

RESEARCH ARTICLE

Dual process impairments in reinforcement learning and working memory systems underlie learning deficits in physiological anxiety

Jennifer D. Senta^{1*}, Sonia J. Bishop^{1,2,3‡}, Anne G.E. Collins^{1,4‡}

1 Helen Wills Neuroscience Institute, University of California, Berkeley, California, United States of America, **2** School of Psychology, Trinity College Dublin, Dublin, Ireland, **3** Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland, **4** Department of Psychology, University of California, Berkeley, California, United States of America

‡ Joint senior author

* jsenta@berkeley.edu



Abstract

Anxiety has been robustly linked to deficits in frontal executive function including working memory (WM) and attentional control processes. However, although anxiety has also been associated with impaired performance on learning tasks, computational investigations of reinforcement learning (RL) impairment in anxiety have yielded mixed results. WM processes are known to contribute to learning behavior in parallel to RL processes and to modulate the effective learning rate as a function of load. However, WM processes have typically not been modeled in investigations of anxiety and RL. In the current study, we leveraged an experimental paradigm (RLWM) which manipulates the relative contributions of WM and RL processes in a reinforcement learning and retention task using multiple stimulus set sizes. Using a computational model of interactive RL and WM processes, we investigated whether individual differences in physiological or cognitive anxiety impacted task performance via deficits in RL or WM. Elevated physiological, but not cognitive, anxiety scores were strongly associated with worse performance during learning and retention testing across all set sizes. Computationally, higher physiological anxiety scores were significantly related to reduced learning rate and increased rate of WM decay. To highlight the importance of modeling WM contributions to learning, we considered the effect of fitting RL models without WM modules to the data. Here we found that reduced learning performance for higher physiological anxiety was at least partially misattributed to stochastic decision noise in 9 out of 10 RL-only models considered. These findings reveal a dual-process impairment in learning in anxiety that is linked to a more physiological than cognitive anxiety phenotype. More broadly, this work also points to the importance of accounting for the contribution of WM to RL when investigating psychopathology-related deficits in learning.

OPEN ACCESS

Citation: Senta JD, Bishop SJ, Collins AGE (2025) Dual process impairments in reinforcement learning and working memory systems underlie learning deficits in physiological anxiety. PLoS Comput Biol 21(9): e1012872. <https://doi.org/10.1371/journal.pcbi.1012872>

Editor: Lyle J. Graham, Centre National de la Recherche Scientifique, FRANCE

Received: February 11, 2025

Accepted: September 2, 2025

Published: September 26, 2025

Copyright: © 2025 Senta et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Data has been made publicly available at <https://osf.io/w8ch2/>.

Funding: The study was supported by the National Institute of Mental Health (NIMH) grant numbers R01MH124108-05 and R01MH122558-04 to SJB. The funders had

no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Individuals with anxiety may have difficulties with learning, memory, and attentional control. Computational studies investigating the specific learning processes that are impaired in anxiety have yielded inconsistent results, possibly because they generally fail to account for the role of supporting systems such as working memory (WM). Moreover, anxiety itself is a multi-dimensional construct with symptoms which vary across individuals. In the current study, we employed an experimental paradigm which accounted for the supporting role of WM in reinforcement learning (RL) in a learning and retention task. We tested physiological and cognitive dimensions of anxiety against RL and supporting processes. We found that physiological, but not cognitive, anxiety was behaviorally linked to reduced learning and testing performance. Computationally, we found that these performance deficits were attributable to both RL (via reduced learning rate) and WM (via increased forgetting rate) systems. To highlight the importance of modeling WM contributions to learning, we tested a series of simpler RL-only models on our data. Here we found that models which did not account for WM tended to at least partially misattribute performance deficits to noise.

Introduction

Our ability to learn from our experiences of the world is a crucial element of successful decision-making and ultimately survival. Along with other dimensions of psychopathology, anxiety has been behaviorally linked to impairments in learning, including slower learning and reduced performance [1]. Models of reinforcement learning (RL; [2]) have been successfully used to investigate the cognitive mechanisms of learning across animals and humans. Extending this work into the clinical domain, RL models have been used to investigate the influence of psychopathology upon learning [3]. Here, there remains a lack of clarity regarding the precise effects of anxiety on the computational mechanisms supporting reinforcement learning [4]. This is complicated by differences between studies in both the experimental paradigms used [5–9] and in whether anxiety symptoms are directly measured or if a stressor manipulation is used as a proxy for anxiety [4,10]. In relation to the latter, prior work suggests that individual differences in anxiety-related psychopathology and induced anxiety may have very different neural signatures [11].

