A core component of psychological therapy causes adaptive changes in computational learning mechanisms

Quentin Dercon[1,2†] ✉, Sara Z. Mehrhof[1†], Timothy R. Sandhu[1,3], Caitlin Hitchcock[1,4], Rebecca P. Lawson[1,3], Diego A. Pizzagalli[5,6], Tim Dalgleish[1,7], and Camilla L. Nord[1] ✉

[1]Medical Research Council (MRC) Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK
[2]Department of Psychiatry, University of Oxford, Oxford, UK
[3]Department of Psychology, University of Cambridge, Cambridge, UK
[4]Department of Psychology, University of Melbourne, Melbourne, Australia
[5]Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[6]Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, MA, USA
[7]Cambridgeshire and Peterborough NHS Foundation Trust, UK

[†]These authors contributed equally to this work.
✉Address correspondence to: Quentin.Dercon@mrc-cbu.cam.ac.uk or Camilla.Nord@mrc-cbu.cam.ac.uk

Abstract

Cognitive distancing is a therapeutic technique commonly used in psychological treatment of various mental health disorders, but its computational mechanisms remain unknown. To determine the effects of cognitive distancing on computational learning mechanisms, we use an online reward decision-making task, combined with reinforcement learning modelling in 935 participants, 49.1% of whom were trained to regulate their emotional response to task performance feedback. Those participants practicing cognitive distancing showed heightened learning from negative events as well as an increased integration of previous choice values. These differences seemed to represent an evolving shift in strategy by the distancing participants during the task, from exploiting optimal choices earlier in the task (as indicated by greater inverse temperature parameters), to a late-stage increase in learning from negative outcomes (represented as higher loss learning rates). Our findings suggest adaptive changes in computational learning mechanisms underpin the clinical utility of cognitive distancing in psychological therapy.

1 Main

> Why do those cliffs of shadowy tint appear More sweet than all the landscape smiling near?—'Tis distance lends enchantment to the view, And robes the mountain in its azure hue.
>
> Thomas Campbell
> *The Pleasures of Hope* (1799)

Cognitive distancing is a psychological skill in which an individual views a negative thought from afar, reducing its emotional impact and enhancing their ability to challenge or reframe the thought. Cognitive distancing strategies are a central component of many psychological therapies[1] including cognitive behavioural therapy (CBT)[2], dialectical behaviour therapy[3], and mindfulness-based cognitive therapy[4]. These strategies enable targeting of emotion regulation difficulties, which occur across mood, substance use, eating, and personality disorders, but improve following effective psychological treatment[1] and may also predict treatment response[5].

Despite decades of research on the psychological and neural bases of emotion regulation[6,7], the specific mechanisms of skills such as cognitive distancing remain unknown. Standard psychological measures cannot distinguish effects of cognitive distancing on subcomponents of motivated behaviour, such as reward sensitivity and learn-

ing[8], nor determine whether distancing from the immediate situation is adaptive over time. This limits our understanding of why cognitive distancing is therapeutically beneficial. Unlike standard measures, theory-driven models from the field of computational psychiatry[8] could enable us to decouple the effects of cognitive distancing on subcomponents of motivated behaviour such as reward sensitivity and learning, potentially elucidating their specific therapeutic mechanisms as they evolve over time[9].

One major proposed mechanism-of-action of pharmacological treatments for psychiatric disorders is through changing reinforcement learning, which appears compromised in numerous psychiatric disorders[10]. For example, in depression, meta-analyses have found evidence for low-to-moderate reinforcement learning decrements[11], perhaps due to reward insensitivity[12], with converging evidence that antidepressant administration in healthy participants improves reward learning[13]. Similar computational approaches have helped elucidate the mechanisms of pharmacological treatments of depression[14], schizophrenia[15], and bipolar disorder[16]. However, there is substantial between-study heterogeneity[11]: individual studies variously report heightened sensitivity to punishment[17], reward[18], feedback across valences[19], and no difference in reinforcement learning between depressed patients and controls[20]. One explanation for this is that general disturbances in reinforcement learning transcend diagnostic boundaries: recent work has found that differences in goal-directed control[21] and aversive learning[22] map onto transdiagnostic symptom dimensions better than diagnoses. This suggests that a dimensional approach may better describe the nature of reinforcement learning in psychiatric disorders and their treatment, echoing calls from the clinical psychological literature for transdiagnostic treatments[23].

Given that emotion regulation strategies alter reward processing[24], cognitive distancing may also affect reward learning transdiagnostically. Recent work indicates that CBT affects reward learning[18], altering how patients assign value to actions, and the extent to which these are updated by different environmental signals. These representations, formalised in a reinforcement learning framework as expected values (EVs) and prediction errors (PEs) respectively[25], have been consistently linked to heterogeneous dopamine signals of midbrain origin[26]. In line with this, neuroimaging studies have suggested that cognitive distancing ("adopt[ing] the position of a neutral observer"[27]) during learning tasks may affect midbrain representations of EVs and PEs, decreasing striatal responses to reward anticipation[24,27] via prefrontal cortex modulation[28,29]. However, the computational underpinnings of these representations – for example, whether cognitive distancing affects the extent to which PEs are used to update EVs (learning rate) or the extent to which differences in EVs between options determine choices (inverse temperature) – are not known.

Here we test whether and how cognitive distancing alters reinforcement learning, using a probabilistic selection task (PST)[30,31] combined with an established computational model to decompose components of participant behaviour. By recruiting a large online sample ($n = 995$) broadly representative of the UK population, we hoped to measure effects of cognitive distancing on reinforcement learning, and quantify associations between reinforcement learning and data-driven transdiagnostic symptom dimensions[21]. Our study methods and analyses were preregistered on OSF.

2 Results

Nine hundred and ninety-five participants completed a learning task (PST[30,31], Figure 1**C**) alongside a psychiatric questionnaire question battery used to predict scores on three transdiagnostic factors[21] (Figure 1**B,D**). Crucially, half the participants ($n = 497$) were randomised to a cognitive distancing manipulation where they were instructed to 'take a step back' in response to feedback during training trials. Q-learning models[32] with single or dual learning rates[31] were included in the model space. These models assume that EVs of the stimuli in each pair are updated in response to feedback; higher learning rate parameter values (α) indicate increased sensitivity of choices to recent feedback (single learning rate), or in dual learning rate models, differential sensitivity to positive or negative feedback[31] (reward or loss learning rate; α_{reward} or α_{loss}). Each model also included an inverse temperature parameter (β), which captures the extent to which differences in EVs determine choices between symbols.

Models were fit to training choices alone, and to training plus test phase choices in a hierarchical Bayesian manner[33,34], separately in distanced and non-distanced participants (see Methods 4.2). As the test phase lacked feedback, in line with previous work[31], results from training-plus-test models were interpreted as the parameter values at the end of training, fit to test phase choices. The effect on learning parameters of the distancing manipulation, and of increases in each transdiagnostic factor score, was then quantified using Bayesian generalised linear models (GLMs) adjusting for age, sex, and digit span.

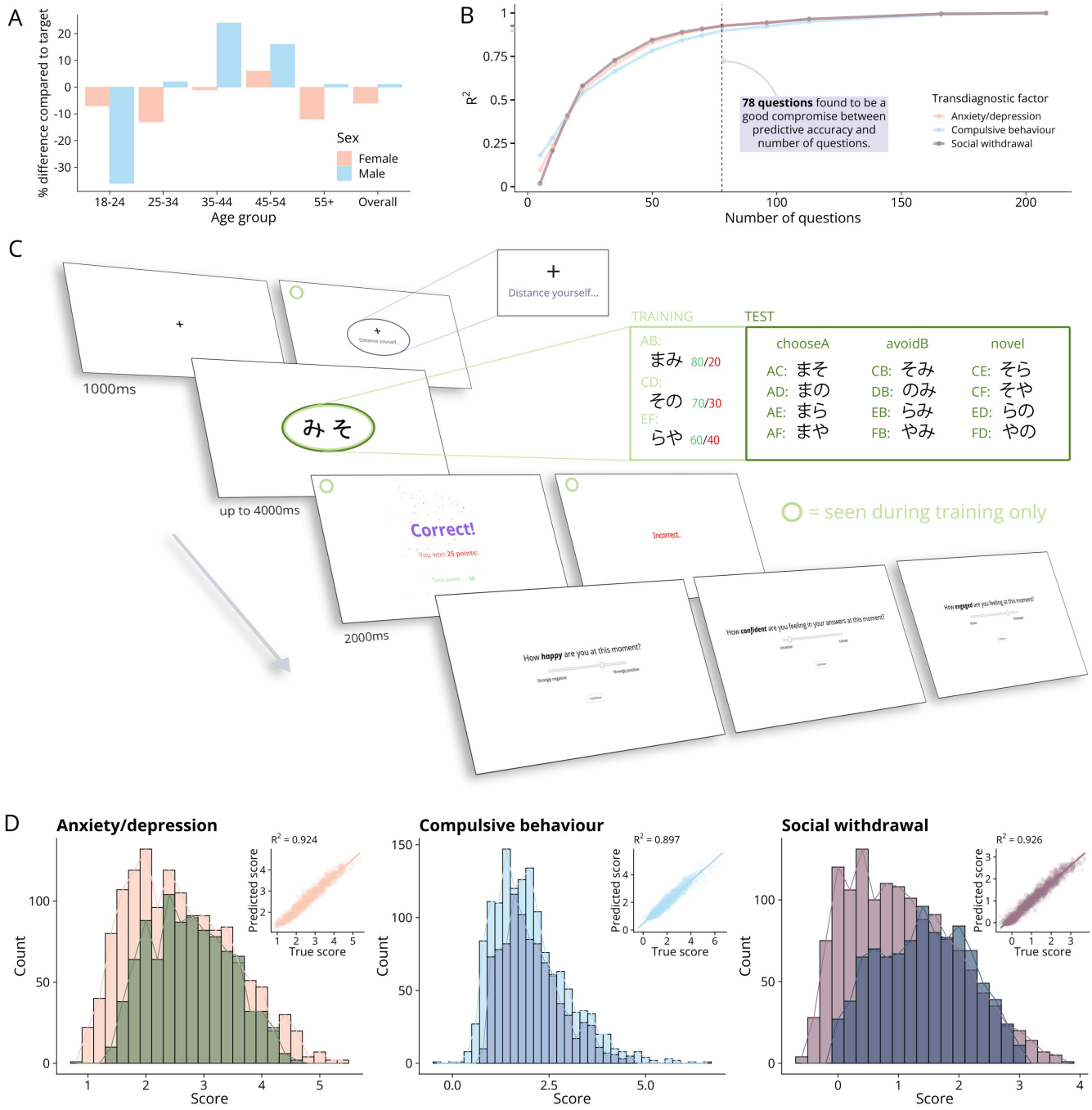

Figure 1: **Task setup, distancing, and transdiagnostic factor derivation**. **A**. Percentage difference between the numbers in each age/sex group (excluding $n = 5$ non-binary individuals) reporting a diagnosed psychiatric disorder, compared to the targeted numbers calculated from UK population data based on 1,000 participants (2011 Census[35] and the 2014 Adult Psychiatric Morbidity Survey[36]). **B**. Multi-target lasso regression with five-fold cross-validation was used to predict the three transdiagnostic dimension factor scores from different subsets of the 209 questions in the original dataset (Gillan *et al.*[21], study 2). 78 questions were found to predict the factor scores with high predictive accuracy. **C**. The probabilistic selection task (PST) in the present study consisted of six blocks of sixty trials (training phase) where participants were instructed to choose between Hiragana characters presented as three pairs (AB, CD, EF; twenty of each per block), and given feedback ('Correct!' or 'Incorrect.'). One character in each of the pairs was consistently more likely to be correct (reward probabilities of 0.8/0.2, 0.7/0.3, and 0.6/0.4 for A/B, C/D, and E/F respectively. 497 participants (49.9%) were randomised to a self-distancing intervention, and additionally received a prompt to "Distance yourself…" with the fixation cross at the start of each training trial. The training phase was followed by a sixty-trial test phase without feedback where the twelve other possible character combinations were added (i.e., four of each of the fifteen pairs). All participants saw the same six characters, but the pairs themselves were randomised for each participant, and the order of the pairs was counterbalanced across trials. One of three affect questions was asked after each trial, with unlimited time to answer. **D**. Comparison between the factor score distributions for the $n = 935$ non-excluded participants in the present study (darker colours), and those previously obtained by Gillan *et al.*[21] in $n = 1413$ participants (lighter colours). Inset plots show the predictive validity of the subset of 78 questions in predicting these scores in the original dataset.

