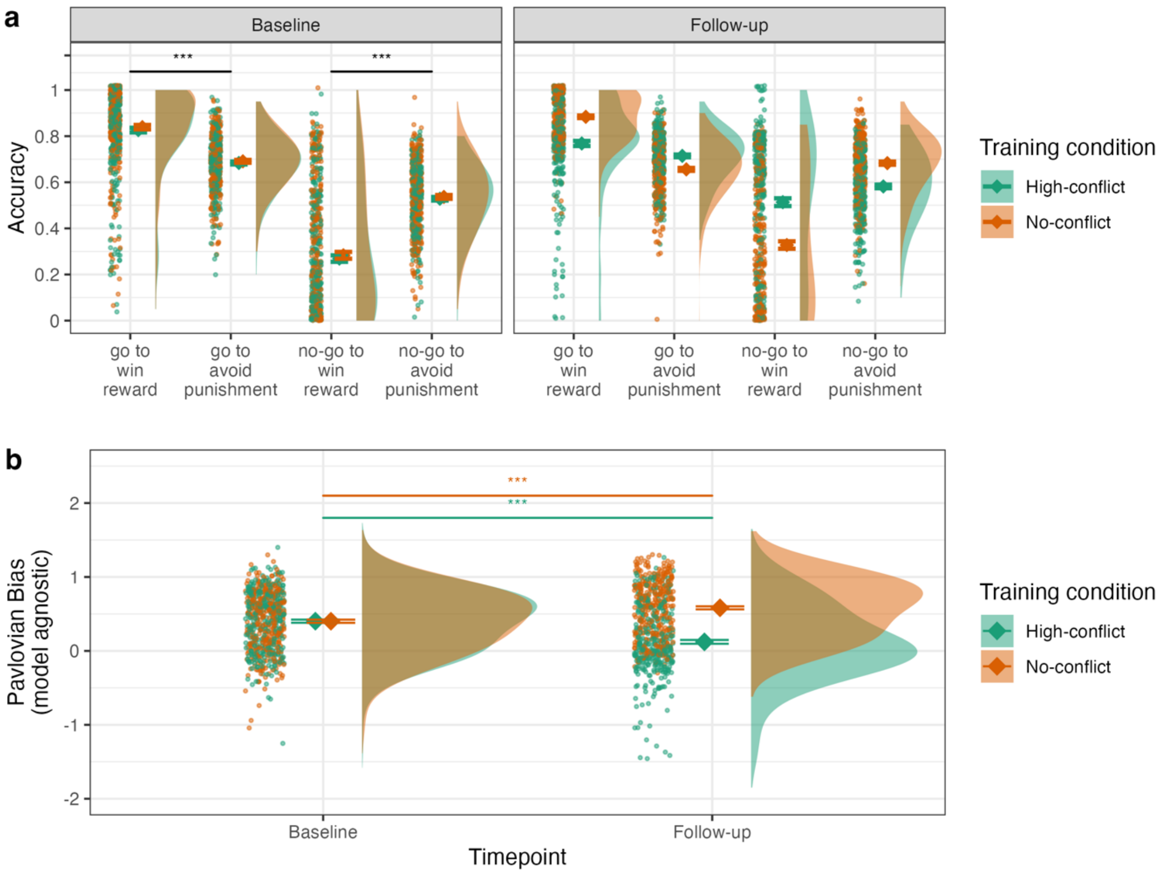


Figure 3. High-conflict training significantly enhanced control over Pavlovian biases. Both plots show the individual datapoints, the mean ± SE (diamonds and horizontal lines) and overall distribution. A) Groups were closely matched at baseline and showed clear signs of Pavlovian bias (impaired accuracy at go to avoid punishment and no-go to win reward trials). Following training, the high-conflict (but not no-conflict) group showed a significant improvement in accuracy. B) The high-conflict training, but not the no-conflict training, led to a significant decrease in the model agnostic measure of Pavlovian bias. *** indicates $p < .001$

Performance during the training phase

Turning to performance during the training phase itself, we observed a significant difference between groups (see Figure 4 and Table S1). While both training groups improved their accuracy over the course of training, there was a significant interaction between training group and timepoint, $F(4, 2752) = 63.0$, $p < .001$, $\eta^2_{partial} = 0.08$, with the high-conflict group improving by a greater amount. Indeed the average improvement in accuracy from the first to the fifth training sessions for the high-conflict training group was 0.16 ($SD = 0.18$), while for the no-conflict group it was 0.05 ($SD = 0.09$), $t(491) = 9.91$, $p < .001$, $d = 0.76$.

Detailed analysis of the sequential differences between training sessions is provided in the Supplement, p.11.

Figure 4. Performance on the Pavlovian bias training, split by trial type. Average accuracy in both groups improved over the course of training, but the improvement was greater in the high-conflict compared with the no-conflict group. Plots show individual data points and distributions (colour) and means±SE (black).

Computational modelling

The best performing model had separate Pavlovian approach and avoidance biases, and separate reward and punishment learning rates (see supplement for full details of the model comparison). The Pavlovian avoidance parameter reflects the degree to which participants tended to withhold action in the presence of stimuli that had a Pavlovian association with punishment. Reductions in this parameter therefore indicate a diminished influence of the Pavlovian system on behaviour. Examining the posterior estimates from this model (Figure 5), we found a significant difference between the high-conflict and no-conflict training groups in their change in avoidance bias ($t[688] = 36.1$, $p < .001$, $d = 2.75$), but not approach bias ($t[688] = 0.35$, $p = .73$). Specifically, the high-conflict training reduced the strength of participants' avoidance bias almost to zero on average (from 1.49 at baseline to 0.09 at follow-up: mean change = –1.40, $SD = 0.55$; $t[344] = 46.8$, $p < .001$, $d = 2.52$), whereas those in the no-conflict group showed a much smaller reduction (mean change = –0.24, $SD = 0.21$; $t[344] = 21.2$, $p < .001$, $d = 1.14$; again, remaining significant after Bonferroni correction). It is worth nothing that, by modelling the approach and avoidance biases separately, we have been able to reveal a much larger change in (avoidance) bias than had been indicated by our model agnostic analysis above.

As before, we also ran an exploratory analysis with ANCOVA to control for any differences in baseline Pavlovian biases, and again this led to the same conclusions (a significant training effect on avoidance bias, $F[1,687] = 1310$, $p < .001$, $\eta^2_{partial} = 0.66$, but not approach bias, $F[1,687] = 0.60$, $p = .4$).

Of note, we also observed a significant decrease between groups in their change in go-bias ($t[688] = 11.4$, $p < .001$, $d = 0.87$). While both groups demonstrated significantly reduced go biases after training, the reduction was substantially greater in the active training group versus the sham group ($t[344] = 22.1$, $p < .001$, $d = 1.19$, and $t[344] = 8.51$, $p < .001$, $d = 0.46$ respectively; a reduction from M = 1.29, SD = 0.43 to M = 0.54, SD = 0.59 for the active group, compared with M = 1.28, SD = 0.44 to M = 1.03, SD = 0.46 for the sham group). This may partially account for the absence of an effect on the approach bias parameter (since both parameters promote go responses), a point we return to in the discussion. (Note that these *post hoc* t-tests remained significant after correction for multiple comparisons).