One crucial aspect of RL is an individual's ability to modulate their effective learning rate in accordance with the current environment. Various forms of environmental uncertainty require dynamic adjustments to learning by incorporating recent information more quickly into value expectations [12–14]. Humans are remarkably adept at appropriately adjusting to different environments [15–17], though there is evidence to suggest that this flexibility might be compromised in anxiety, and internalizing psychopathology more generally [5,18,19].

One way in which effective learning rates may be modulated is by an adjustment of the relative recruitment of different neural processes within the brain. In particular, working memory (WM) processes are known to contribute to fast learning behavior in parallel to slower RL processes, and thus to modulate the effective apparent learning rate as a function of load [20–22]. While RL processes have been neurally linked to several brain regions including the ventral tegmental area (VTA), the striatum and cortico-basal-ganglia circuitry [23,24], WM processes have primarily been linked to the brain's prefrontal cortical executive control systems [25–28].

Although reinforcement learning has often been studied without consideration of WM, research has shown that reinforcement learning processes are supplemented by WM systems [20,22,29,30], and WM capacity has been positively associated with performance on reinforcement learning tasks, particularly during early-stage learning [22,23,31–33]. Further, under certain circumstances, WM processes can additionally interfere with RL [20,34,35].

To address the parallel and contributory processes of working memory during reinforcement learning, a line of recent research has successfully introduced a new experimental paradigm which differentially manipulates the relative contributions of working memory and reinforcement learning systems during a learning and retention task. These studies have shown that jointly modeling the WM and RL processes can reveal key features of learning which cannot be accounted for with stand-alone RL models [21,36]. This is achieved using models where separate modules for short-term WM (parameterizing WM capacity, decay rate, and confidence weighting) and longer-term RL (parameterizing learning rate and other variables of interest) can both variably contribute to task performance. Research on individual differences has further shown that using the RLWM framework can help more precisely identify the distinct mechanisms underlying apparently similar learning impairments; for example, learning impairments in schizophrenia appear driven by WM capacity limitations [21], in older adults by faster WM decay [37], and in young children by slower RL and weaker WM involvement [36]. Thus, failure to take WM processes into account might also contribute to inconsistencies in findings within the anxiety RL literature.

The need to consider the contribution of working memory processes to anxiety-related deficits in learning is further supported by findings directly linking anxiety to aberrations in the use of aspects of prefrontal cortical executive systems [38,39] including specific deficits in working memory [40]. In a study of N-back task performance in both safe and threatening environments, patients diagnosed with anxiety disorders (ADs) had impaired performance and reaction times across both environments [41]. In addition, AD patients showed impaired recruitment of prefrontal cortical regions during these WM tasks. Other studies examining the same frontal regions have also hinted at a dimensional specificity to the role of anxiety in impaired executive function. An fMRI investigation of performance and cortical activity during a sustained attention task found that a general measure of trait anxiety was linked to impoverished recruitment of frontal regions at points where adjustments of attentional control were required, whereas a specific measure of worry separately impacted frontal-default mode connectivity during periods where attention lapsed [39]. In another study, only anxious arousal, and not other indices of anxiety or negative affect, was linked to impaired perceptual decision-making [42]. These findings highlight the possibility that concurrent investigation of different subdimensions of anxiety may be crucial for understanding anxiety-related differences in cognitive function.

Here we sought to use the enhanced RLWM experimental and modeling framework to disentangle pure RL from working memory contributions to differences in learning as a function of dimensions of anxiety. Given prior findings indicating that physiological and cognitive dimensions of anxiety may be differentially associated with impairments in processes spanning attentional control, working memory, and reinforcement learning, we sought to test our hypotheses with respect to specific characterizations of each symptom domain. We used scores on self-report measures of anxious arousal (Mood and Anxiety Symptom Questionnaire [43] (MASQ) Anxious Arousal (AA) subscale) and worry (Penn State Worry Questionnaire [44] (PSWQ)) to evaluate the influence of these two distinct dimensions of anxiety on learning and working memory involvement during reinforcement learning. Given the high rate of comorbidity between anxiety and depression, we performed a supplemental exploratory analysis of the relationship between the two anxiety measures mentioned, plus two

common self-report measures of depressive affect (CES-D [45] and BDI-II [46,47]), with all the parametric mechanisms of the winning RLWM model to inform potential directions for future research across the two highly comorbid disorders (see Supplemental Analyses in [S1 Text](#)).