2.1 Sample characteristics

After applying exclusion criteria (see Methods 4.1.7), a total of 935 participants (49.1% distanced) were included in analyses. We recruited a sample that closely resembled the UK population in terms of age, sex, and history of a diagnosed psychiatric disorder (Table 1; Figure 1**A**). Mean compulsive behaviour factor scores in non-excluded participants were comparable to a similarly-large online sample[21] (2.05 vs. 2.00; $t(2286.1) = 1.24$, $p = 0.22$); mean scores for the anxiety/depression factor were slightly higher (2.76 vs. 2.59; $t(2287) = 5.04$, $p < 0.001$); and mean scores for social withdrawal markedly higher (1.41 vs. 1.06; $t(2225.7) = 10.2, p < 0.001$) (Figure 1**D**).

Table 1: Demographic characteristics of the sample, by group.

	Non-distanced	Distanced	Excluded
Cohort size	476	459	60
Demographics			
Age, mean (standard deviation (SD); range)	45.3 (14.9; 18-86)	45.5 (15.6; 18-83)	40.0 (16.1; 18-74)
Gender, number (%)			
Male	231 (48.5)	229 (49.9)	25 (41.7)
Female	243 (51.1)	229 (49.9)	33 (55.0)
Non-binary	2 (0.4)	1 (0.2)	2 (3.3)
Ethnicity, number (%)			
White	431 (90.5)	405 (88.2)	51 (85.0)
Asian	23 (4.8)	28 (6.1)	6 (10.0)
Black	9 (1.9)	9 (2.0)	2 (3.3)
Mixed	12 (2.5)	13 (2.8)	1 (1.7)
Other	1 (0.2)	4 (0.9)	0 (0.0)
Subjective socioeconomic status (/9), median (interquartile range (IQR))	5 (4-6)	5 (4-6)	5 (4-6)
English proficiency, number (%)			
Intermediate or lower ($\leq$ B1)	0 (0.0)	0 (0.0)	3 (3.3)
Can read Hiragana characters, number (%)	3 (0.6)	9 (2.0)	0 (0.0)
Comorbidities			
Neurological disorder, number (%)	0 (0.0)	0 (0.0)	18 (30.0)
Psychiatric disorder (ever diagnosed), number (%)			
Generalised or social anxiety disorder	97 (20.4)	101 (22.0)	16 (26.7)
Major depressive disorder	33 (6.9)	37 (8.1)	5 (8.3)
Any	122 (25.6)	132 (28.8)	20 (33.3)
Current medications, number (%)			
Antidepressant (any)	61 (12.8)	69 (15.0)	14 (23.3)
Anti-hypertensive, anti-coagulant, or statin	49 (10.3)	62 (13.5)	6 (10.0)
Contraceptive pill	24 (5.0)	26 (5.7)	2 (3.3)
Painkiller	25 (5.3)	30 (6.5)	11 (18.3)
Task performance			
Time taken (minutes), mean (SD)	93.9 (25.1)	93.9 (25.5)	90.6 (24.0)
Mean RT during training, mean (SD)	1117.8 (356.1)	1157.9 (363.7)	1002.0 (344.4)
Digit span, median (IQR)	7 (6-8)	7 (6-8)	6.5 (6-8)

2.2 Model comparison

Models were compared using two numerical metrics of out-of-sample predictive accuracy: the expected log posterior density (higher values indicate a better model), and the leave-one-out information criterion (lower values indicate a better model)[37]. In both non-distanced and distanced participants, and across models fit to training alone and models fit to training-plus-test, there was consistent evidence suggesting Q-learning models with dual learning rates had better estimated out-of-sample predictive accuracy than more parsimonious single learning rate models (Figure 2**A**). We report results from all models, as they are largely comparable.

2.3 Associations between learning parameters and transdiagnostic psychiatric symptom dimensions

Individual-level posterior mean parameter values from single or dual learning rate Q-learning models fit to training

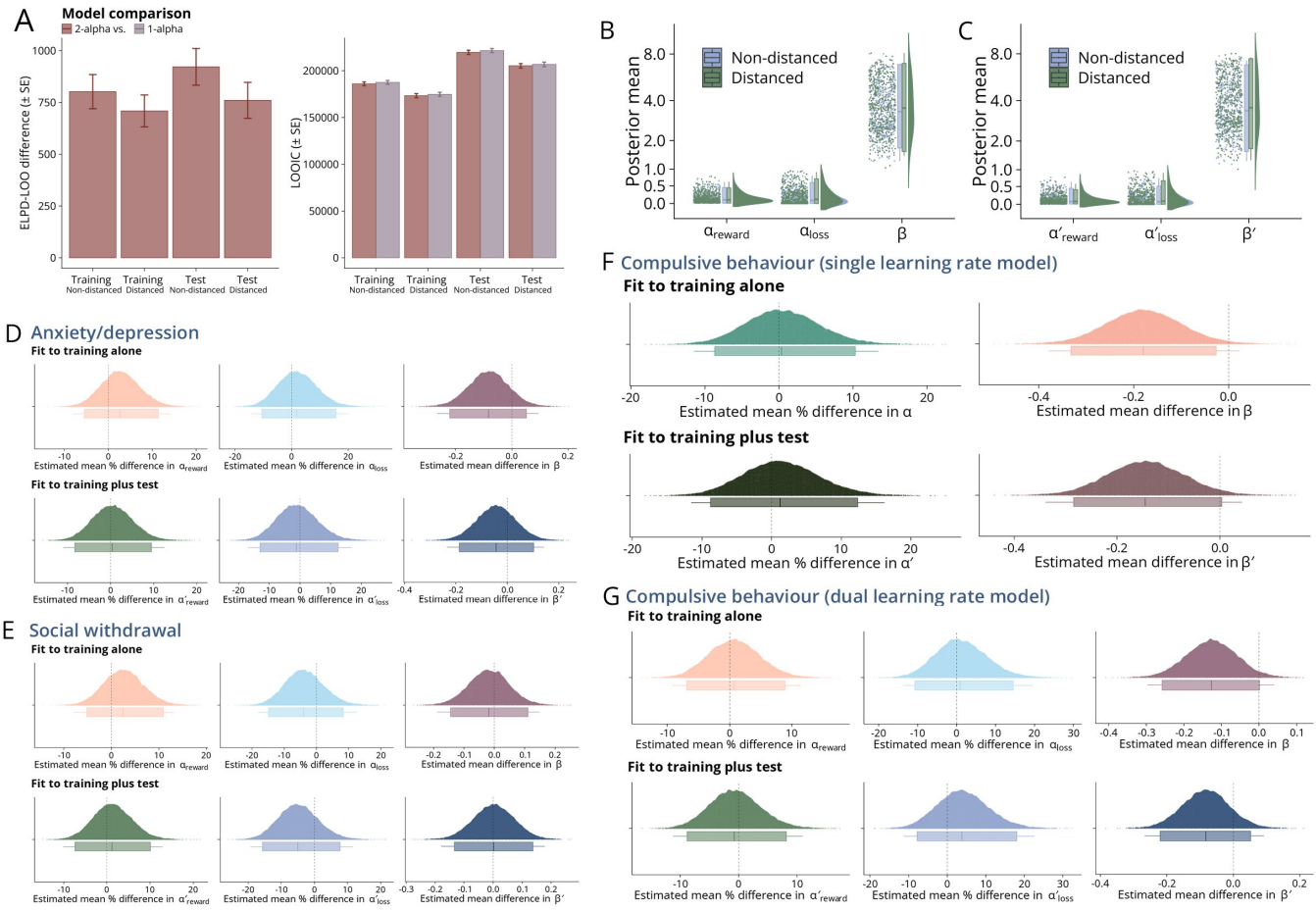

Figure 2: **Model comparison, parameter distributions, and associations between learning parameters and transdiagnostic psychiatric symptom dimensions**. **A**. Difference in numerical fit metrics between the dual and single learning rate models fit to training data alone, or training-plus-test, by distancing group. expected log posterior density (ELPD) is the expected log posterior density (higher is better), and leave-one-out information criterion (LOOIC) is the leave-one-out information criterion (lower is better). **B-C**. Distributions of individual-level posterior means for learning parameters from the fits to training (**B**) and test (**C**) data. **D-G**. Coefficient posterior distributions from Bayesian GLMs (adjusted for age, sex, digit span and distancing status) reflecting the estimated percentage change in the learning rate parameter or the estimated mean change in the inverse temperature for a unit increase in each of the three transdiagnostic factor scores, from Q-learning models fit to training alone or training-plus-test (parameters from fits to training-plus-test denoted prime) In **D-G**, boxplot boxes denote 95% highest density intervals (HDIs), and lines denote 99% HDIs.

alone or training-plus-test (parameters from fits to training-plus-test denoted prime) were then related to the three transdiagnostic factors using Bayesian GLMs adjusted for age, sex, digit span and distancing status (see Methods 4.2.4). Results from the GLMs suggested there was limited evidence for an association between any Q-learning model parameter and scores on either the anxiety/depression or social withdrawal factors (Figure 2**D-E**). An exception was the negative association between the anxiety/depression factor score and the inverse temperature parameter β (from the dual learning rate model fit to training alone), such that unit increases in this score were associated with lower β values, indicating lower sensitivity to value differences between symbols in those scoring higher on this transdiagnostic symptom dimension (Figure 2**D**; posterior mean coefficient$= -0.08$). Still, the evidence for this was very weak, with the 95% HDI wide and including zero $(-0.22, 0.05)$. There was also little evidence of any association between β' and anxiety/depression from models fit to training-plus-test, nor of associations between learning rate parameters from any model and any of the three factor scores.

We did find evidence that increases in compulsive behaviour factor scores were associated with lower inverse temperature values (β and β'; Figure 2**F-G**). The strongest evidence for this came from the single learning rate models (Figure 2**F**) (β: 95% HDI $= (-0.33, -0.03)$); β': 95% HDI $= (-0.28, 0.004)$)); evidence from the dual learning rate model was consistent but weaker (Figure 2**G**) (β: 95% HDI $= (-0.26, 0.002)$); β': 95% HDI $= (-0.22, 0.05)$).