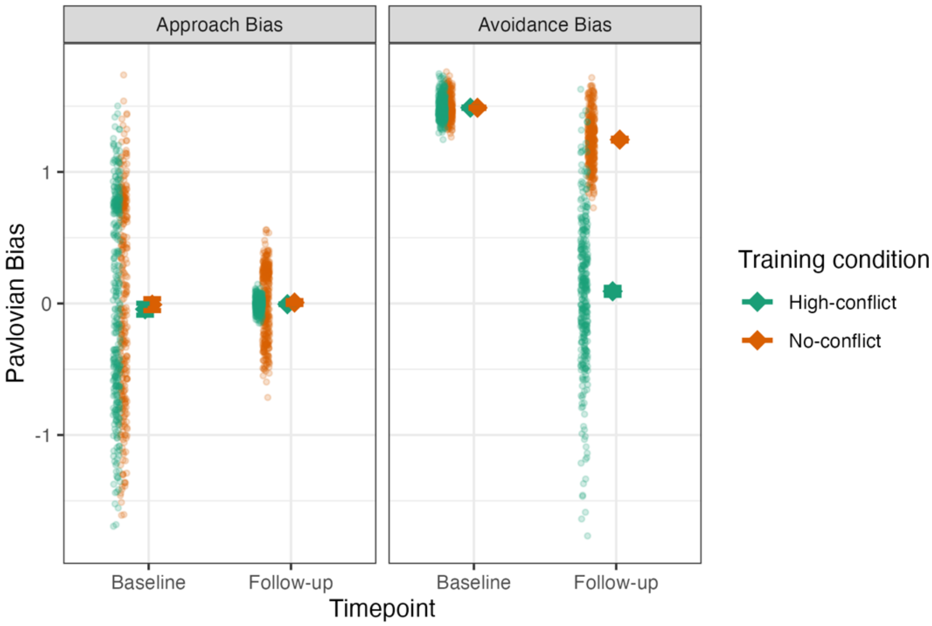


Figure 5. Subjects' Pavlovian bias values, according to the winning model (Base plus two Pavlovian bias plus two learning rates). Plot shows each participant's mean bias and the mean ± SE in each condition.

To assess the validity of these model-based inferences, we conducted a parameter recovery analysis, by fitting the model to simulated data generated using the observed parameter values. This analysis tests whether the model can recover the original parameter estimates, which is essential for interpreting individual and group-level differences in behaviour. The go bias parameter, both Pavlovian bias parameters and the punishment learning rate parameters could all be recovered reliably ($r > 0.8$). In contrast, reward learning rate, reward sensitivity and punishment sensitivity parameters showed poorer recovery. Correlation analyses (see Supplement) suggested that this was not due to collinearity, implying that poor recovery of these parameters is unlikely to have distorted estimation of the reliably-recovered ones. To further validate our findings, we re-ran the primary analysis using the simpler Base model (which does not have the poorly-recovering parameters) and obtained identical results (see Figure S15). These checks confirm that the training effect on Pavlovian bias is robust and not dependent on model complexity or parameterization.

Finally we investigated a possible mechanistic account of this change in Pavlovian bias, based on the model proposed in Dorfman & Gershman (2019). In that paper, the authors proposed an algorithm in which the balance between Pavlovian and instrumental controllers is updated dynamically, based on their relative predictive power, through a process of Bayesian model averaging. After simulating new participants with their model (details are provided in the Supplement), we found that it provides a qualitatively good match to our empirical results, reproducing an improvement following training which is specific to the conflict training condition only. This suggests that an adaptive process, which updates the weight given to the Pavlovian and instrumental controllers based on their ongoing predictive accuracy, could underlie the training effects which we observed here.

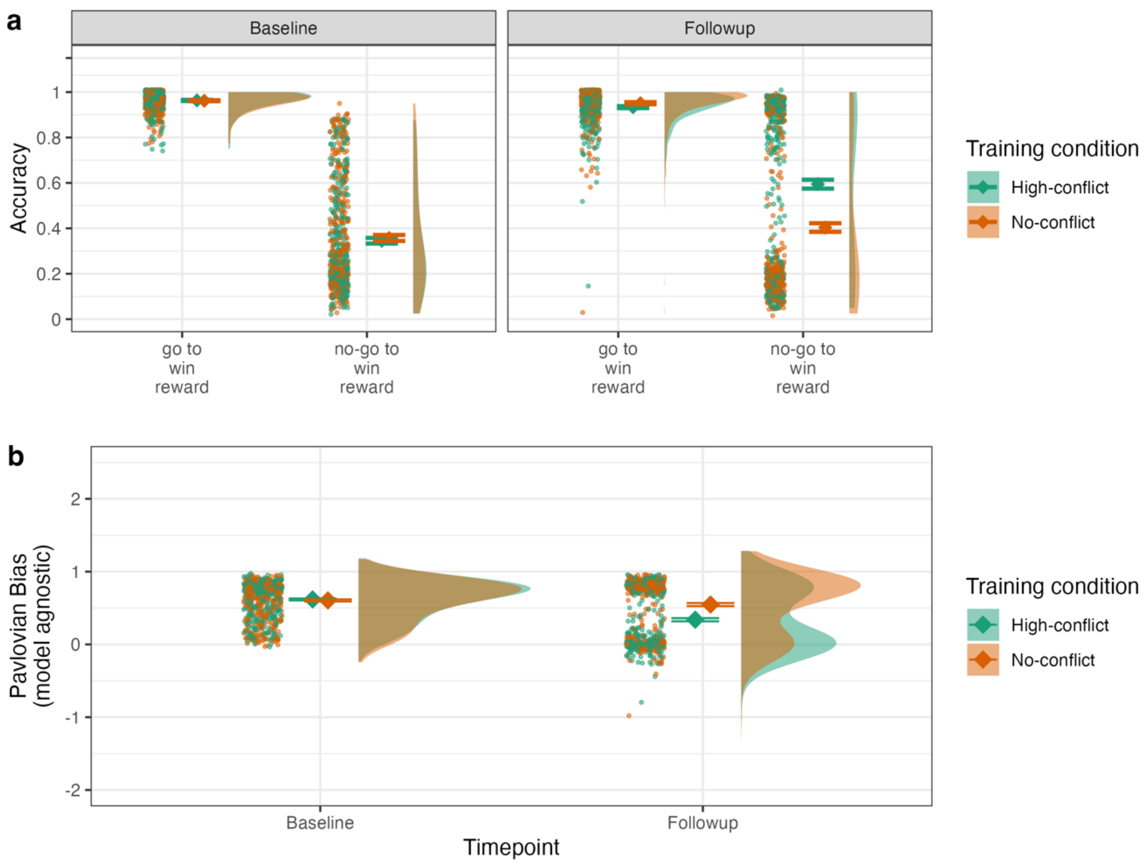


Figure 6. Simulated accuracy and Pavlovian bias from the adaptive model of Pavlovian-instrumental control (Dorfman and Gershman, 2019). Comparing with the empirical data (Figure 3) indicates that this model provides a qualitatively good fit and so may help to explain the mechanism by which the training had its effect. Note that the model does not at present handle punishment outcomes, and so we simulated just the two reward conditions.

Transfer effects

We did not observe transfer to the other cognitive tasks or self-reported mood: the change in affective bias between timepoints did not differ between groups, $t(688) = 0.11$, $p = .91$; nor was there an interaction between timepoint and group for the Risk Taking task, $F(1,688) = 0.03$, $p = .87$; depression, $F(1,688) = 0.67$, $p = .41$; state anxiety, $F(1, 688) = 1.25$, $p = .26$; or trait anxiety, $F(1, 688) = 0.56$, $p = .46$. Other analyses of these tasks are reported in the Supplement.

Discussion

In this double-blind study, we examined whether control over Pavlovian biases can be learned. Over five sessions, participants in the high-conflict group practiced the control-demanding, Pavlovian-instrumental conflict trials of the Orthogonal Go/No-Go task. Converging evidence from both model-agnostic and computational modelling analyses revealed that these participants were able to substantially reduce their Pavlovian biases (compared with participants in the no-conflict group); indeed the avoidance bias was almost entirely eliminated, according to the results of the modelling. This clearly demonstrates that people can learn to control their Pavlovian biases through training.