Finally, to highlight the importance of factoring in WM contributions to learning, we considered the effect of fitting RL models without WM modules to the experimental data for the current study. Here we examined two classes of RL models (using single or set size dependent learning rates) with varying parameterizations to include mechanisms of choice perseveration, forgetting, and negative feedback neglect. We assessed the mechanistic attribution of impaired learning performance and compared/contrasted it to the findings from the best-fit joint-process RLWM model to determine whether the RLWM model was able to improve specificity of findings relative to more common RL model formulations.

Results

Accuracy

In the current study, we employed a deterministic reward-based stimulus-action learning task which presents varying set sizes of stimuli by block, with stimulus set sizes (nS) ranging from 2 to 6 (see [Fig 1a](#)), in order to differentially engage WM versus RL systems during learning. During the initial “learning” phase, participants completed a series of 13 blocks. In each block participants learned associations between images and three key presses ('J', 'K', 'L'). The number of images presented in each block was varied, ranging from 2 to 6 images per block, with 13 trial repetitions for each image. Following a distractor task, a surprise “testing” phase was presented which measured learning retention after WM had decayed. A total of $n = 164$ participants were included in the analysis; see Methods for complete details.

We first sought to replicate previous findings of joint RL and WM involvement at a group level. During learning, participants performed better than chance (chance = $1/3$), with mean accuracy across all set sizes of 78.2%. All set sizes had high mean overall performance above 70% accuracy (see [Fig 1b](#)), with set size 2 having the highest mean performance (mean = 85%) and set size 6 the lowest (mean = 75%). In line with previous literature, we observed that as set size increased, accuracy decreased: Overall set size slope (see Methods) median was 0.204 and was significantly greater than 0 for the group (Wilcoxon one-sample test statistic = 571.0, $p = 0.000$), indicating a significant effect of set size in (decreasing) performance accuracy.

The testing phase followed a distraction task, the aim of which was to eliminate the contribution of working memory to performance. In line with this, participant performance in the testing phase was substantially lower than during learning (Wilcoxon one-sided test statistic = 13003.0, $p = 6.42e-25$). Overall performance during testing still exceeded chance (chance = $1/3$) with a mean accuracy of 65.2% across participants and set sizes. Despite greater accuracy during the learning phase for small set size stimuli, accuracy during testing was not significantly modulated by set size (set size slope median = -0.017, Wilcoxon one-sided test statistic = 6176.000, $p = 0.334$), reflecting a greater relative retention of learned associations at higher set sizes (see [Fig 1b](#)). Indeed, the difference in performance between learning and testing (which effectively factors out set size effects in learning accuracy) was significantly negatively associated with set size (Kruskal-Wallis = 80.075, $p = 1.65e-16$; see [Fig 1c](#)). This superficially counter-intuitive finding replicates previous findings and is consistent with greater reliance on working memory in the initial learning phase at smaller set sizes, but also of interference of WM blocking RL in smaller set sizes [20,37,48,49].

Relationship of anxiety with performance

We used the MASQ AA subscale, which assesses self-report items specific to anxious arousal, to obtain a measure of physiological anxiety. We note that this captures somatic symptoms in a similar manner to the STICSA somatic anxiety subscale which has recently been used elsewhere in the computational decision-making literature [8,9]. The overall group mean score (\pm standard deviation) on the MASQ AA was 27.25 \pm 10.86; see Fig A in [S1 Text](#). Scores on the MASQ