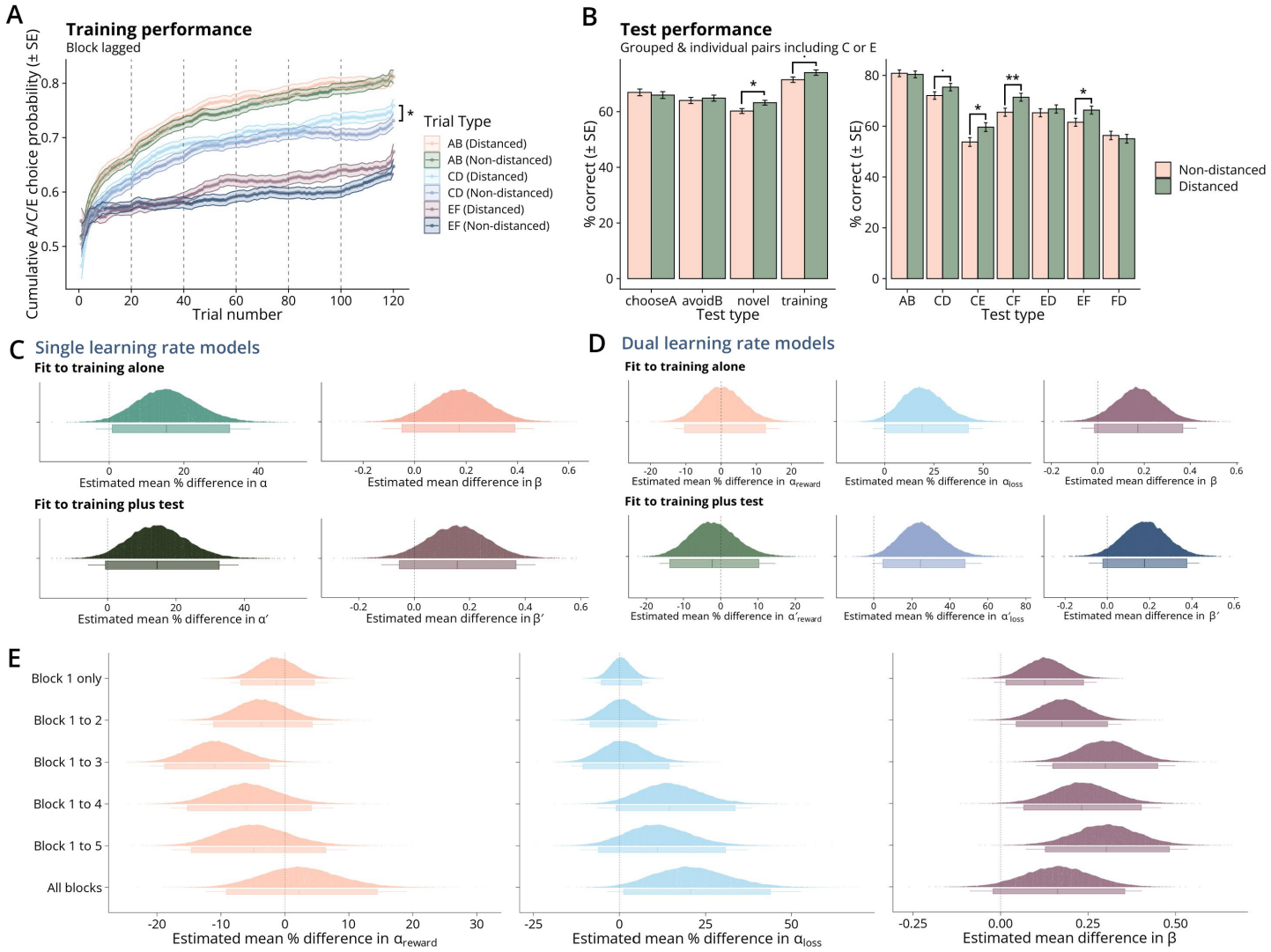

Figure 3: **Raw training and test phase performance and model-derived comparisons between distanced and non-distanced participants**. **A**. Model-free training performance (cumulative probability of choosing the higher-probability symbol A/C/E), by group and stimulus pair, lagged by twenty trials (i.e., block-lagged, as each pair is presented twenty times per block). **B**. Raw test phase performance (% correct) by test type, plus test phase performance on individual training pairs and novel pairs including symbols C or E. **C-D**. Coefficient posterior distributions from Bayesian GLMs (adjusted for age, sex, and digit span) reflecting the estimated percentage change in the learning rate parameter α (**C**) or $\alpha_{reward}/\alpha_{loss}$ (**D**), and the estimated mean change in the inverse temperature β, comparing distanced and non-distanced participants. Parameters were estimated from Q-learning models with single (**C**) or dual (**D**) learning rates, fit to training alone or training-plus-test (parameters from fits to training-plus-test denoted with prime). **E**. Parameter differences between distanced and non-distanced participants, estimated from dual learning rate models fit to increasing numbers of training blocks (sixty trials per block). In **D-G**, boxplot boxes denote 95% HDIs, and lines denote 99% HDIs.

2.4 Associations between learning parameters and cognitive distancing

2.4.1 Differences in raw training and test phase performance, and affect ratings

Next, we assessed whether cognitive distancing affected learning during the task. When comparing the tendency of participants to choose the symbol most likely to be correct in each pair, distanced participants appeared to perform marginally better than non-distanced individuals on more "difficult" pairs, where the symbols' reward probabilities were closer together (Figure 3**A-B**). Across training trials, distanced participants performed better on medium-difficulty "CD" trials (reward probabilities 0.7/0.3) in the final training block (% difference= 4.6; $t(931.6) = 2.29, p = 0.022$; Figure 3**A**). Distanced individuals also performed on average 5.1% better when tested on difficult novel pairs (those excluding the symbols most or least associated with reward): $t(932.65) = 2.48, p = 0.013$; Figure 3**B**).

Distancing also had subtle effects on self-reported affect throughout training. Linear mixed effects models were used to relate average affect ratings per block to group, block, and

their interaction (adjusting for age, sex, and digit span). All affect ratings decreased over time, but there was a block-by-distancing interaction for happiness and engagement (but not confidence): both happiness and engagement declined less per block in distanced participants than non-distanced participants, by an estimated 0.44 and 0.41 points (out of 100) per block (95% confidence intervals (CIs) $= (0.14, 0.73)$ and $(0.05, 0.78)$ for happiness and engagement respectively).

2.4.2 Results from models fit to training alone

To quantify potential differences in Q-learning model parameters between groups, we compared individual-level posterior means between distanced and non-distanced individuals using adjusted Bayesian GLMs.

There was weak but consistent evidence that distanced participants had higher inverse temperature values (β) (Figure 3**C-D**), suggesting less stochastic choices throughout training, with the mean difference estimated at 0.17 from both models (95% HDIs $= (-0.05, 0.39)$ and $(-0.01, 0.37)$ for single and dual learning rate models, respectively). Distanced participants also showed increased sensitivity to recent feedback, as indicated by estimated 15.3% higher α values (single learning rate model; 95% HDI for multiplier $= (1.01, 1.32)$) (Figure 3**C**). Notably, the dual learning rate model showed this increased sensitivity was specific to negative feedback trials (Figure 3**D**), affecting loss learning rates (estimated multiplier for $\alpha_{loss} = 1.19$, 95% HDI $= (1.0002, 1.42)$) but not positive learning rates (estimated multiplier for $\alpha_{reward} = 1.004$, 95% HDI $= (0.90, 1.13)$).

2.4.3 Results from models fit to training plus test data

In line with results from models fit to training alone, we found weak evidence for differences in inverse temperature parameters from models fit to training-plus-test (β'), with distanced participants having an estimated 0.16 (95% HDI $= (-0.05, 0.37)$) and 0.18 higher β' $(-0.02, 0.38)$ values from single and dual learning rate models respectively (Figure 3**C-D**). There was also evidence for differences in sensitivity to feedback from the single learning rate model additionally fit to test phase choices (estimated multiplier for α'=1.15, 95% HDI $= (0.99, 1.33)$; Figure 3**C**). Results from the dual learning rate model confirmed this specificity to negative feedback, with distanced participants estimated to have 24.6% higher loss learning rates at end of training (estimated multiplier for $\alpha'_{loss} = 1.25$; 95% HDI $= (1.05, 1.48)$; Figure 3**D**).

2.4.4 Temporal emergence of differences in learning parameters

Although we found consistent evidence of increases in loss learning rates in distanced individuals, particularly towards the end of training, there was little evidence of group differences in raw test performance on novel pairs that included the "B" symbol – the symbol with the lowest reward probability ($t(932.92) = 0.66, p = 0.51$; Figure 3**B**) – which would be expected in the context of a higher loss learning rate[31]. To investigate this, we fit dual learning rate models to increasing numbers of training blocks, and then compared the individual-level parameter values in the distanced and non-distanced groups with adjusted GLMs as before (Figure 3**E**). We found that group differences in loss learning rates emerged as the task progressed (with 95% HDIs for its multiplier between groups including 0 for all fits to fewer than all six blocks; Figure 3**E**). In contrast, there was strong evidence that distanced participants had consistently higher inverse temperature parameter values throughout training, with all 95% HDIs excluding 0 except for the final fit to all six blocks, and some evidence that they had 10.9% lower positive learning rates over the first three training blocks (estimated multiplier for $\alpha_{reward} = 0.89$; 95% HDI $= (0.81, 0.98)$; Figure 3**E**).

3 Discussion

Learning to predict rewards and avoid punishments is essential to adaptive behaviour. Disruptions in this fundamental process have been found across psychiatric disorders[10,38] and may be a target of psychological therapy[18]. In a large, broadly representative online sample, we found that a common psychotherapeutic strategy, cognitive distancing, enhanced performance in a reinforcement learning task. Our results indicate two computational mechanisms may underpin the therapeutic effects of cognitive distancing. Cognitive distancing facilitates integration and subsequent exploitation of previously-reinforced choices (via increased inverse temperature parameters); and, with time or practice, adaptively enhances the ability to update decisions following negative feedback (emerging increased loss learning rates).

Cognitive distancing is an antecedent-focused means of emotional regulation[39] which can reduce distress[40] and depressed thoughts[41] via disengagement from intense emotions in favour of a more experiential perspective. Similar cognitive reappraisal strategies reduce subjective experiences of both negative[42] and positive[43] affect, and

blunt physiological and neural responses to reward expectation[24,44]. Previously, cognitive distancing was specifically found to attenuate the contrast in encoding low- versus high-reward stimuli[27], regulated by top-down dorsolateral prefrontal input[29]. However, unlike our task, participants had learnt all relevant information about the task during pre-scan practice[27]. Reward signals are most relevant during learning, although they continue to be encoded by the dopaminergic midbrain after learning has taken place[45].

We found that cognitive distancing was consistently associated with heightened inverse temperatures, indicating clearer representations of differences in true (latent) values between choice options[46]. This may have enabled more deterministic choosing of the 'better' option in each pair, over those with less reward-certainty. Higher inverse temperature values can also be interpreted as a bias towards exploitation of current task knowledge, versus exploration of uncertain options[46]. Notably, a reduction in inverse temperature in patients with major depressive disorder has been found in previous studies using similar tasks and computational models[12,47]. Though our symptom-behaviour associations did not replicate this, our results indicate that cognitive distancing interventions may improve depressive symptoms by increasing exploitation of rewarding outcomes. This is a key hypothesis emerging from our results.