The existing literature has historically regarded Pavlovian biases as being highly persistent and resistant to change. Our results here suggest that, while such biases are indeed strong, they can nevertheless be overcome through training. The inclusion of a sham training condition was essential for showing that the training effect resulted specifically from practicing the high-conflict trials, and was not simply a continuation of the asymptotic performance improvement seen at baseline. The earlier work of Cavanagh et al. (2013) suggests a possible mechanism for this effect: they found that accuracy on the Go/No-Go task covaries with EEG frontal theta, a neural signature of top-down control; similarly, Guitart-Masip et al. (2012) found using fMRI that activity in the inferior frontal gyrus is associated with successful performance, and speculated that this region may help to regulate the balance between the Pavlovian and instrumental systems. Conceivably, then, in our study the training may have taught participants how and when to engage these control signals in order to mitigate Pavlovian bias and maximise performance on the task.

Training could have affected control in at least two ways: it could have increased participants' general ability to monitor for Pavlovian-instrumental conflict and deploy cognitive control as and when needed (termed 'reactive control' by Braver, 2012); or it could have led to narrower learning, in the sense that participants may simply have learned which stimuli predicted the need to exert greater control later in the trial ('proactive control'). In the present study we are not able to discriminate between these two possibilities, although we might speculate that reactive control requires more extensive training specifically aimed at transferring learning to novel situations and stimuli. A third possibility is that the training may not have impacted on control at all, and instead led to an improvement in performance through some other process, such as habit learning (Everitt & Robbins, 2005).

An important question is why the present study produced a clear training effect, while earlier attempts—such as Ereira et al. (2001)—did not. Both studies build on the original Orthogonal Go/No-Go task (Guitart-Masip et al., 2011), but differ in several critical respects. Ereira et al. conducted four experiments with motor variants of the task and found no training effects; in a fifth experiment, they reported a reduced 'semantic' bias—based on choices between words implying approach or avoidance—but this did not replicate in a non-gamified version of their task. Their task also varied stimuli across sessions, which may have disrupted consistent learning. In contrast, our study followed the original task design but held all stimulus-response mappings constant, while the training phase focussed solely on the conflict conditions, isolating the effect of repeated high-conflict exposure. Crucially, our findings go beyond methodological refinement: we provide the first causal demonstration that behavioural expressions of Pavlovian bias—particularly avoidance—can be deliberately suppressed through targeted training. This supports the view that cognitive control over these biases is not only possible, but trainable. Although constraining the training (e.g., one stimulus per condition, fixed mappings) likely limited the transfer to other contexts, it enabled a clean test of whether bias reduction could occur at all. Future studies can now build on this foundation to explore whether richer training regimes support generalisation.

The present study also constitutes a successful proof of principle for cognitive bias modification, which may have wider applications. Patients with depression or anxiety (Mkrtchian et al., 2017; Nord et al., 2018) have been shown to have enhanced Pavlovian biases, possibly as a result of deficits in cognitive control (Robinson et al., 2013), and this is thought to contribute to symptoms through the maintenance of avoidance behaviour. Our results open the possibility that cognitive bias training could eventually be used to treat some symptoms of depression and anxiety. This is a particularly exciting and significant avenue for future research because, as we have demonstrated here, such training is low-cost and can be deployed at scale through online platforms. We note that although in the current study we did not see any effects on depression or anxiety, we had specifically recruited participants with no history of psychiatric illness, so it is likely there was already a floor to improvement in symptom scores; nevertheless, we will need to examine how these results extend to patient groups in future research. Additionally, the current design focused on immediate training effects, so further work is required to assess whether reductions in Pavlovian bias are sustained over time.

One further limitation of this study is that as a control condition, we included a sham training condition only, and not a passive control (i.e. one with no training at all, which would indicate the simple effect of passage of time on the Go/No-Go task). It is conceivable that people may improve over time even without training, and that we observed a difference between the high-conflict and no-conflict groups not because the high-conflict training was effective but conversely because the no-conflict training interfered with this improvement. While we cannot definitively refute this explanation here, we suggest that the very fact that Pavlovian biases have until now been widely considered to be extremely resistant to modification indicates that improvement simply due to the passage of time is unlikely.

Finally, it is notable that the high-conflict training led to a large reduction in avoidance bias but no change in approach bias. This may be because avoidance is more amenable to change. However, it may also be driven by floor effects, as the approach bias parameters were already around zero at baseline (the effect of training was then to shrink the variance in these estimates, rather than shift the mean). It is also possible that some of the variance in approach-related responding was absorbed by the go bias parameter, thereby obscuring potential training effects on the approach bias itself. In any case, our sensitivity analysis using the Base model (which includes a single, combined Pavlovian bias term) produced qualitatively identical results. This provides reassurance that the training effect on Pavlovian bias is robust and not dependent on parametrisation or model structure.

Conclusions

In sum, this study provides causal evidence that Pavlovian bias—particularly avoidance bias—can be selectively reduced through targeted, repeated training on control demanding trials. While previous work has shown that such biases can change with context or decline over time, we show that they can be actively suppressed through focussed training. Crucially, our sham-controlled, randomised design allows us to rule out general explanations for the improvement, such as increased task familiarity, and suggest instead that it was the repeated practice of employing cognitive control which was important. These findings establish that Pavlovian biases are not fixed and instead are amenable to training, which provides a foundation for future research aimed at generalising such training to new stimuli, wider environments and clinical populations where Pavlovian biases contribute to maladaptive behaviour.

Acknowledgements

This research was supported by a Wellcome Trust 4-year PhD Studentship awarded to HF. Open Access was supported by the UK Medical Research Council.

Open Access

References

Beck, A.T., Steer, R.A., & Brown, G. K. (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. *ArXiv*:1701.02434. https://arxiv.org/abs/1701.02434.

Boureau, Y.-L., & Dayan, P. (2011). Opponency Revisited: Competition and Cooperation Between Dopamine and Serotonin. *Neuropsychopharmacology*, 36(1), 74–97. https://doi.org/10.1038/npp.2010.151

Braver, T. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Science*, 16, 106-113. doi:10.1016/j.tics.2011.12.010.

Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal Theta Overrides Pavlovian Learning Biases. *The Journal of Neuroscience*, 33(19), 8541–8548. https://doi.org/10.1523/JNEUROSCI.5754-12.2013

Chen, P., Geurts, D., Maata, J., van den Bosch, R., Hofmans, L., Papadopetraki, D., den Ouden, H., & Cools, R. (2023). Effect of striatal dopamine on Pavlovian bias. A large [18F]-DOPA PET study. *Behavioural Neuroscience, 137*(3), 184-195. https://doi.org/10.1037/bne0000547

Daniel-Watanabe, L., McLaughlin, S., Robinson, O. (2022). Association between a directly translated cognitive measure of negative bias and self-reported psychiatric symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 7 (2), 201-209. https://doi.org/10.1016/j.bpsc.2020.02.010

Dayan, P., & Huys, Q. (2008). Serotonin, inhibition and negative mood. *PLoS Computational Biology*. 4(2), e4. https://doi.org/10.1371/journal.pcbi.0040004

Dayan, P., & Balleine, B. (2002). Reward, motivation and reinforcement learning. *Neuron*, 36(2), 285-298. https://doi.org/10.1016/S0896-6273(02)00963-7

Dayan, P., Niv, Y., Seymour, B., & Daw, N. (2006). The misbehaviour of value and the discipline of the will. *Neural Networks*, 19, 1153-1160. http://dx.doi.org/10.1016/j.neunet.2006.03.002

Dickinson, A., & Balleine, B. (2002). The role of learning in the oepration of motivational systems. In H. Pashler & R. Gallistel (Eds.), Steven's handbook of experimental psychology: Learning, motivation, and emotion (pp. 497–533). John Wiley & Sons Inc.