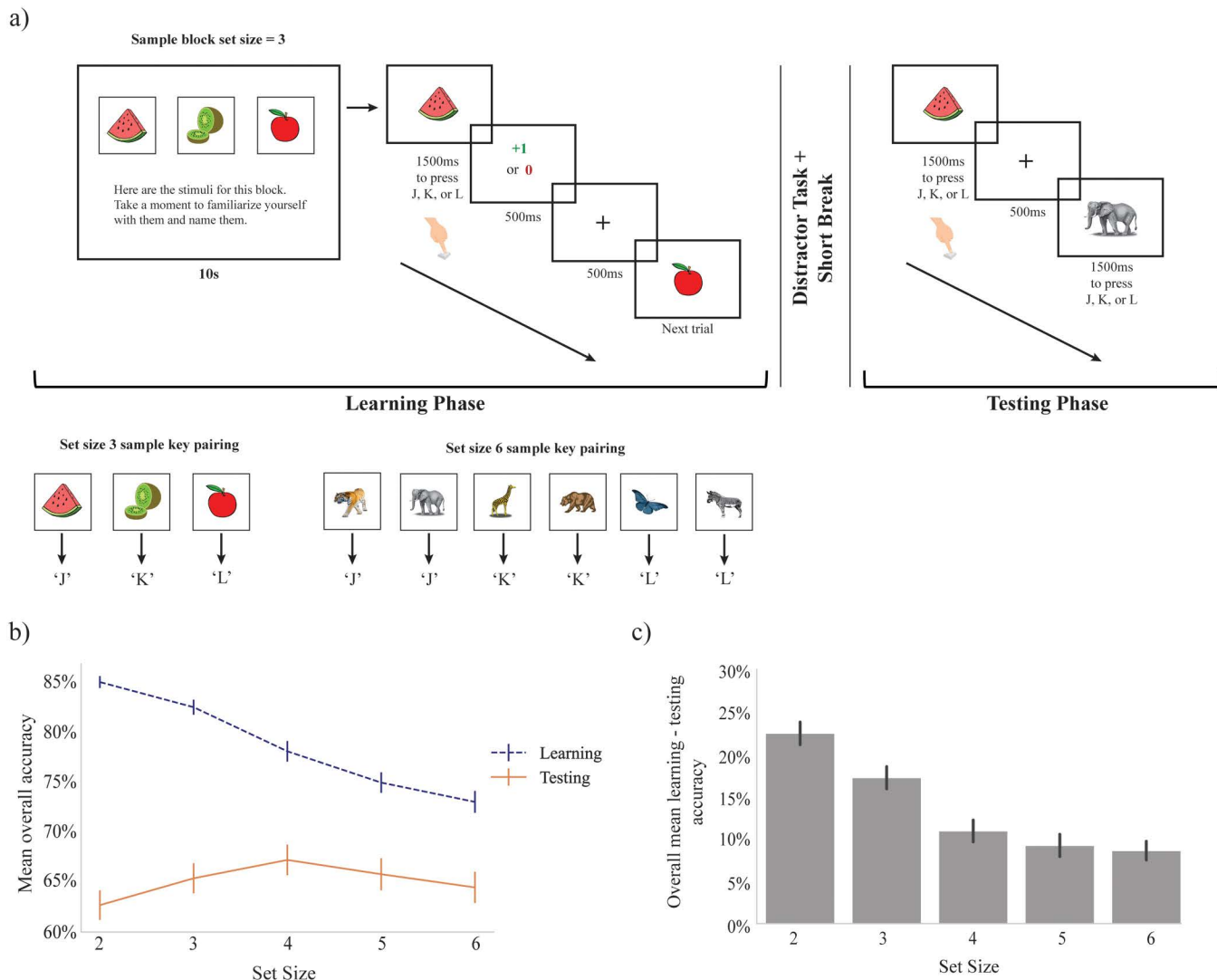


Fig 1. RLWM task and behavior. a) Participants completed 13 blocks of trials. Each block had a set size of 2, 3, 4, 5, or 6 images to be learned, and each stimulus was shown 13 times within each block for up to 1.5s each presentation (terminating when a key was pressed). Images for a block were related to each other (e.g., “Instruments”) and were unique for each block. Each image in a block was associated with one correct response key from 3 possible keys ('j', 'k', 'l'). Participants learned the correct key to press for each image via trial and error, with immediate feedback displayed for 500ms after each key press (+1 if correct, 0 if incorrect). After a distractor task (an N-back task using unrelated shapes, not used for main analysis), participants were given a surprise testing phase in which each stimulus from the learning phase was shown 3 times (in shuffled random order across all stimuli). No feedback was provided in the testing phase. b) Mean overall accuracy during learning was significantly lower for high versus low set sizes, while mean overall accuracy during testing was not significantly modulated by set size. c) There was a significant drop in performance between mean overall learning versus testing accuracy at each set size; this difference was negatively associated with increasing set size.

<https://doi.org/10.1371/journal.pcbi.1012872.g001>

AA subscale were significantly negatively correlated with mean overall performance during learning (Spearman rho(162) = -0.299, $p=9.79e-5$; see Fig 2a for illustration using median split on MASQ AA scores). Further, individuals with higher MASQ AA scores also showed significantly greater effect of set size (as measured by set size slope; see Methods) on performance during learning (Spearman rho(162) = 0.220, $p=0.005$). See Fig 2g for illustration of learning curves by set size using median split on MASQ AA scores.