Distanced participants adaptively altered reward and loss learning throughout the task, possibly adjusting to changing task dynamics. Initially, higher inverse temperatures compared to non-distanced participants suggest they more quickly developed preferences for certain symbols. However, deterministic preferences may not be the best strategy for the entire task: initial impressions can be wrong, especially for the harder-to-distinguish stimuli. Notably, towards the end of the task, distanced participants became more sensitive to negative feedback, and showed weaker evidence for inverse temperature differences. At this stage, losses may be more informative: assuming preferences are largely correct for the easier pairs, negative feedback will be rarer, and primarily experienced when choosing between the harder-to-distinguish pairs. Previous work shows that participants adjust dual learning rates independently, increasing learning rate when one type of feedback becomes more informative[48]. By increasing loss sensitivity, and testing out preferences by exploring more, distanced participants may be able to discern values of harder-to-distinguish pairs more accurately by the end of training, enabling better performance when subsequently tested on these pairs.

A higher loss learning rate seems therapeutically counterintuitive, given several mental health disorders are marked by heightened punishment sensitivity[49,50]. Critically, however, aberrant loss learning may represent a "catastrophic response to perceived failure"[49,51]. In contrast, a high loss learning rate in distanced participants was accompanied by better test phase performance. It is notable that shifts in learning occurred late in the task: only once participants had practiced and developed their ability to self-distance were they particularly able to use it to their advantage. This suggests that persistent use of cognitive distancing may potentially drive symptom change by improving the ability to learn from negative experiences, and applying that learning to more adaptive behaviour. Clinically, this suggests the mechanism underlying distancing may be *more effective engagement with negative information*, rather than reduced engagement with negative information.

From our primary analyses, we found some evidence linking higher compulsive behaviour factor scores to lower inverse temperature parameter values. This is consistent with previous work reporting an association between compulsive behaviour and deficits in goal-directed control, characterised by the (in)ability to integrate information over time[21]. Contrary to our hypothesis, we found limited evidence of associations between learning parameters and transdiagnostic dimensions capturing anxiety/depression and social withdrawal symptoms. These results are at odds with a considerable body of literature reporting dysfunctional reinforcement learning across psychopathologies[10], especially depression[11]. However, most previous work has compared clinical populations to healthy controls, which may result in better detection of a true effect due to greater symptomatic delineation between groups. Alternatively, together with publication bias in the field[11], true effects may be smaller than those reported by small-scale clinical studies[52]. Additionally, 'loss' trials in our task lacked actual punishment, which is experienced differently to the absence of reward[12], and restricted our ability to detect associations between punishment processing and mental health[53].

Other limitations accompany this work. Firstly, our model space was limited to two simple Q-learning models, differing only in the use of a single vs. dual learning rate(s). While extending the model space to more complex models could have been informative, we employed this simpler strategy to aid interpretation of our results and comparison to previous literature. Secondly, we assessed symptoms through self-report questionnaires in the general population. Although our sample's variation in psychopathology was comparable to similar studies[21,22], it may not be compa-

rable to more severe clinical populations. Thirdly, we took a fully transdiagnostic approach to psychopathology[54], hoping to clarify mixed results previously reported by studies relating differences in reinforcement learning to diagnostic constructs. This precluded us from additionally deriving categorical psychiatric measures and investigating any potential taxonic associations with learning parameters.

In this study, by using a relatively simple learning paradigm with an established computational model, we were able to measure the effects of cognitive distancing on well-understood choice behaviour. This echoes clinical reports that distancing causes an adaptive shift in depressed people's processing of negative experiences[41]. This demonstrates the utility of computational approaches for back-translational studies to illuminate the mechanisms of existing treatments, in turn enabling improved augmentation and eventually personalisation of treatment for mental health disorders.

4 Methods

4.1 Online study

The protocol for the online study consisted of four components: a demographic questionnaire; the probabilistic selection task; a test of working memory (digit span); and a psychiatric questionnaire question battery. The tasks were written in JavaScript using the jsPsych library[55] (v6.2.0), and the study was hosted on a departmental server running Just Another Tool for Online Studies (JATOS)[56] (v3.3.1).

4.1.1 Ethics

This study was approved by the University of Cambridge Human Biology Research Ethics Committee (HBREC.2020.40) and jointly sponsored by the University of Cambridge and Cambridge University Hospitals NHS Foundation Trust (IRAS ID 289980).

4.1.2 Recruitment

Participants were recruited on Prolific[57] over three and a half weeks in April-May 2021. Data from the 2011 UK Census[35], and from the 2014 Adult Psychiatric Morbidity Survey[36] were used to calculate expected numbers by sex for individuals with and without a prior diagnosis of a psychiatric disorder across five age-groups: 18-24, 25-34, 35-44, 45-54, and 55+. We then used Prolific pre-screeners for age, sex, and self-reported history of a diagnosed mental health disorder to recruit batches of the target size (assuming total $n = 1000$) for each of the twenty groups (i.e., with/without prior mental health diagnosis, for each of the five age groups in males and females). We additionally restricted our sample to UK nationals. Participants were paid a fixed rate of £9 (approximately £6/hour on average), plus a performance bonus contingent on task performance (£1 if in the top 30% of points, or £2 if in the top 10%). Ultimately, 1,002 individuals completed most of the study and were paid for their participation, of whom 995 had complete data for all components.

4.1.3 Probabilistic selection task

The PST in the present study was broadly similar to its original conceptualisation[30,31], with minor adaptations, and consisted of a training and test phase (Figure 1**C**).

Before the start of the task, instructions were presented in written form as well as in an explanatory video that was optional to watch. To ensure that participants understood the task, they had to pass a multiple-choice quiz preceding the experiment. This consisted of six statements of which three were correct. While participants had an unlimited number of attempts, they could only move on to the task once all statements were marked correctly. Note that this quiz was only implemented after data from 100 participants had already been collected, meaning that these participants did not have to pass a quiz to start the experiment (see also 4.1.8).

The training phase consisted of six blocks of sixty trials. On each trial, participants were presented with one of three pairs of Japanese Hiragana characters, and were instructed to choose the left or right character via key press ('F' or 'J' key respectively). They were told that while there is no absolute right answer, some characters are more likely to result in a win, so they should try to pick the symbol they think is more likely to be correct. Participants received feedback after each trial ("Correct!" or "Incorrect."), and correct answers were rewarded with 25 points; if no choice was made within 4000ms, the trial ended.

Each stimulus pair was associated with different reward probabilities, for instance, in the 'AB' pair, the 'A' stimulus was associated with an 80% chance of winning, and the 'B' stimulus with a 20% chance of winning. The 'CD' pair of symbols was associated with a 70% or 30% chance of winning (respectively), and the 'EF' pair was associated with a 60% and 40% chance of winning (respectively). After each trial, participants were asked to indicate their current feelings (0-100) on one of three questions (analysis of affect questions will be reported separately): "How happy are you at this moment?", "How engaged are you feeling at this mo-

ment?", or "How confident are you feeling in your answers at this moment?". Presentation of questions was randomised in groups of three such that each question was asked twenty times per block and no question was asked more than two times in a row. At the end of each block, participants were also asked to rate how fatigued they felt compared to the beginning of the block, and received a reminder to try and win as many points as possible by choosing their preferred symbol in each pair. Each stimulus pair was presented exactly twenty times per block, with the order of the characters within each pair randomised on each trial. While the same six Hiragana characters were seen by all participants, the specific pairings and reward associations were randomised for each individual. Only twelve participants (1.2%) reported prior familiarity with Japanese Hiragana characters.

The test phase that followed consisted of a single block of sixty trials. The characters and reward probabilities associated with them remained the same, but participants were presented with all fifteen possible stimulus combinations in a random order (i.e., four of each pair). These included the three training pairs, four others including 'A' ("chooseA"), four others including 'B' ("avoidB"), and the remaining four pairs not including 'A' or 'B' ("novel"). Importantly, though participants were instructed to still choose the character that "feels correct" in every trial, no feedback was given, though affect questions were still asked after each trial to maintain the pace of the training blocks.

4.1.4 Digit span task

To control for working memory differences, a visual digit span task (adapted from here[58]) was administered following the PST. Participants were instructed to memorize sequences of numbers presented one at a time (1000ms per digit). After completing a practice trial (three numbers), the task began with a single digit, and sequences were extended by one digit when correctly reported. The task ended when two sequences of same length could not be reported, or if a sequence length of twenty-five digits was correctly recalled.

4.1.5 Questionnaire questions

Following a method described previously[22], we identified a subset of the 209 questionnaire questions used to first identify the three transdiagnostic dimensions of interest[21] which could accurately predict the three factor scores for all individuals (Figure 1**B,D**). Specifically, a multi-target lasso regression model was trained on original raw question responses (with the three factor scores as the responses), with differing numbers of question coefficients fixed to zero; five-fold cross-validation was then used to assess the predictive accuracy of these question subsets[22]. We found that 78 questions from eight different questionnaires[59–66] (Table S1), represented a good compromise between number of questions and predictive accuracy ($R^2 \geq 0.9$ for all three dimensions). Coefficient estimates from the multi-target lasso regression model (Figure S1) were then used to predict factor scores on each of the transdiagnostic symptom dimensions for all individuals in our sample, based on their answers to the included 78 questions.

In addition, full questionnaires were included for anhedonia (dimensional anhedonia rating scale (DARS)[67]), schizotypy (schizotypal personality questionnaire - brief revised (updated) (SPQ-BRU)[68]), body perception (body perception questionnaire - very short form (BPQ-VSF)[69]), and fatigue impact (modified fatigue impact scale (MFIS)[70]). Questionnaire order was randomised across participants, and the individual questions within each questionnaire were asked one at a time to prevent straightlining.

4.1.6 Cognitive distancing manipulation

Half of the participants were randomly allocated to the cognitive distancing manipulation. The concept of cognitive distancing was introduced in a short video, which explained to participants that they should told to try to "take a step back" from their immediate emotional reactions to positive or negative feedback, perhaps by thinking of themselves as an external observer, viewing the events from a distance. The video had to be watched to proceed to the task, with the following written instructions provided if the participant had issues loading the video.

> "When you're doing the task, you might feel various emotions – for example, irritation, engagement, or happiness. But throughout the task, we would like you to practice a mental strategy called **self-distancing**.
>
> Self-distancing is the ability to take mental 'step back' from your immediate reactions to events, and view these events from a broader, calmer, and less emotional perspective.
>
> One way of practising this is to imagine yourself as an external observer, watching yourself perform the task from a distance, and seeing the results of each of your decisions in the task.
>
> You'll still learn which symbols win you more points than others, and you should still try to win

as many points as possible. But whenever you feel irritated, happy, or any other emotion, even if it feels minor, try to distance yourself from your immediate reaction, by taking a step back from how you are feeling. We understand that this will be tricky, and so if you are unable to distance yourself from your emotional reaction on a particular trial, that is completely fine! Simply honestly report how you're feeling when we ask, and then try again to distance yourself next time."

After the instructions, distanced participants began the PST as detailed above. The task itself was identical to that taken by non-distanced participants, except that they received an additional reminder to "Distance yourself..." with the fixation cross that preceded each training trial (Figure 1**C**). In the test phase, this reminder was omitted, as there was no feedback.