Dorfman, H., & Gershman, S. (2019). Controllability governs the balance between Pavlovian and instrumental action selection. *Nature Communications*, 10, 5826. https://doi.org/10.1038/s41467-019-13737-7

Ereira, S., Pujol, M., Guitart-Masip, M., Dolan, R., & Kurth-Nelson, Z. (2021). Overcoming Pavlovian bias in semantic space. *Scientific Reports*, *11*(3416). https://doi.org/10.1038/s41598-021-82889-8

Everitt, B., & Robbins, T. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience, 8*, 1481-1489.

Fleming, H., Feng, G., Rutledge, R., Roiser, J., & Robinson, O. (2025). Preregistration for 'Cognitive training reduces the strength of cognitive biases'. *OSF.* https://doi.org/10.17605/OSF.IO/M3Y8U

Fleming, H., Feng, G., Rutledge, R., Roiser, J., & Robinson, O. (2025). Task code for 'Cognitive training reduces the strength of cognitive biases'. *Gorilla Open Materials.* https://app.gorilla.sc/openmaterials/669092

Fleming, H., Feng, G., Rutledge, R., Roiser, J, & Robinson, O. (2025). Repository containing data and analysis scripts for 'Cognitive training reduces the strength of cognitive biases'. *Gorilla Open Materials.* https://doi.org/10.17605/OSF.IO/7MSVW

Gabry, J., & Češnovar, R. (2021) cmdstanr: R interface to 'CmdStan'. https://mc-stan.org/cmdstanr, https://discourse.mc-stan.org

Gershman, S., Guitart-Masip, M., & Cavanagh, J. (2021). Neural signatures of arbitration between Pavlovian and instrumental action selection. *PLoS Comput Biol, 17*(2), e1008553.

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, 18(4), 194–202. https://doi.org/10.1016/j.tics.2014.01.003

Guitart-Masip, M., Economides, M., Huys, Q., Frank, M., Chowdhury, R., Duzel, E., Dayan, P., & Dolan, R. (2014). Differential but not opponent effects of L-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology, 231*, 955-966. https://doi.org/10.1007/s00213-013-3313-4

Guitart-Masip, M., Fuentemilla, L., Bach, D., Huys, Q., Dayan, P., Dolan, R., & Duzel, E. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *The Journal of Neuroscience*, 31(21), 7867-7875. https://dx.doi.org/10.1523%2FJNEUROSCI.6376-10.2011

Guitart-Masip, M., Huys, Q., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *Neuroimage, 62(*1), 154-166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Hershberger, W. A. (1986). An approach through the looking glass. *Animal Learning and Behaviour. 14, 443-451.*

Kaser, M., Zaman, R., Sahakian, B. (2017). Cognition as a treatment target in depression. *Psychological Medicine, 47*, 987-989. https://doi.org/0.1017/S0033291716003123

Kassambara, A. (2021). Rstatix: Pipe friendly framework for basic statistical tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., & Robinson, O. J. (2017). Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. *Biological Psychiatry*, 82(7), 532–539. https://doi.org/10.1016/j.biopsych.2017.01.017

Nord, C., Lawson, R., Huys, Q., Pilling, S., &Roiser, J. (2018). *Scientific Reports*, 8, 12582. https://doi.org/10.1038/s41598-018-30828-5

Peters, S., Lumsden, J., Peh, O., Penton-Voak, I., Munafò, M. R., & Robinson, O. J. (2017). Cognitive bias modification for facial interpretation : a randomized controlled trial of transfer to self-report and cognitive measures in a healthy sample *R. Soc. Open Sci*., 4(170681). https://doi.org/10.1098/rsos.170681

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II, A.H. Black & W.F. Prokasy, Eds., pp. 64–99. Appleton-Century-Crofts.

Robinson, O. J., Vytal, K., Cornwell, B. R., & Grillon, C. (2013). The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Frontiers in Human Neuroscience*, 7(May), 1–21. https://doi.org/10.3389/fnhum.2013.00203

Rutledge, R. B., Smittenaar, P., Zeldman, P., Brown, H., Adams, R., Lindenberger, U., Dayan, P., & Dolan, R. (2016). Risk Taking for Potential Reward Decreases across the Lifespan. *Current Biology*, 26, 1634-1639. https://doi.org/10.1016/j.cub.2016.05.017

Schultz, W., Dayan, P., & Montague, R. (1997) A neural substrate of prediction and reward. *Science*, 1593-1599.

Schurr, R., Reznik, D., Hillman, H./, Bhui, R., & Gershman, S. (2024). Dynamic computational phenotyping of human cognition. *Nature Human Behaviour, 8*, 917-931.

Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, J. G. (1983). Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press.

Stan Development Team. (2023). Stan modelling language users guide and reference manual. https://mc-stan.org

Sutton, R. (1988). Learning to predict by the methods of temporal differences. *Machine Learning, 3*, 9-44.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely application information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571-3594. doi:10.5555/1756006.1953045.

Williams, D., & Williams, H. (1969) Automaintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behaviour, 12*, 511-520.

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

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

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

$$pGo_t = logistic\big(w(s_t)\big) \tag{2a}$$

$$response_t = Bernoulli(pGo_t) \tag{2b}$$

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

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

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

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

Priors

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

$$\begin{aligned} \mu_{GoBias} &\sim Normal(0,2) & \sigma_{GoBias} &\sim Exponential(1) \\ \mu_{PavBias} &\sim Normal(0,2) & \sigma_{PavBias} &\sim Exponential(1) \\ \mu_{LR} &\sim Normal(0,1) & \sigma_{LR} &\sim Exponential(1) \end{aligned}$$

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

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

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

}

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

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

}

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.

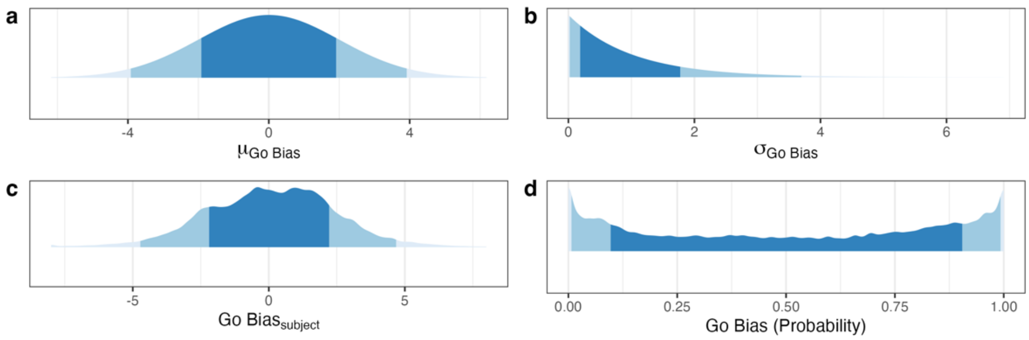


Figure S2. Prior predictions for the Pavlovian bias parameters, with distributions $\mu_{PavBias} \sim Normal(0,2)$, $\sigma_{PavBias} \sim Exponential(1)$ and $PavBias_{subject} \sim Normal(\mu_{PavBias}, \sigma_{PavBias})$. (a) and (b) show the analytical distributions of the population mean and standard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