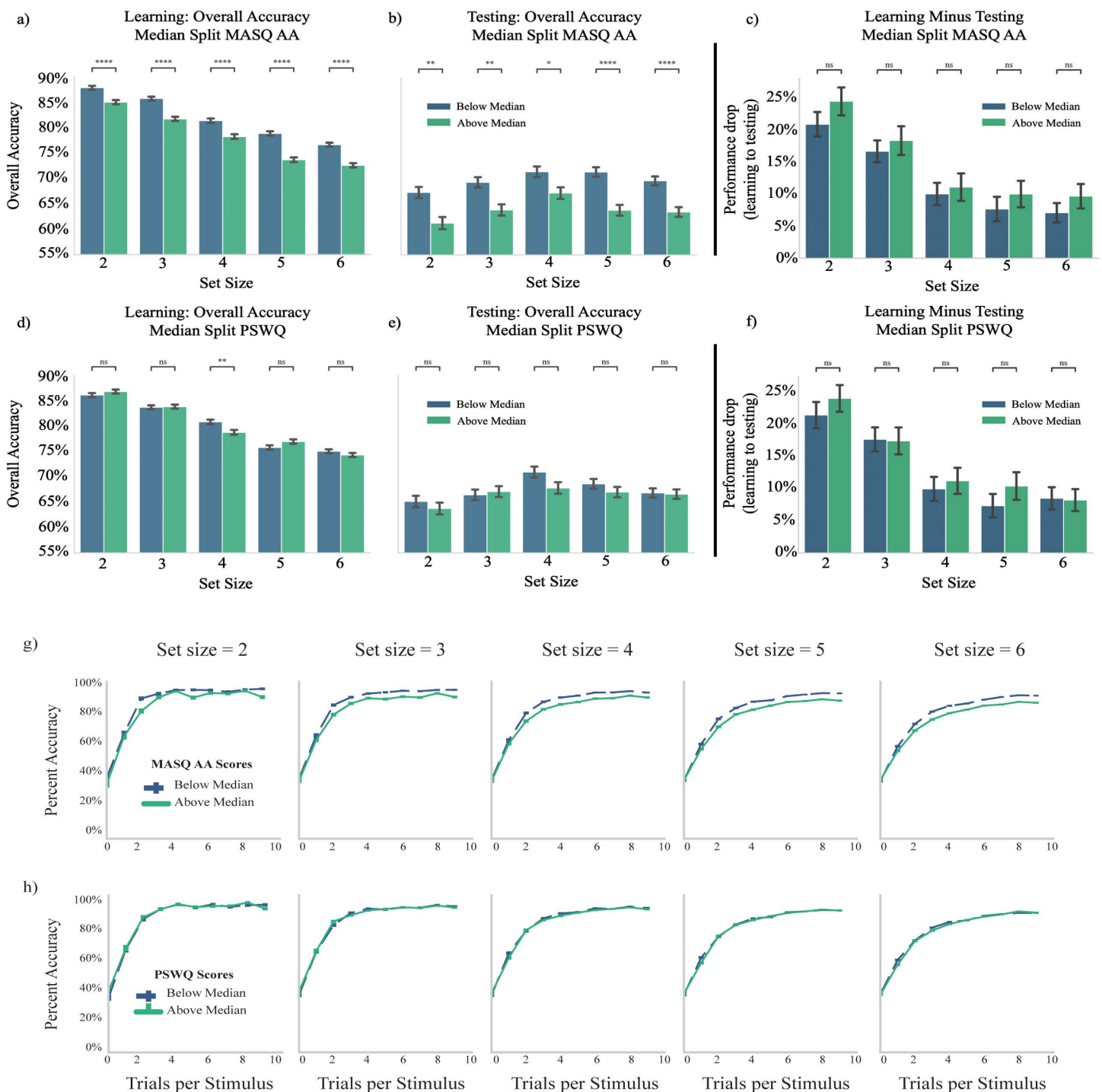


Fig 2. Median split of learning and testing performance on measures of trait anxiety. a,b) Individuals with above-median scores on the Mood and Anxiety Symptom Questionnaire (MASQ) Anxious Arousal (AA) subscale show significantly lower performance accuracy relative to below-median scores at every set size during both learning (a) and testing (b) phases of the RLWM task. c) There was no significant difference in performance drop between learning and testing phases for above-median MASQ AA relative to below-median MASQ AA scoring participants at any set size. d,e) Individuals with above-median scores on the Penn State Worry Questionnaire (PSWQ) do not have significantly different performance relative to below-median scores at any set size during learning except set size = 4, and no significantly different performance at any set size during testing. f) There is no significant difference in performance drop between learning and testing phases for above-median PSWQ relative to below-median PSWQ scoring participants at any set

size. g) Learning curves for each set size split by median scores on MASQ AA reveal relative differences in learning speed and ultimate performance. h) Learning curves for each set size split by median scores on PSWQ reveal no relative differences in learning speed and ultimate performance.