4.1.7 Attention checks and exclusion criteria

Participants were excluded if they either self-reported a diagnosis of a neurological disorder ($n = 18$), or English proficiency below a B2 level (good command or working knowledge; $n = 3$). We also included two task-based exclusion criteria: a digit span of 0 ($n = 5$), or $> 95\%$ preference for a single key (which no one met). Lastly, based on recent recommendations[71], two harder catch questions were included in the questionnaires (e.g., "In the past week, I would (have) avoided... Eating mouldy food"; expected answer 'Usually'), in addition to two standard catch questions (e.g., "Please answer 'Strongly Agree' to this question."). Participants were excluded if they got one of the standard questions wrong ($n = 16$), or both harder questions wrong ($n = 20$). In total, 60 participants of the 995 met at least one of these criteria and were excluded from all analyses (Table 1).

4.1.8 Deviations from preregistration

The demographic exclusion criteria (low English proficiency and/or any neurological disorder), plus the catch questions were included in our preregistration, while the task-based exclusion criteria (key-mashing and digit spans of 0) were not included. Instead, we had initially opted to exclude poor performers through an accuracy criterion ($\geq$60% correct on the AB pair). However, after running the first batch of 100 participants, we found that over 35% had been excluded, largely due to the accuracy criterion. We consulted with experts in the field who had run similar studies online via Twitter who advised us that excluding based on accuracy may unfairly bias our sample as mental health symptoms are commonly associated with cognitive changes[72] that could affect accuracy and suggested that we remove this exclusion criterion and replace it with the task-based exclusion criteria we used for all subsequent participants. In addition, as aforementioned, we were advised to add a multiple-choice quiz on the task instructions, which had to be answered correctly to begin the task.

4.2 Computational modelling

4.2.1 Models fit to training data alone

Model-free reinforcement learning in the PST is commonly modelled using Q-learning models[25,31,32]. In these models, the weight or Q-value $Q_t(s_t, a_t)$ for a given action a in state s at time t is an estimate of the state-action value, which can in turn be understood as an estimate of the expected sum of future rewards, conditional on that action at time t[25]. Q-values are updated trial-by-trial based on prediction errors (δ_t), where α is the learning rate. In bandit tasks such as the PST, δ_t is simply the difference between the Q-value for the chosen action (i.e., picking a certain symbol in pair) and the observed reward r_t (1), as selecting a certain action is assumed not to affect the transition to future states[25].

$$\delta_t = r_t - Q_t\left(s_t, a_t\right) \tag{1}$$

The two established models of interest in our case were a standard Q-learning model with a single learning rate α (2), and an extended Q-learning model with dual learning rates, α_{reward} and α_{loss} (3)[31].

$$Q_{t+1}\left(s_t, a_t\right) = Q_t\left(s_t, a_t\right) + \alpha\delta_t \tag{2}$$

$$Q_{t+1}(s_t, a_t) = \begin{cases} Q_t(s_t, a_t) + \alpha_{reward}\delta_t & \text{if } \delta_t \geq 0, \text{ or} \\ Q_t(s_t, a_t) + \alpha_{loss}\delta_t & \text{if } \delta_t < 0 \end{cases} \tag{3}$$

In the single learning rate model, α can be interpreted as the sensitivity to recent feedback, with higher values indicating that Q-values are being rapidly updated in response to both positive and negative feedback. In the dual learning rate model, Q-values are assumed to be updated separately depending on whether δ_t is negative, which in turn occurs only when feedback is negative (i.e., $r_t = 0$). Higher α_{loss} can hence be interpreted as an increased sensitivity to recent negative feedback (and so reduced integration over trials), while higher α_{reward} values suggest increased sensitivity to recent positive feedback[31].

In both models, the Q-values can be converted to probabilities (i.e., of choosing one symbol over another) using a softmax logistic function as follows (4)

$$P_t(s_t, a_t) = \frac{1}{1 + e^{-\beta[Q_t(s_t,a_t) - Q_t(s_t,\bar{a}_t)]}}, \quad (4)$$

where $\bar{a}_t$ is the alternative (avoided) choice in the pair and β is an inverse temperature parameter, lower values of which indicate more deterministic choices. In all models, the choice of one symbol over the other was assumed to follow a Bernoulli logistic distribution with the chance-of-success parameter equal to $logit[P_t(s_t, a_t)] = \beta[Q_t(s_t, a_t) - Q_t(s_t, \bar{a}_t)]$, as the absence of other symbols from the choices on each trial was assumed not to affect the probability of choosing one over another[73].

4.2.2 Models fit to training plus test data

Both the single and dual learning rate models can be extended to include test phase trials[31]. However, in the absence of feedback, Q-values are assumed to be fixed at the end of training; as such, the learning rate(s) and inverse temperature parameters can be thought of as those at the end of training which best fit the subsequent test phase choices[31]. This means that the probability of choosing one option over any other in the test phase is given by (5)

$$P_t^{test}(s_t, a_t) = \frac{1}{1 + e^{-\beta'[Q_{final}(s_t,a_t) - Q_{final}(s_t,\bar{a}_t)]}}, \quad (5)$$

where β' and Q_{final} correspond to the inverse temperature parameter and Q-value at the end of training respectively.

4.2.3 Model fitting, checks, and comparisons

Models were fit in a hierarchical Bayesian manner using CmdStan[74], via its R interface cmdstanr, enabling simultaneous estimation of group-level and individual-level parameters[75]. Weakly informative Gaussian group-level and individual-level priors were assumed for all learning parameters, and models were fit separately for distanced and non-distanced participants. Stan code for the training data models was taken from the hBayesDM package[34] repository; these training data models were then adapted to additionally incorporate test phase trials for the training plus test data analyses.

In all cases, models were fit using Markov-Chain Monte Carlo (MCMC), with 4 parallel chains and 4,000 warm-up draws plus 20,000 sampling draws per chain for all models. Numerical diagnostics, namely effective sample size (ESS) and split R-hat[76] were used to assess chain mixing and convergence. Individuals' parameters were omitted from subsequent analyses if they had either split R-hat ≥ 1.1 or bulk ESS < 100 for any parameter (applied to no more than two individuals per fit). Visual MCMC diagnostics (traces and rank histograms) were also checked for model convergence. Models were then compared in terms of out-of-sample predictive accuracy, using numerical metrics (the expected log posterior density and the leave-one-out information criterion[37]). We also conducted posterior predictive checks, generating predictions from posterior distributions, and comparing these to observed choices for all individuals and trials (Figure S2-S5). Finally, the ability of the models to recover model parameters was assessed by simulating data for known parameter values, fitting each model to these data, and comparing the fitted parameter values to those used to simulate the data (Figure S6-S7).

4.2.4 Outcome generalised linear models and exploratory analyses

To assess the evidence for associations between model parameters and the outcomes of interest, namely the transdiagnostic factor scores and distancing, we used Bayesian GLMs. Specifically, due to their positively skewed distributions (Figure 2**B-C**), the association between the learning rates and the outcomes of interest was assessed using gamma family GLMs with log link functions, while the association between the inverse temperatures and outcomes was assessed using Gaussian family GLMs with identity link functions (i.e., linear regression). All models were adjusted for age, digit span, and sex (male or female; imputed as birth-assigned for non-binary individuals due to low numbers), plus distancing status in models relating transdiagnostic factor scores to Q-learning model parameters. The outcome GLMs were fit using CmdStan[74], with 2,000 warm-up draws and 10,000 sampling draws per chain, using Stan models and priors from the rstanarm R package[77]. Lastly, an exploratory analysis was run to assess at what stage differences in training performance emerged between distanced and non-distanced participants. To do so, the dual learning rate Q-learning model was fit separately in each group as before to trials from increasing numbers of training blocks (i.e., block one, block one to two, block one to three, etc.), and Bayesian GLMs run once again to assess the evidence for group differences in learning parameters at each stage.

Open code and data

All data, task, and analysis code are shared openly on GitHub to encourage replication and extension of our findings. Results can be interactively reproduced and extended step-by-step within a Docker container via Jupyter notebooks — see the README for more details. Our study preregistration can be found on OSF.

Acronyms

AES	Apathy Evaluation Scale
AUDIT	Alcohol Use Disorders Identification Test
BIS	Barratt Impulsiveness Scale
BPQ-VSF	Body Perception Questionnaire - Very Short Form
CBT	Cognitive Behavioural Therapy
CI	Confidence Interval
DARS	Dimensional Anhedonia Rating Scale
EAT	Eeating Attitudes Test
ESS	Effective Sample Size
ELPD	Expected Log Posterior Density
EV	Expected Value
IQR	Interquartile Range
GLM	Generalised Linear Model
HDI	Highest Density Interval
JATOS	Just Another Tool for Online Studies
LOOIC	Leave-One-Out Information Criterion
LSAS	Liebowitz Social Anxiety Scale
MFIS	Modified Fatigue Impact Scale
MRC	Medical Research Council
MCMC	Markov-Chain Monte Carlo
PE	Prediction Error
PST	Probabilistic Selection Task
OCIR	Obsessive Compulsive Inventory Revised
SD	Standard Deviation
SDS	Self-rating Depression Scale
SE	Standard Error
STAI	State-Trait Anxiety Scale
SPQ-BRU	Schizotypal Personality Questionnaire - Brief Revised (Updated)

References

1. Sloan, E. *et al.* Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review* **57,** 141–163 (2017).
2. Papa, A., Boland, M. & Sewell, M. T. in *Cognitive Behavior Therapy: Core Principles for Practice* (eds O'Donohue, W. & Fisher, J. E.) 273–323 (John Wiley & Sons, Inc., 2012).
3. Neacsiu, A. D., Eberle, J. W., Kramer, R., Wiesmann, T. & Linehan, M. M. Dialectical behavior therapy skills for transdiagnostic emotion dysregulation: A pilot randomized controlled trial. *Behaviour Research and Therapy* **59,** 40–51 (2014).
4. Hamidian, S., Omidi, A., Mousavinasab, S. M. & Naziri, G. The Effect of Combining Mindfulness-Based Cognitive Therapy with Pharmacotherapy on Depression and Emotion Regulation of Patients with Dysthymia: A Clinical Study. *Iranian Journal of Psychiatry* **11,** 166 (2016).
5. Siegle, G. J., Carter, C. S. & Thase, M. E. Use of fMRI to predict recovery from unipolar depression with cognitive behavior therapy. *American Journal of Psychiatry* **163,** 735–738 (2006).
6. Sheppes, G., Suri, G. & Gross, J. J. Emotion Regulation and Psychopathology. *Annual Review of Clinical Psychology* **11,** 379–405 (2015).
7. Etkin, A., Büchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nature Reviews Neuroscience* **16,** 693–700 (2015).
8. Huys, Q. J., Maia, T. V. & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience* **19,** 404–413 (2016).
9. Moutoussis, M., Shahar, N., Hauser, T. U. & Dolan, R. J. Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. *Computational Psychiatry* **2,** 50 (2018).
10. Maia, T. V. & Frank, M. J. From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience* **14,** 154–162 (2011).
11. Halahakoon, D. C. *et al.* Reward-Processing Behavior in Depressed Participants Relative to Healthy Volunteers: A Systematic Review and Meta-analysis. *JAMA Psychiatry* **77** (2020).
12. Huys, Q. J., Pizzagalli, D. A., Bogdan, R. & Dayan, P. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders* **3,** 1–16 (2013).
13. Scholl, J. *et al.* Beyond negative valence: 2-week administration of a serotonergic antidepressant enhances both reward and effort learning signals. *PLOS Biology* **15,** e2000756 (2017).
14. Hales, C. A., Houghton, C. J. & Robinson, E. S. Behavioural and computational methods reveal differential effects for how delayed and rapid onset antidepressants effect decision making in rats. *European Neuropsychopharmacology* **27,** 1268–1280 (2017).
15. Insel, C. *et al.* Antipsychotic dose modulates behavioral and neural responses to feedback during reinforcement learning in schizophrenia. *Cognitive, Affective and Behavioral Neuroscience* **14,** 189–201 (2014).
16. Volman, I. *et al.* Lithium modulates striatal reward anticipation and prediction error coding in healthy volunteers. *Neuropsychopharmacology* **46,** 386–393 (2020).
17. Pizzagalli, D. A., Jahn, A. L. & O'Shea, J. P. Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological Psychiatry* **57,** 319–327 (2005).
18. Brown, V. M. *et al.* Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA Psychiatry* **78,** 1113–1122 (2021).
19. Steele, J. D., Kumar, P. & Ebmeier, K. P. Blunted response to feedback information in depressive illness. *Brain* **130,** 2367–2374 (2007).