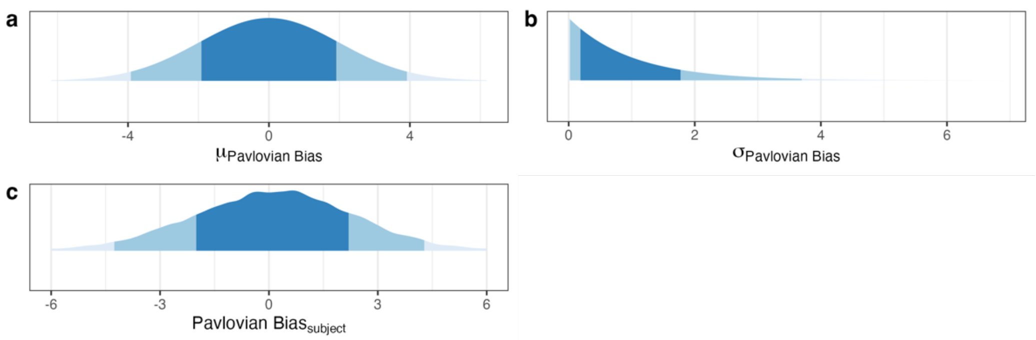


Figure S3. Prior predictions for the outcome sensitivity parameters, with distributions $\mu_{sensitivity} \sim Normal(0,1.5)$, $\sigma_{sensitivity} \sim Exponential(1)$ and $sensitivity_{subject} \sim Normal(\mu_{sensitivity}, \sigma_{sensitivity})$. (a) and (b) show the analytical distributions of 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).

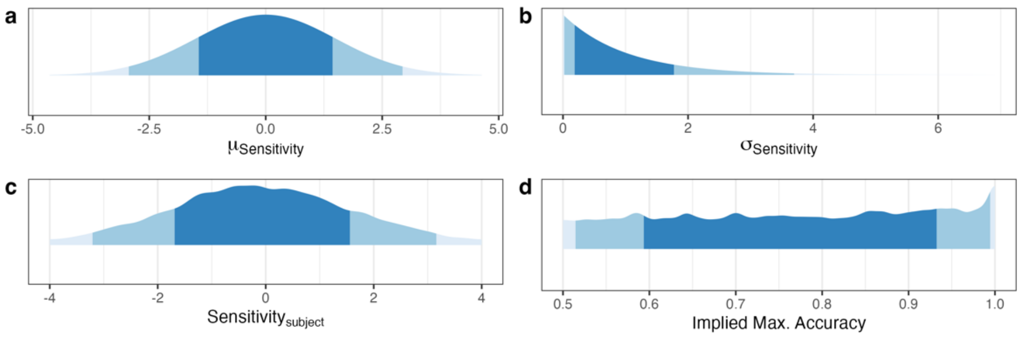
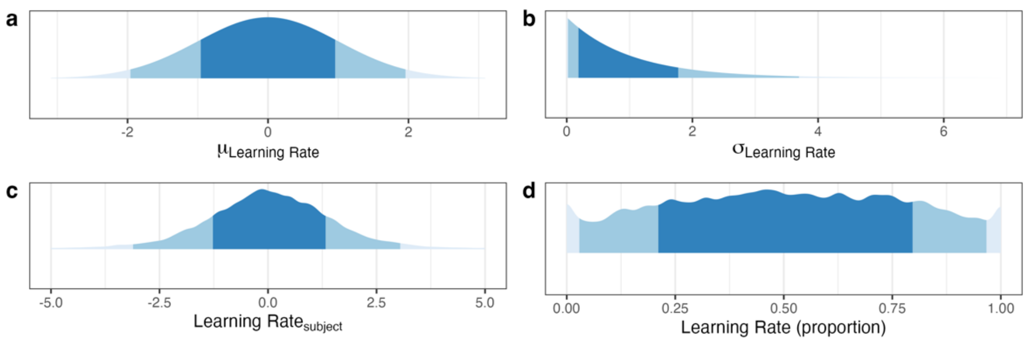


Figure S4. Prior predictions for the learning rate parameters, with distributions $\mu_{LR} \sim Normal(0,1)$, $\sigma_{LR} \sim Exponential(1)$ and $LR_{subject} \sim Normal(\mu_{LR}, \sigma_{LR})$. (a) and (b) show the analytical distributions of the population mean and standard deviation; (c) shows 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: $p = .70$).

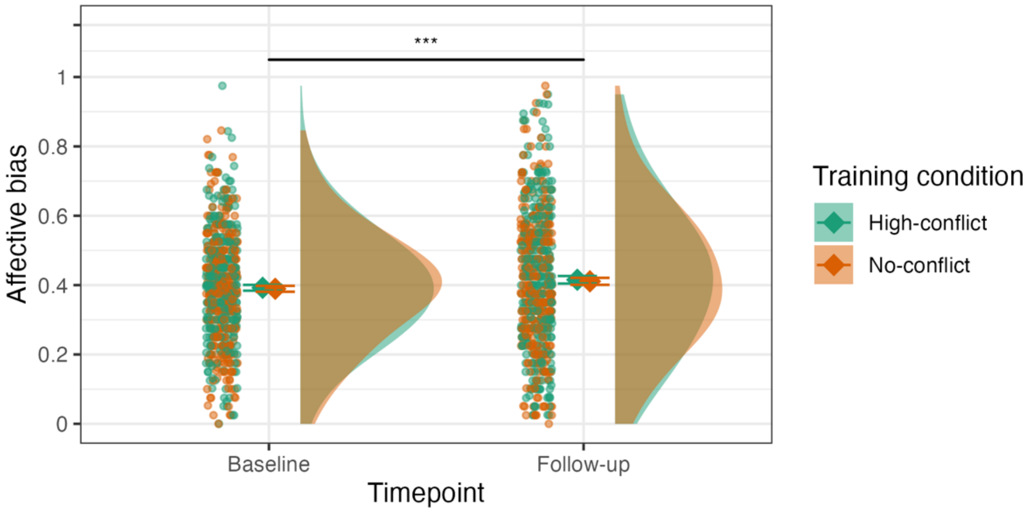


Figure S6. Affective bias before and after training. Affective bias is measured by the proportion of responses matching the ambiguous stimuli to the high-reward exemplar (a value of 0.5 is neutral, <0.5 is a negative bias and >0.5 is a positive bias). There was a significant main effect of timepoint 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.

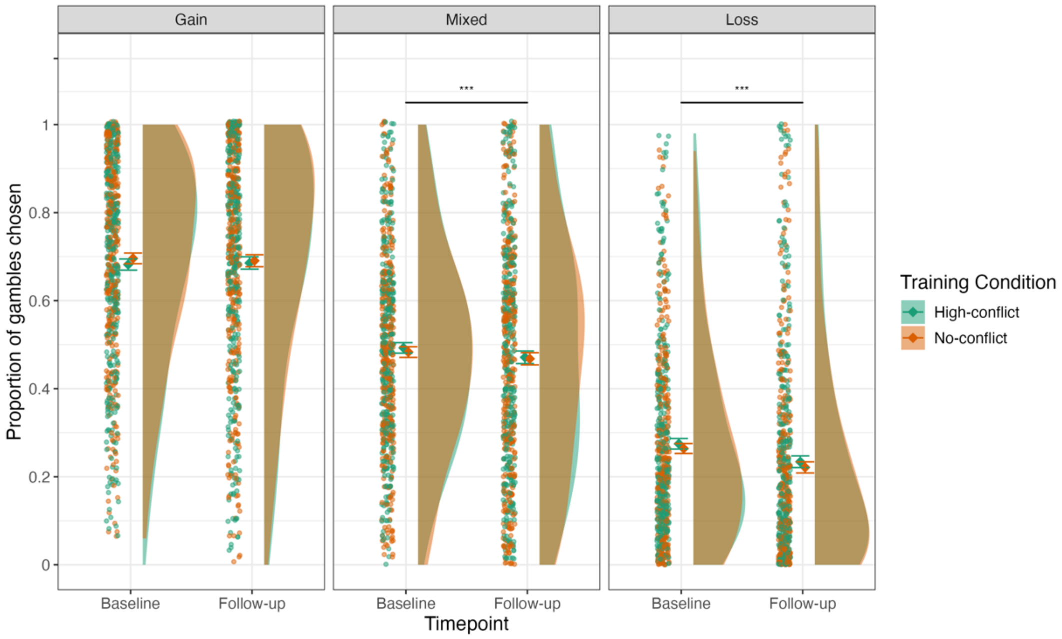


Figure S7. Risk Taking task: Proportions of gambles chosen. Participants chose to gamble less often in the Follow-up session, but only in the mixed and loss frames. In addition, overall the rates of gambling were higher in the gain frame compared with the mixed frame, which in turn was higher than in 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.