<https://doi.org/10.1371/journal.pcbi.1012872.g002>

As a complementary analysis, we conducted a repeated measures ANCOVA with within-subject factor of set size, covariate of z-scored MASQ AA scores, and dependent variable of mean learning performance. This revealed a significant within-subject effect of set size (within-subject set size effect with Greenhouse-Geisser correction applied $F(3.430, 555.588) = 66.143$, $p < 0.001$); a significant main effect of MASQ AA (between-subjects MASQ AA effect $F(1, 162) = 37.729$, $p < 0.001$); and a significant interaction of set size with MASQ AA (within-subject set size x MASQ AA effect with Greenhouse-Geisser correction applied $F(3.430, 555.588) = 3.844$, $p = 0.007$).

We next investigated whether physiological anxiety also affected long term retention in the testing phase. Indeed, higher scores on the MASQ AA subscale were also significantly associated with reduced overall testing performance (Spearman $\rho(162) = -0.289$, $p = 0.000$; see [Fig 2b](#) for illustration using median split on MASQ AA scores). Of note, MASQ AA scores were not associated with any effect of set size on performance during the test phase (Spearman $\rho(162) = 0.099$, $p = 0.205$). This may reflect the differing role of WM during learning and testing and hints at the possibility that high and low MASQ AA participants might show differential reliance on WM during learning. A repeated measures ANCOVA with within-subject factor of set size, covariate of z-scored MASQ AA scores, and dependent variable of mean testing performance revealed a significant main effect of set size (within-subjects set size effect with Greenhouse-Geisser correction $F(3.744, 606.481) = 4.000$, $p = 0.004$) and a significant main effect of MASQ AA scores (between-subjects MASQ AA effect $F(1, 162) = 17.956$, $p < 0.001$), but only a weakly trending effect of interaction between set size and MASQ AA (within-subjects set size x MASQ AA effect with Greenhouse-Geisser correction $F(3.744, 606.481) = 2.046$, $p = 0.091$).

To more directly investigate whether the effect of physiological anxiety on testing performance was due to differences during learning, we next analyzed test phase accuracy relative to asymptotic learning phase accuracy. Higher MASQ AA subscale scores were significantly associated with reduced relative asymptotic performance between set size 2 and set size 6 during learning (Spearman $\rho(162) = 0.207$, $p = 0.008$), but not with relative overall performance between set size 2 and set size 6 during testing (Spearman $\rho(162) = 0.079$, $p = 0.314$). We additionally tested the relationship between MASQ AA subscale scores and drop between asymptotic accuracy over last 3 trials during learning and overall performance during testing. A repeated measures ANCOVA with within-subject factor of set size, covariate of z-scored MASQ AA scores, and dependent variable of drop in performance between mean of final 3 learning trials and mean overall testing performance revealed a main effect of set size (within-subjects set size effect $F(4, 648) = 8.967$, $p < 0.001$), but no main effect of MASQ AA (between-subjects MASQ AA $F(1, 162) = 0.135$, $p = 0.713$) or interaction of MASQ AA with set size (within-subjects set size x MASQ AA effect $F(4, 648) = 0.537$, $p = 0.708$). See [Fig 2c](#) for illustration using median split on MASQ AA scores. This may indicate that testing phase effects were primarily driven by differences in learning experience.

Unexpectedly, scores on the cognitive measure of anxiety (the PSWQ) did not appear to significantly impact task performance. The overall group mean score (\pm standard deviation) on the PSWQ was 52.12 ± 13.81 ; see [Fig A in S1 Text](#)). During learning, PSWQ scores were not significantly associated with mean overall aggregate learning performance (Spearman $\rho(162) = -0.091$, $p = 0.248$), but had a trending relationship with learning phase set size slope (Spearman $\rho(162) = 0.149$, $p = 0.057$). There were no significant effects of PSWQ scores on performance during testing, either across set sizes (Spearman $\rho(162) = -0.074$, $p = 0.347$) or as a function of set size (Spearman $\rho(162) = 0.009$, $p = 0.910$). See [Fig 2d–2f, 2h](#) for illustrations using median split on PSWQ scores. See Supplemental Analyses in [S1 Text](#) for the associated ANCOVA results; here neither the main effect of PSWQ nor the interaction of PSWQ by set size was significant for either phase of the task.

Computational modeling

To investigate the computational mechanisms underlying learning and testing performance, we fitted a series of RLWM models [20–22] to the study data. Unlike RL-only models with multiple learning rates, these RLWM models, which use separate but interacting RL and WM modules, are able to replicate the asymmetric drop in accuracy between low and high set sizes seen in participants (see Fig 3 for illustration). Varying combinations of parameters were considered (see Methods), and model fit was measured by the Akaike Information Criteria (AIC; see Methods).