20. Chase, H. W. *et al.* Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychological Medicine* **40,** 433–440 (2010).

21. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5** (ed Frank, M. J.) e11305 (2016).

22. Wise, T. & Dolan, R. J. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications* **11** (2020).

23. Dalgleish, T., Black, M., Johnston, D. & Bevan, A. Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology* **88,** 179 (2020).

24. Delgado, M. R., Gillis, M. M. & Phelps, E. A. Regulating the expectation of reward via cognitive strategies. *Nature Neuroscience* **11,** 880 (2008).

25. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* ISBN: 978-0-262-03924-6 (MIT Press, Cambridge, 1998).

26. Schultz, W. Behavioral dopamine signals. *Trends in Neurosciences* **30,** 203–210 (2007).

27. Staudinger, M. R., Erk, S., Abler, B. & Walter, H. Cognitive reappraisal modulates expected value and prediction error encoding in the ventral striatum. *NeuroImage* **47,** 713–721 (2009).

28. Goldin, P. R., McRae, K., Ramel, W. & Gross, J. J. The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion. *Biological Psychiatry* **63,** 577–586 (2008).

29. Staudinger, M. R., Erk, S. & Walter, H. Dorsolateral Prefrontal Cortex Modulates Striatal Reward Encoding during Reappraisal of Reward Anticipation. *Cerebral Cortex* **21,** 2578–2588 (2011).

30. Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306,** 1940–1943 (2004).

31. Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T. & Hutchison, K. E. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 16311–16316 (2007).

32. Watkins, C. J. C. H. & Dayan, P. Q-learning. *Machine Learning* **8,** 279–292 (1992).

33. Gelman, A. *et al. Bayesian Data Analysis* 3rd ed. ISBN: 978-1-4398-9820-8 (CRC Press, New York, NY, 2013).

34. Ahn, W.-Y., Haines, N. & Zhang, L. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry* **1,** 24 (2017).

35. Office of National Statistics. *2011 Census - Sex by age by IMD2004 by ethnic group* [Online; accessed 2021-03-17]. 2016.

36. Stansfeld, S. *et al.* in *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014* 37–68 (NHS Digital, 2016).

37. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27,** 1413–1432 (2016).

38. Lee, D. Decision Making: from Neuroscience to Psychiatry. *Neuron* **78,** 233 (2013).

39. Gross, J. J. Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology* **74,** 224–237 (1998).

40. Mennin, D. S., Ellard, K. K., Fresco, D. M. & Gross, J. J. United we stand: Emphasizing commonalities across cognitive-behavioral therapies. *Behavior Therapy* **44,** 234–248 (2013).

41. Kross, E., Gard, D., Deldin, P., Clifton, J. & Ayduk, O. "Asking why" from a distance: Its cognitive and emotional consequences for people with major depressive disorder. *Journal of Abnormal Psychology* **121,** 559 (2012).

42. Ochsner, K. N. *et al.* For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage* **23,** 483–499 (2004).

43. Sang, H. K. & Hamann, S. Neural correlates of positive and negative emotion regulation. *Journal of Cognitive Neuroscience* **19,** 776–798 (2007).

44. Sokol-Hessner, P. *et al.* Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 5035–5040 (2009).

45. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47,** 129–141 (2005).

46. Pedersen, M. L. & Frank, M. J. Simultaneous Hierarchical Bayesian Parameter Estimation for Reinforcement Learning and Drift Diffusion Models: a Tutorial and Links to Neural Data. *Computational Brain and Behavior* **3,** 458–471 (2020).

47. Pike, A. C. & Robinson, O. J. Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry* (2022).

48. Pulcu, E. & Browning, M. Affective bias as a rational response to the statistics of rewards and punishments. *eLife* **6** (ed Frank, M. J.) e27879 (2017).

49. Elliott, R., Sahakian, B. J., Herrod, J. J., Robbins, T. W. & Paykel, E. S. Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis specific impairment. *Journal of Neurology, Neurosurgery & Psychiatry* **63,** 74–82 (1997).

50. Jappe, L. M. *et al.* Heightened sensitivity to reward and punishment in anorexia nervosa. *International Journal of Eating Disorders* **44,** 317–324 (2011).

51. Roiser, J. P. & Sahakian, B. J. Hot and cold cognition in depression. *CNS Spectrums* **18,** 139–149 (2013).

52. Gelman, A. & Carlin, J. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science* **9,** 641–651 (2014).

53. Lawson, R. P. *et al.* Disrupted habenula function in major depression. *Mol Psychiatry* (2016).

54. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *The American Journal of Psychiatry* **167,** 748–751 (2010).

55. De Leeuw, J. R. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* **47,** 1–12 (2015).

56. Lange, K., Kühn, S. & Filevich, E. "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE* **10,** e0130834 (2015).

57. Palan, S. & Schitter, C. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* **17,** 22–27 (2018).

58. Luthra, M. & Todd, P. M. *Role of Working Memory on Strategy Use in the Probability Learning Task* in (Montreal, Canada, 2019), 721–728.

59. Foa, E. B. *et al.* The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological Assessment* **14,** 485–496 (2002).

60. Garner, D. M. & Garfinkel, P. E. The Eating Attitudes Test: an index of the symptoms of anorexia nervosa. *Psychological Medicine* **9,** 273–279 (1979).

61. Marin, R. S., Biedrzycki, R. C. & Firinciogullari, S. Reliability and validity of the apathy evaluation scale. *Psychiatry Research* **38,** 143–162 (1991).

62. Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R. & Grant, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. en. *Addiction* **88,** 791–804 (1993).

63. Zung, W. W. K. A Self-Rating Depression Scale. *Archives of General Psychiatry* **12,** 63–70 (1965).

64. Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R. & Jacobs, G. A. *Manual for the State-Trait Anxiety Inventory* (Consulting Psychologists Press, Palo Alto, CA, 1983).

65. Patton, J. H., Stanford, M. S. & Barratt, E. S. Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology* **51,** 768–774 (1995).

66. Liebowitz, M. R. Social Phobia. *Anxiety* **22,** 141–173 (1987).

67. Rizvi, S. J. *et al.* Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression. *Psychiatry Research* **229,** 109–119 (2015).

68. Davidson, C. A., Hoffman, L. & Spaulding, W. D. Schizotypal personality questionnaire - brief revised (updated): An update of norms, factor structure, and item content in a large non-clinical young adult sample. *Psychiatry Research* **238,** 345–355 (2016).

69. Cabrera, A. *et al.* Assessing body awareness and autonomic reactivity: Factor structure and psychometric properties of the Body Perception Questionnaire-Short Form (BPQ-SF). *International Journal of Methods in Psychiatric Research* **27** (2018).

70. Fisk, J. D. *et al.* Measuring the functional impact of fatigue: Initial validation of the fatigue impact scale. *Clinical Infectious Diseases* **18,** S79–S83 (1994).

71. Zorowitz, S., Niv, Y. & Bennett, D. Inattentive responding can induce spurious associations between task behavior and symptom measures. *PsyArXiv* (2021).

72. Millan, M. J. *et al.* Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. en. *Nature Reviews Drug Discovery* **11,** 141–168 (2012).

73. Luce, R. D. *Individual Choice Behavior* (John Wiley, Oxford, England, 1959).

74. Stan Development Team. *Stan Modelling Language Users Guide and Reference Manual* (2021).

75. Ahn, W. Y., Krawitz, A., Kim, W., Busemeyer, J. R. & Brown, J. W. A Model-Based fMRI Analysis With Hierarchical Bayesian Parameter Estimation. *Journal of Neuroscience, Psychology, and Economics* **4,** 95–110 (2011).

76. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Burkner, P. C. Rank-Normalization, Folding, and Localization: An Improved R-hat for Assessing Convergence of MCMC. *Bayesian Analysis* **16,** 667–718 (2021).

77. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. *rstanarm: Bayesian applied regression modeling via Stan.* (2020).

Acknowledgments

The authors would like to thank Dr. Becky Gilbert for her assistance with jsPsych programming.

Funding declaration

This study was funded by an AXA Research Fund Fellowship awarded to C.L.N. (G102329) and the Medical Research Council (SUAG/077, SUAG/043 G101400) and partly supported by the National Institute for Health Research Cambridge Biomedical Research Centre. R.P.L. is supported by a Royal Society Wellcome Trust Henry Dale Fellowship (206691) and is a Lister Institute Prize Fellow.

Author contributions

Conceptualization, C.L.N.; Methodology, C.L.N., Q.D., S.M., T.S., R.P.L., D.A.P., T.D.; Investigation, Q.D., S.M., C.L.N.; Project Administration, Q.D., S.M., C.L.N.; Writing – Original Draft, Q.D., S.M., C.L.N.; Formal Analysis, Q.D., S.M.; Software, Q.D.; Writing – Review & Editing, C.H., D.A.P., T.D., T.S., R.P.L.; Funding Acquisition, C.L.N., T.D., Supervision: C.L.N., T.D.

Conflicts of interest

Over the past 3 years, D.A.P. has received consulting fees from Albrights Stonebridge Group, Boehringer Ingelheim, Compass Pathway, Concert Pharmaceuticals, Engrail Therapeutics, Neumora Therapeutics (former BlackThorn Therapeutics), Neurocrine Biosciences, Neuroscience Software, Otsuka Pharmaceuticals, and Takeda Pharmaceuticals; one honorarium from Alkermes, and research funding from NIMH, Dana Foundation, Brain and Behavior Research Foundation, Millennium Pharmaceuticals. In addition, he has received stock options from BlackThorn Therapeutics and Compass Pathways. All other authors report no financial relationships with commercial interest.

Supplementary material

Questionnaire subsets used to predict transdiagnostic factors

Table S1: Subset of 78 questionnaire questions used to predict transdiagnostic factors. For the Liebowitz social anxiety scale (LSAS), each question was asked twice to quantify both 'fear' and 'avoidance', and the average of the two taken.