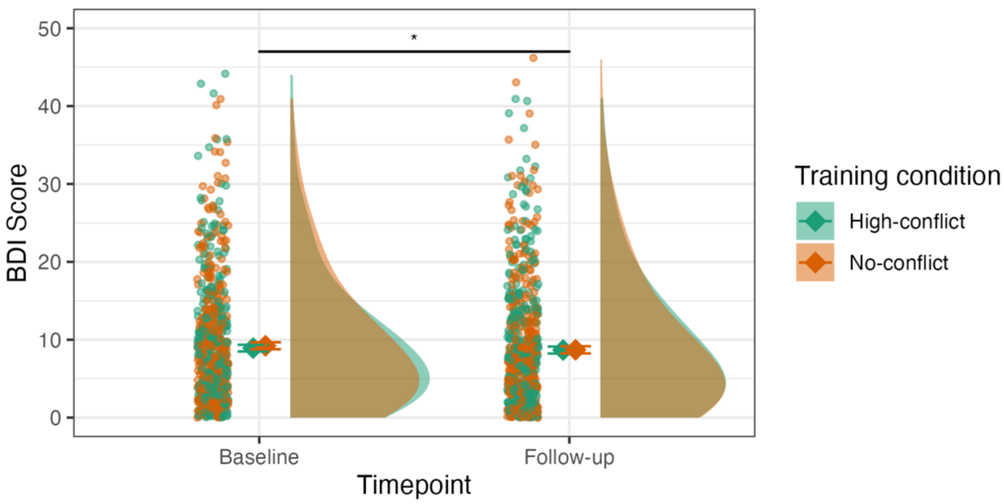


Figure S8. Beck Depression Inventory scores. Scores decreased significantly from Baseline to Follow-up, for both training groups. Plot shows (left to right) individual data points, mean±SE and distributions. * $p < .05$

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

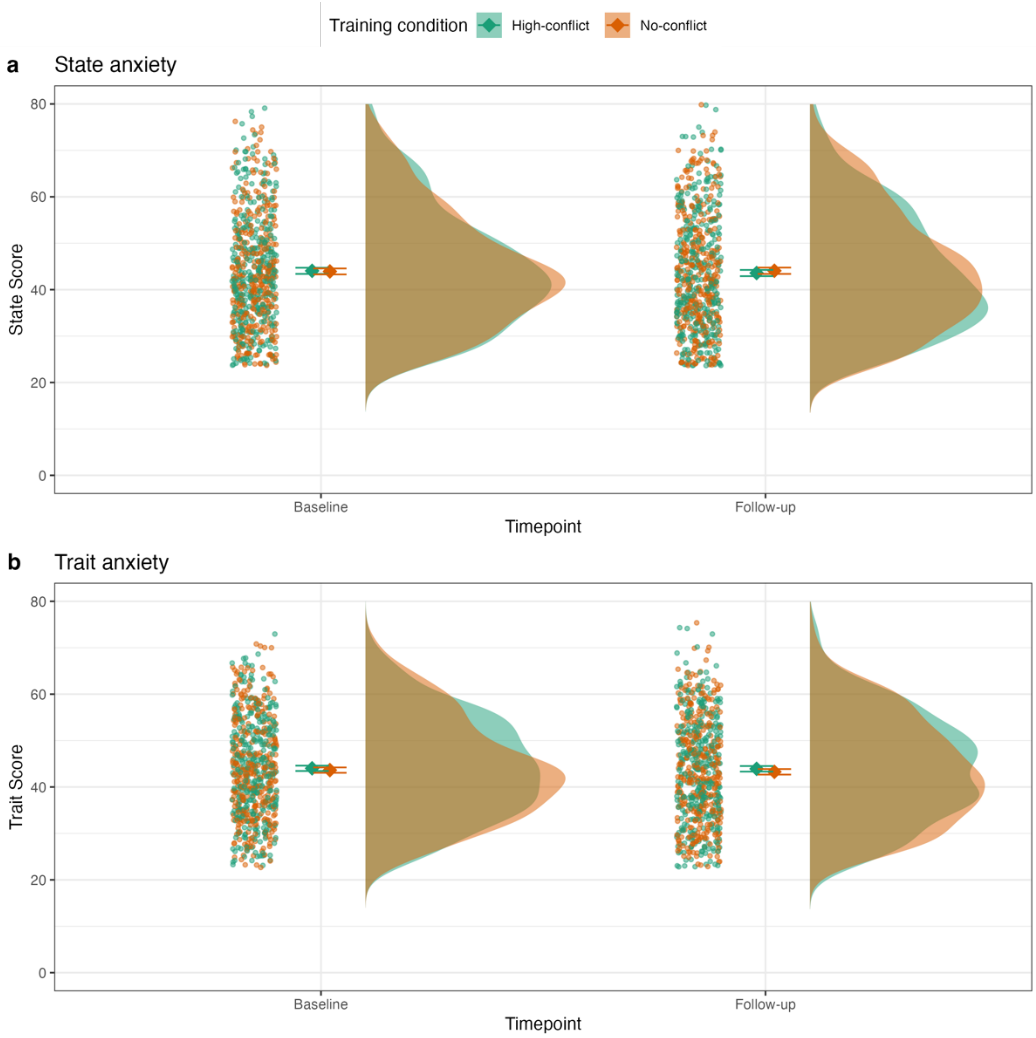


Figure S9. State-Trait Anxiety Inventory Scores. There were no differences between either timepoints or training groups. Plots show (left to right) individual data points, mean±SE and distributions.

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.

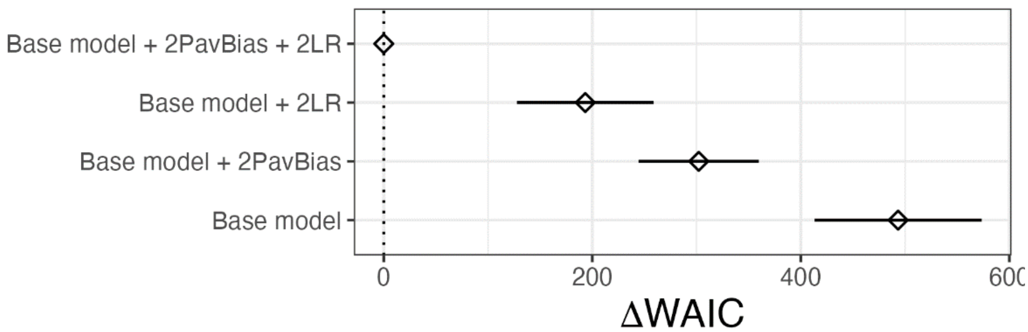


Figure S10. Model comparison results for the Orthogonal Go/No-Go task. Plots show the difference in WAIC (and SE of this difference) between the best performing model, indicated by the vertical dotted line, and each of the models in turn. The best model by nearly three standard errors of difference constituted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