The model which best represented the overall behavioral data as measured by (lowest) total AIC was Model #5 (RLWM- $\eta_{asym_2\rho K}$; see Table 1, Fig 7, and Methods). This model included the following parameters: RL learning rate α_{RL} , which applied when reward = 1 (RL learning rate = 0 when reward = 0 in winning model variant); stochastic choice noise ε ; softmax inverse temperature for testing phase choice (β_{test} ; softmax inverse temperature for learning phase was fixed at $\beta = 50$ to improve parameter fitting, and choice noise in learning is captured by ε); bias parameter η_{WM} which controlled negative feedback neglect in WM module; WM decay/forgetting parameter φ_{WM} ; and parameter i , which controlled information sharing between WM and RL for RL prediction error calculation. See Fig B in S1 Text for distributions of parameter values. Based on recent work regarding potential bias in some RL models [50], we performed a robustness check of the

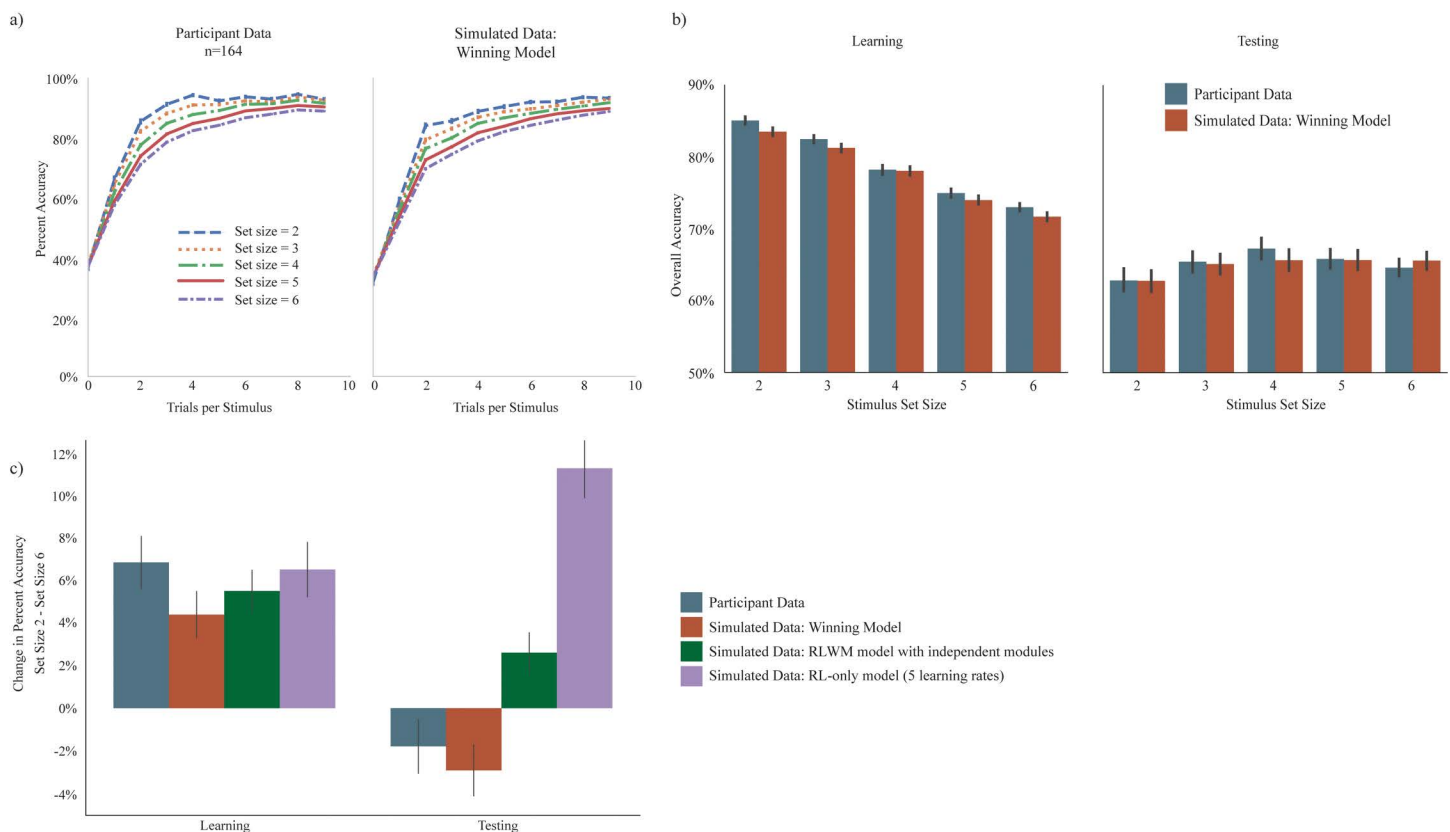


Fig 3. Model validation. a) Average learning curves across each set size represented as mean percent accuracy per stimulus presentation count. Left panel a; actual participant data. Right panel a; data simulated using the winning model (Model #5). b) Overall accuracy at each set size during learning (left panel b) and testing (right panel b) for actual participant data versus data simulated using the winning model (Model #5). c) Performance difference between set size = 2 and set size = 6 (mean percent accuracy SS2 – mean percent accuracy SS6) during learning (left) and testing (right) phases of the RLWM task. Participant data shows a reversal of this difference between learning and testing. This pattern is reproduced by simulated data from the winning model (Model #5), but is not captured by either an RLWM model variant with fully independent RL and WM modules or by an RL-only model with separate learning rates for each set size (Model RL- $5\alpha_{RL}\varphi_{RL}\eta_{RL}$; see methods for model details).

<https://doi.org/10.1371/journal.pcbi.1012872.g003>

Table 1. Computational models of reinforcement learning evaluated.

Model	Parameters	Description	Total AIC
Model #1: RLWM	$\alpha, \beta_{\text{test}}, \varepsilon, \rho, \varphi, K, i$	Basic RL and WM model with parameter i to coordinate information sharing between modules	146,311
Model #2: RLWM _{bias}	$\alpha, \beta_{\text{test}}, \varepsilon, \rho, \varphi, K, \eta, i$	Same as RLWM, but updates following negative feedback are partially neglected according to parameter η	143,910
Model #3: RLWM _{bias_2r}	$\alpha, \beta_{\text{test}}, \varepsilon, \rho_{\text{low_ss}}, \rho_{\text{high_ss}}, \varphi, K, \eta, i$	Same as RLWM _{bias} , but allows WM choice weight to vary between set sizes below versus above capacity	143,799