Questionnaire	Question numbers (**reversed scoring**)	Scale
Obsessive compulsive inventory revised (OCIR)	2-9, 11-16, 18	0 to 4
Eeating attitudes test (EAT)	1-4, 6-8, 10-12, 14, 15, 18, 20-24	0 to 5
Apathy evaluation scale (AES)	**17**, **18**	0 to 3
Alcohol use disorders identification test (AUDIT)	3	0 to 4
Self-rating depression scale (SDS)	1, **11**, **12**, **14**, **16-18**, **20**	1 to 4
State-trait anxiety scale (STAI)	**1**, **3**, 4, 5, 8, **10**, 12, **13**, 15, **16**, 17, **19**, 20	1 to 4
Barratt impulsiveness scale (BIS)	6, **9**, **10**, **13**, 14	1 to 4
Liebowitz social anxiety scale (LSAS)	2, 6-12, 15, 16, 18-21, 23, 24	0 to 3

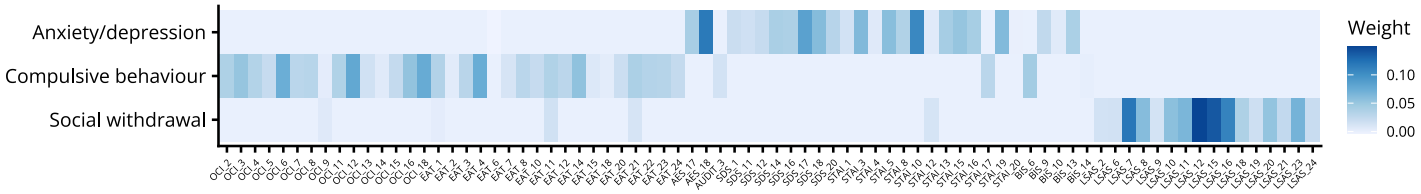

Figure S1: Five-fold cross-validated multi-target lasso regression coefficient weights for each of the included 78 questions, used to predict the three transdiagnostic symptom dimensions.

Model validation and checks

Posterior predictive checks

To assess the predictive validity of each model, choices were sampled from the posterior distribution for each of the 80,000 sampling iterations (20,000 draws by four chains), for each individual and trial. Choices (0 or 1 if the most likely correct symbol A/C/E was chosen for a given pair) were then averaged over posterior draws (i.e., $\frac{\sum_1^n choice}{n}$, where choice is 1 or 0, and n is the total number of draws) for each trial/individual to obtain the model prediction. These model-derived predictions were then visually and numerically compared to the observed data as seen below for each of the models/groups.

Models fit to training data alone

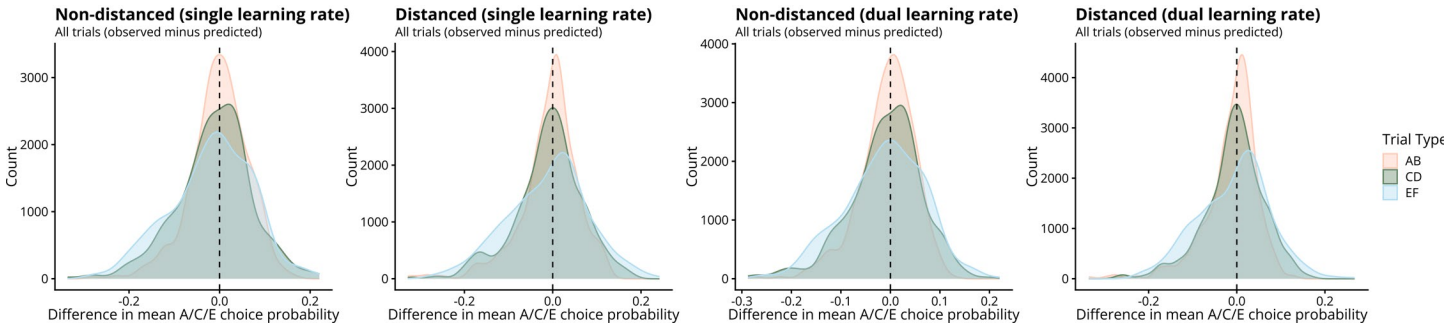

Figure S2: Difference in mean A/C/E choice probability (observed minus predicted) averaged across all trials for each individual, by symbol pair.

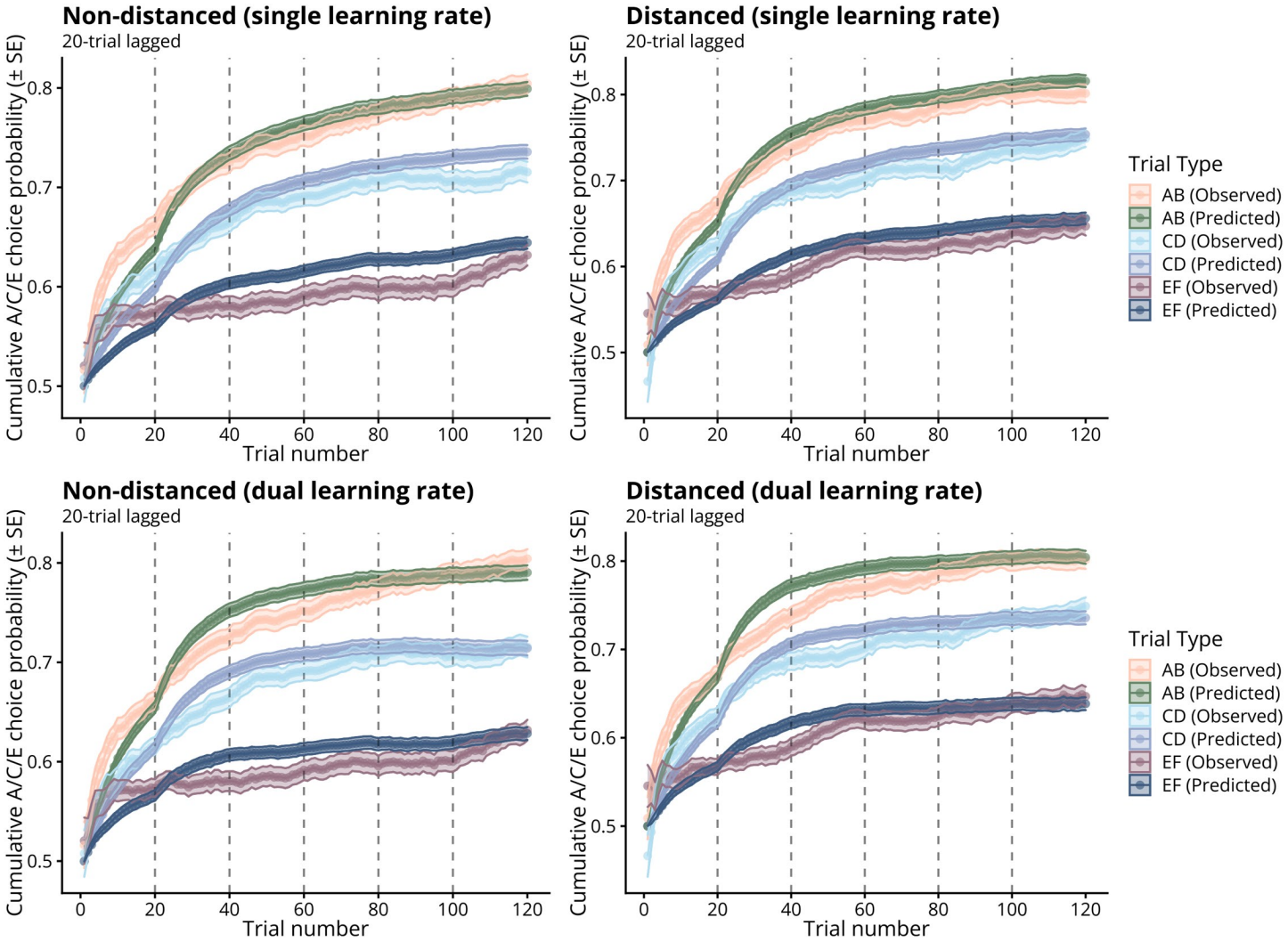

Figure S3: Comparison between observed and predicted mean (± standard error (SE)) cumulative (twenty-trial lagged) probability of choosing the most likely correct option in each pair (symbol A, C, or E in pairs AB, CD, or EF respectively), averaged across all individuals.

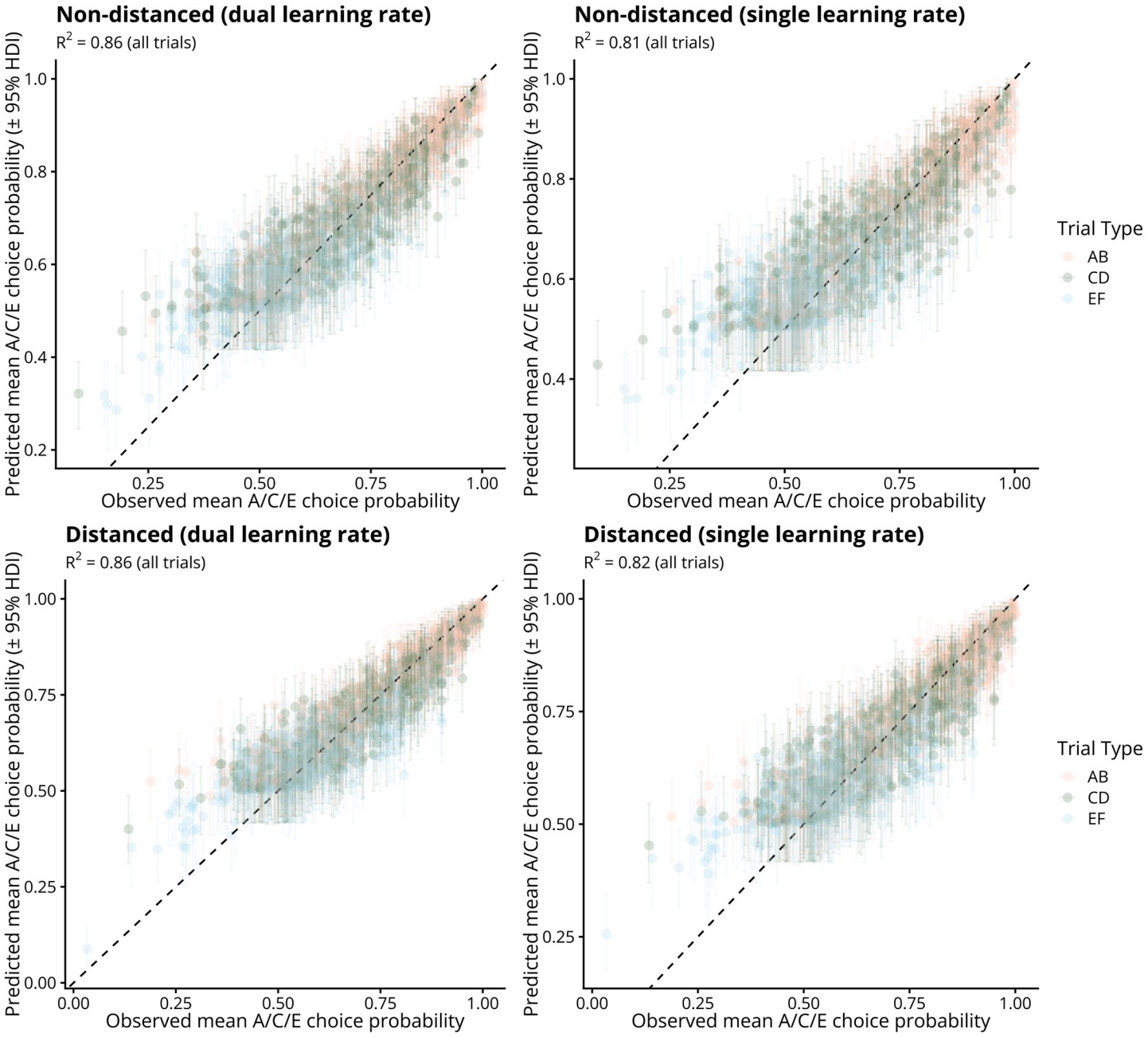

Figure S4: Posterior mean A/C/E (predicted) choice probability (± 95% HDI) for each individual on each pair type, plotted against the observed values after all six training blocks.