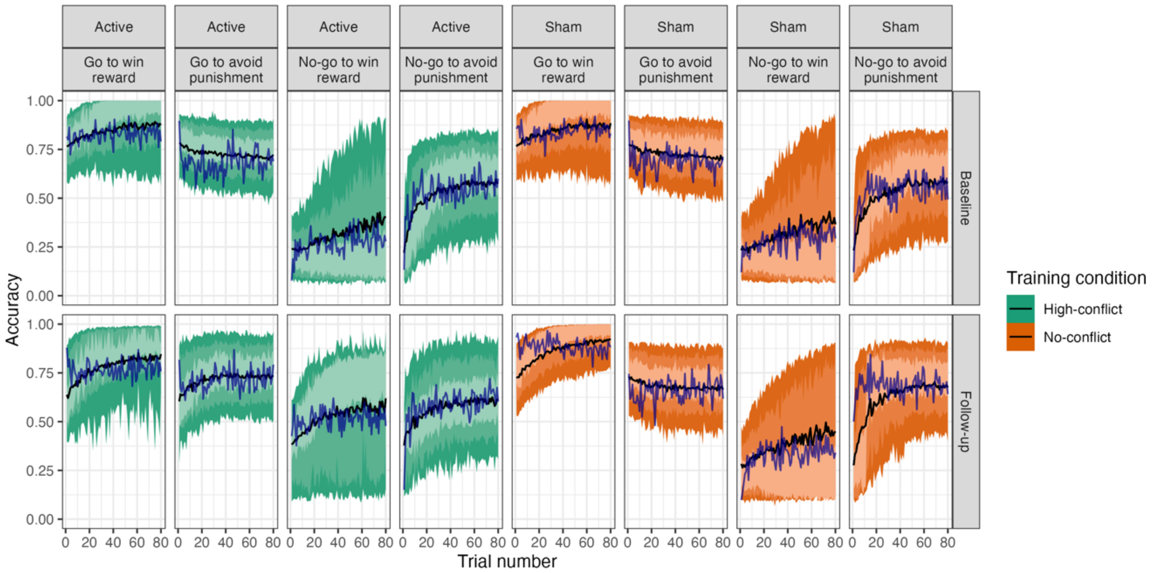


Figure S11. Posterior predictions from the winning model (Base model +2Pavbias + 2LR). Plots show the mean (black line) and 50/80/95% highest density continuous intervals (HDCI) for the posterior predictions, as well as the empirical data (blue line) for comparison.

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.

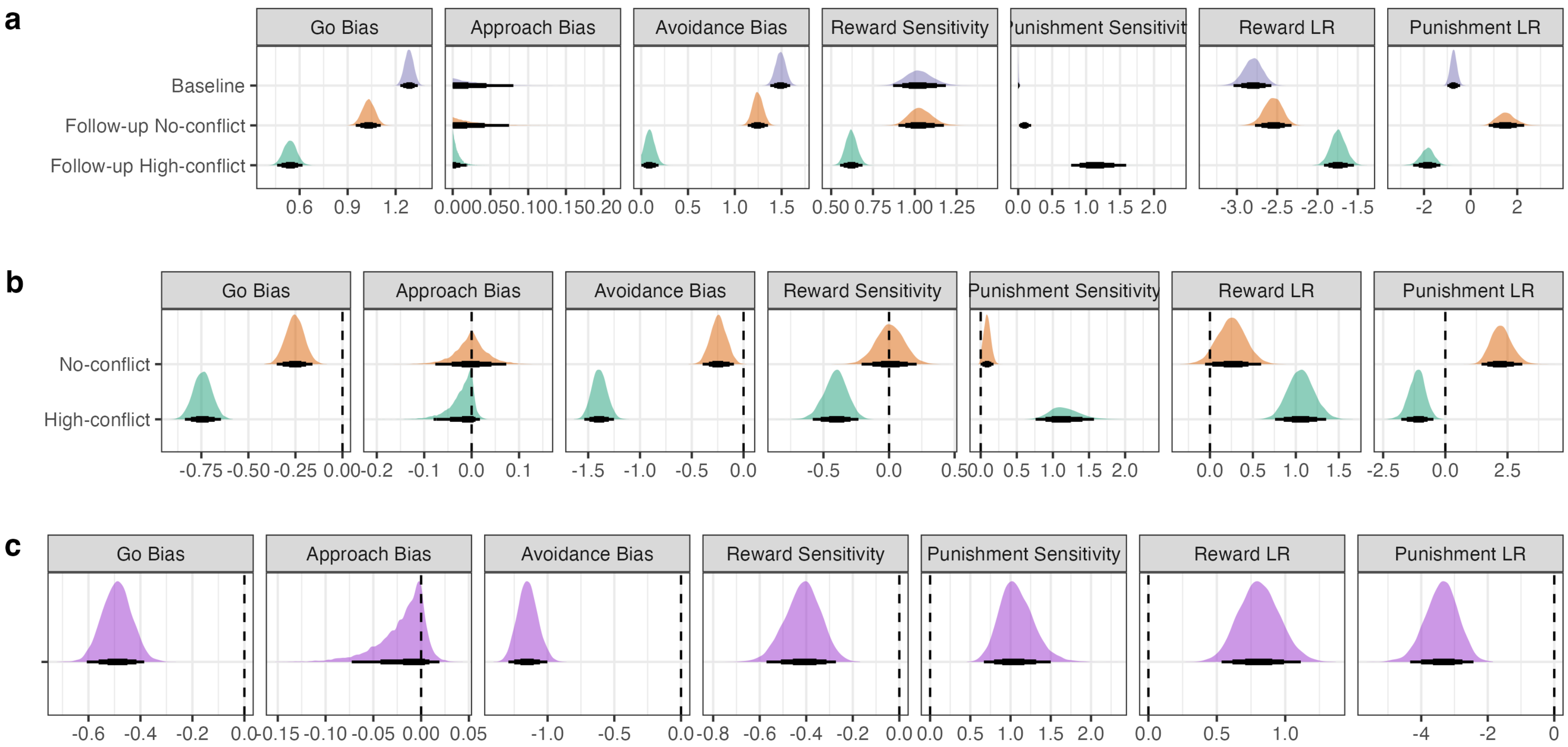


Figure S12. Posterior estimates of the population-level parameters according to the winning model (Base plus two Pavlovian biases plus two learning rates). Each plot shows the posterior distribution and the 50/80/95% highest density continuous intervals (black lines) for the parameter shown. (a) The 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, $r = 0.37$, $p < .001$