Models fit to training plus test data

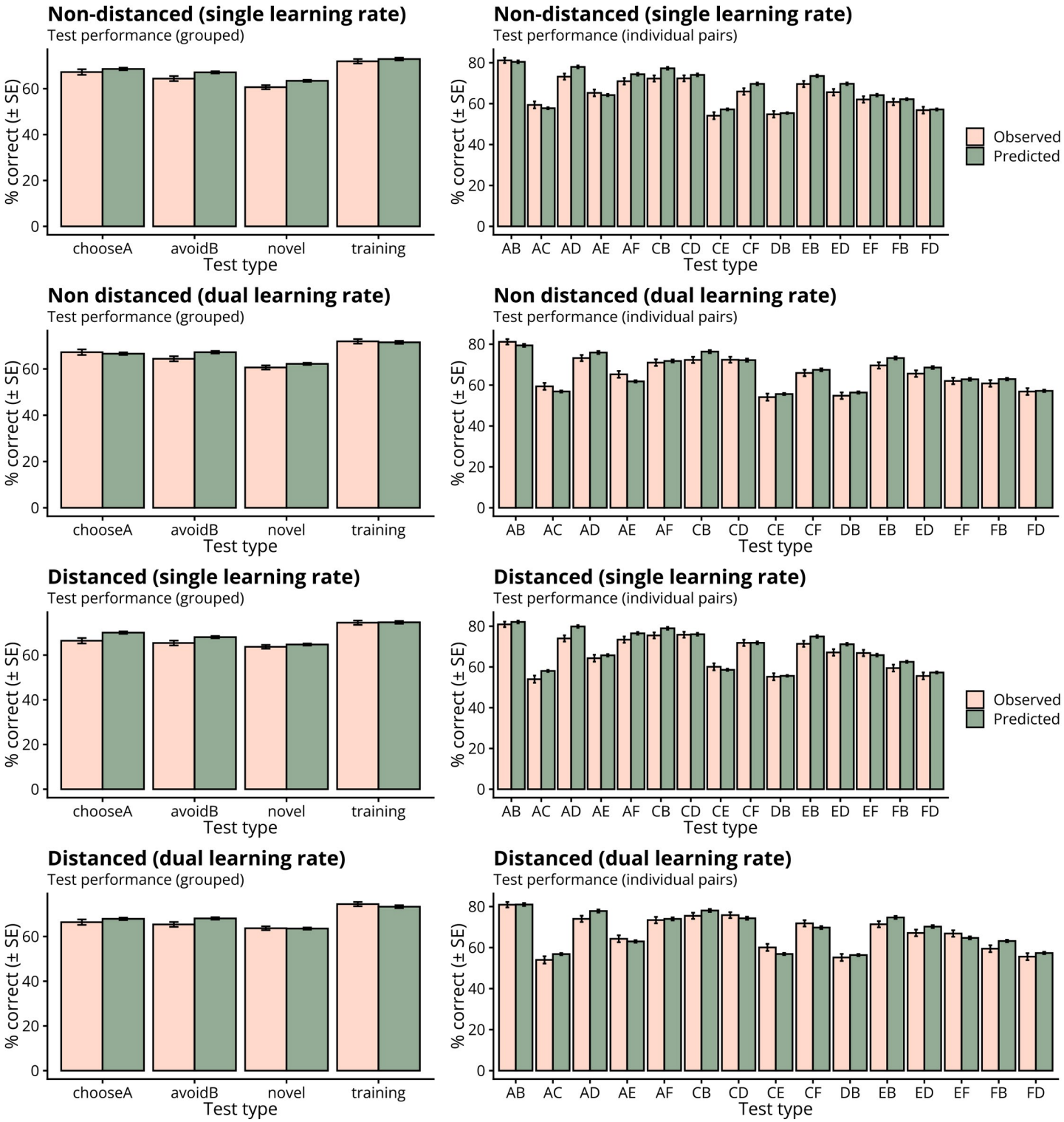

Figure S5: Predicted and observed test choices (mean % accuracy ± SE), averaged across all participants and test trials, by test trial group and for each individual test phase stimulus.

Parameter recovery

To assess whether the parameter values obtained from our models were meaningful, we simulated task data for a randomly sampled (known) set of parameter values ($n = 500$ simulated individuals per model). Learning rate parameters (α) were

drawn from a $Gamma(k = 2, \Theta = 0.1)$ distribution (i.e., positively skewed, bounded by 0), while inverse temperature parameters were drawn from a $Gaussian(\mu = 3, \sigma = 1)$ distribution. Each of the four models (i.e., single and dual learning rate models fit to training alone, or training plus test) were then fit to the simulated data, and the "recovered" parameter values compared to the known parameter values for each individual.

Models fit to training data alone

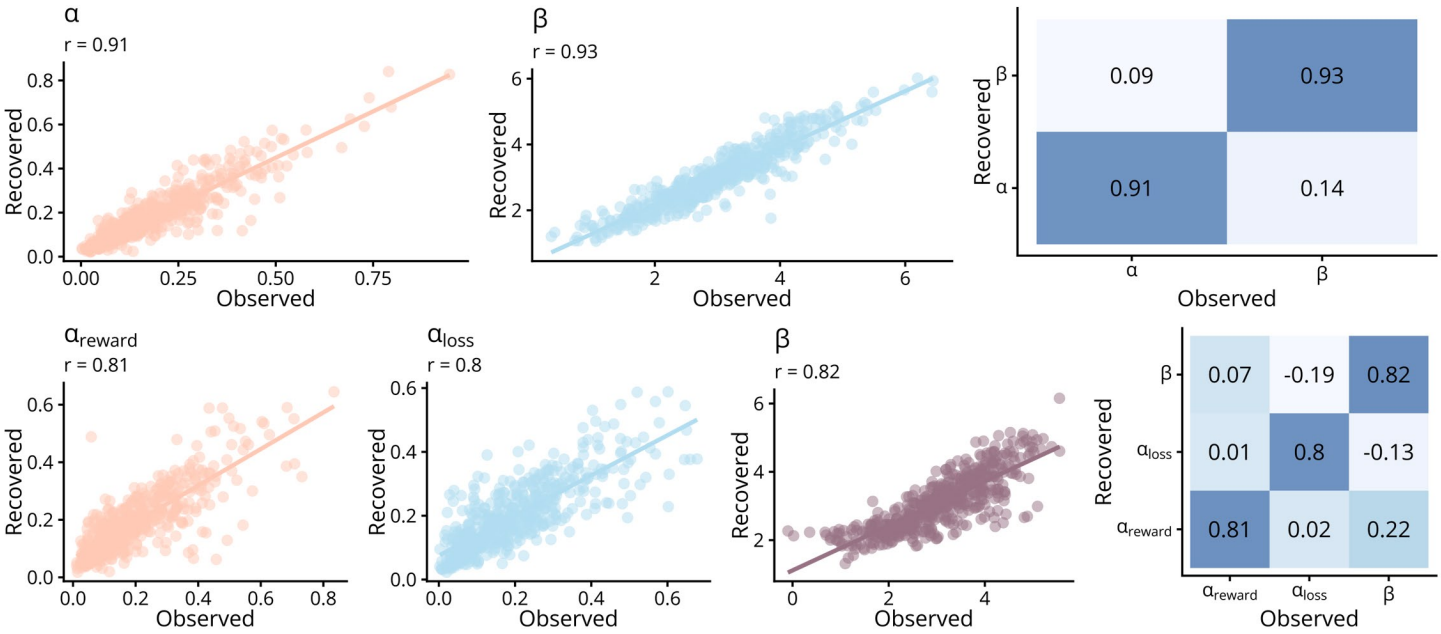

Figure S6: Plots of recovered parameter values against the observed values used to generate the task data, plus correlation plots for the single (*top*) and dual (*bottom*) learning rate models fit to training alone.

Models fit to training plus test data

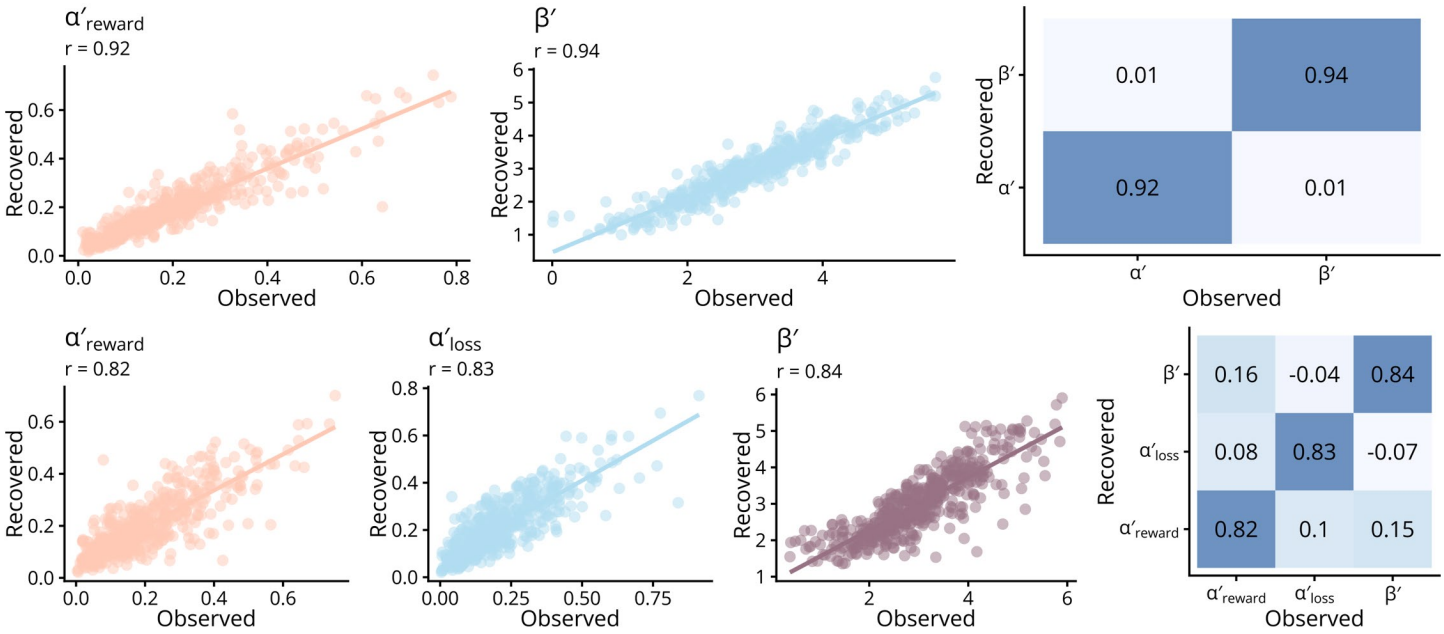

Figure S7: Plots of recovered parameter values against the observed values used to generate the task data, plus correlation plots for the single (*top*) and dual (*bottom*) learning rate models fit to training plus test choices.