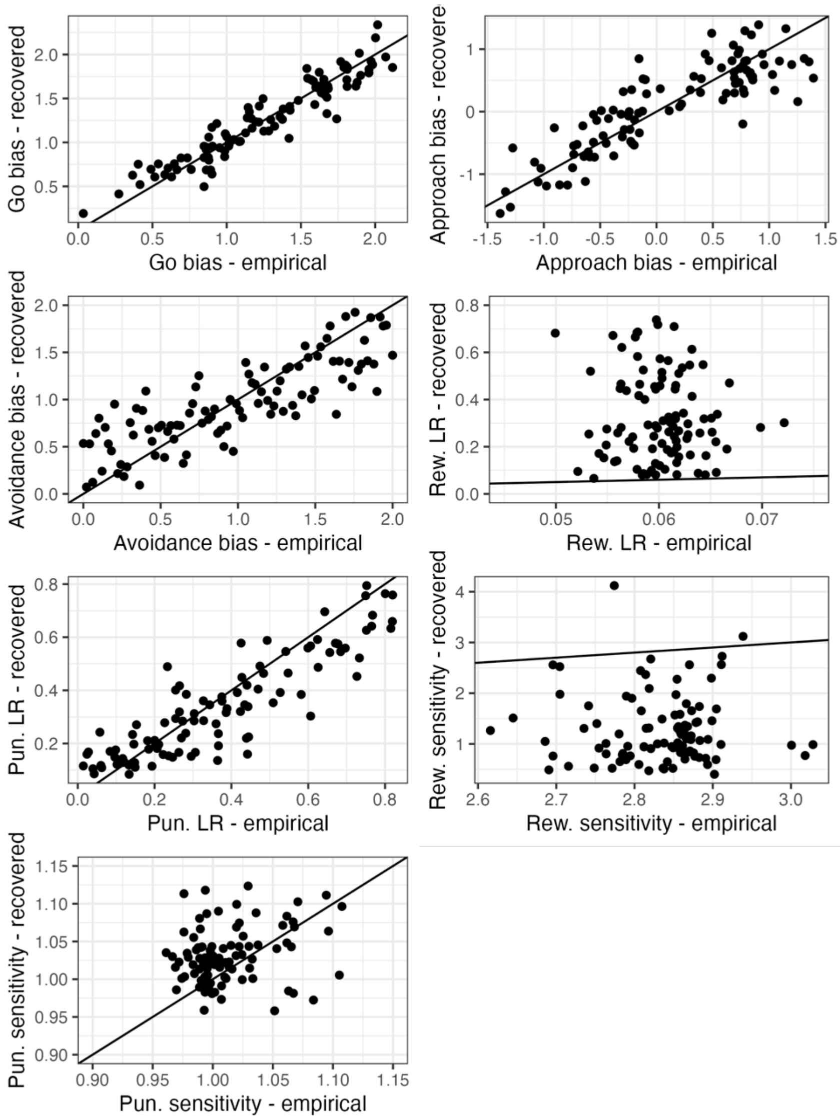


Figure S13. Results of the parameter recovery for the winning model, Base+2Pav+2LR. The black line indicates perfect recovery, y = x.

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.

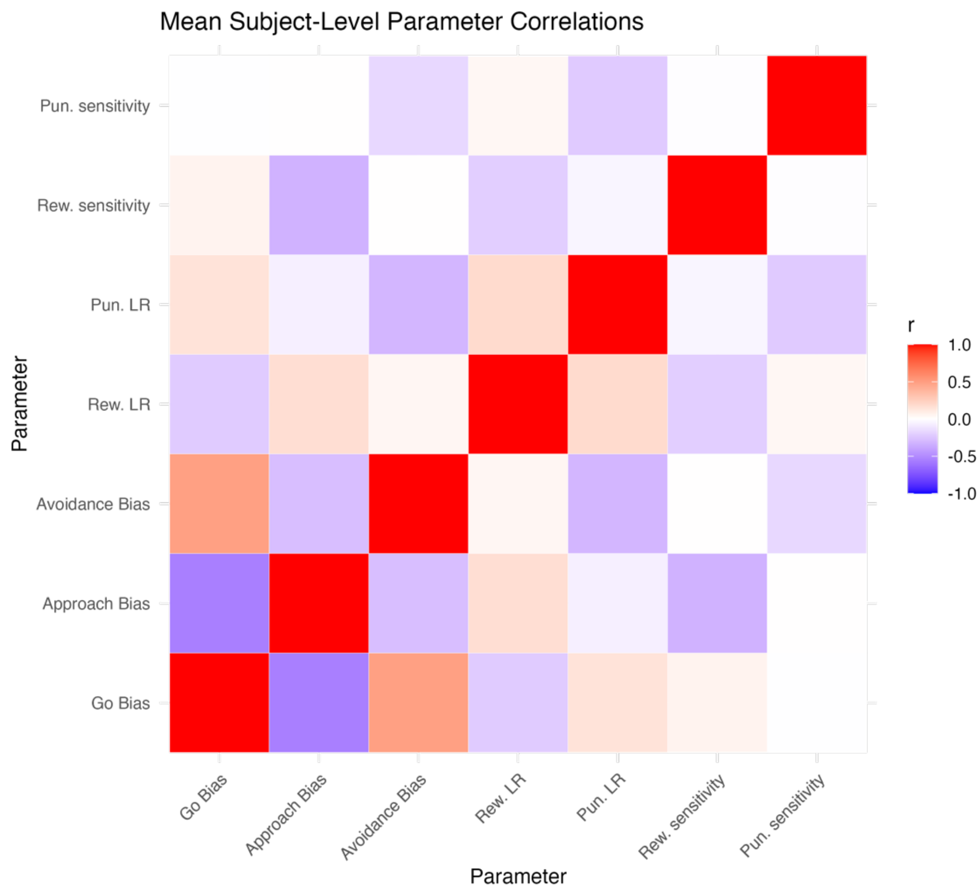


Figure S14. Average within-subject correlations for each pair of parameters. Although the reward sensitivity and learning rate parameters had been the hardest to recover, they showed relatively low correlations with other parameters, suggesting that they were not trading off against each other or other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

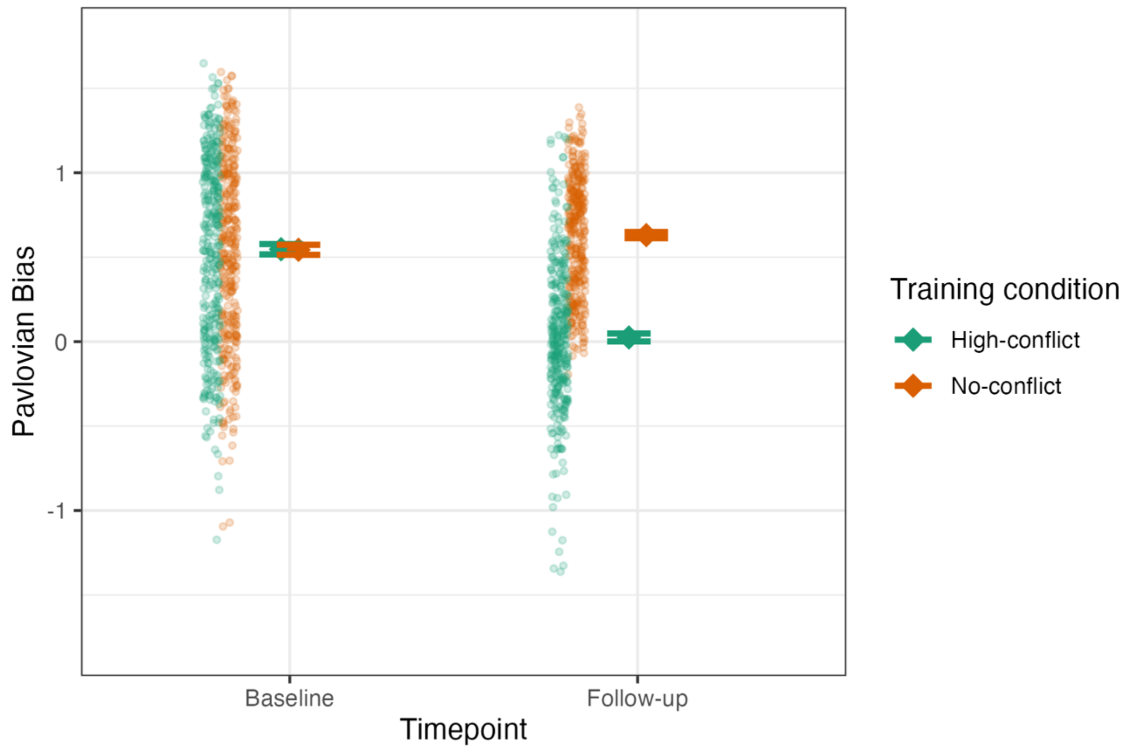


Figure S15. Subjects' Pavlovian bias values according to the simplest 'Base' model. Results are equivalent to those from the winning model; compare with Figure 4 in the main text.

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.