Model #4: RLWM _{asymbias}	$\alpha, \beta_{\text{test}}, \varepsilon, \rho, \varphi, K, \eta_{\text{asym}}, i$	Same as RLWM _{bias} , but negative feedback is neglected asymmetrically for RL (totally neglected) versus WM (partially neglected via η)	143,718
Model #5: RLWM _{asymbias_2r}	$\alpha, \beta_{\text{test}}, \varepsilon, \rho_{\text{low_ss}}, \rho_{\text{high_ss}}, \varphi, K, \eta_{\text{asym}}, i$	Same as RLWM _{asymbias} , but allows WM choice weight to vary between set sizes below versus above capacity	143,629
Model #6: RLWM _{asymbias_2r_ck}	$\alpha, \beta_{\text{test}}, \varepsilon, \rho_{\text{low_ss}}, \rho_{\text{high_ss}}, \varphi, K, \eta_{\text{asym}}, i, \tau, \kappa$	Same as RLWM _{asymbias_2r} , but includes choice kernel to reflect decaying influence of past action choices on current choice	143,770

For each model: model key, list of parameters, short description, and total group AIC are shown.

<https://doi.org/10.1371/journal.pcbi.1012872.t001>

winning model by confirming that the addition of a perseveration choice kernel (Model #6) did not improve model fit, and did not result in significant perseveration or significant changes in main model parameters; see Supplemental Analyses in [S1 Text](#) and Fig E in [S1 Text](#) for details.

The total AIC of the winning model (Model #5; 143,629) was very close to the total AIC of the second-best model (Model #4; 143,718). We therefore confirmed in supplemental analysis that all results below remained significant when tested in Model #4; see Supplemental Analyses in [S1 Text](#).

Model based analyses

We first verified that model parameters were identifiable in the winning model by simulating data with fixed parameter values and assessing the accuracy of the parameter values fit to the simulated data (see Fig C in [S1 Text](#)). Additionally, we verified that models themselves were identifiable by performing a model recovery analysis, simulating three data sets for each of three selected models, and fitting each of the resulting nine data sets with each of the selected models to confirm that the generative model best fit its own data in each case (see Fig D in [S1 Text](#)). Furthermore, model validation showed that the winning model (Model #5) could capture the learning dynamic and testing phase accuracy well (see [Fig 3](#)).

Reduced learning and elevated WM decay associated with physiological anxiety

Given the significant behavioral impairment to learning and testing phase performance for higher scores on the MASQ AA, we examined the relationship of these scores with specific cognitive mechanisms of learning and working memory use as quantified by the winning RLWM computational model (Model #5). All results held when separately tested for significance in the second-best fitting model (Model #4); see Supplemental Analyses in [S1 Text](#).

We performed initial hypothesis testing for participant scores on physiological arousal (MASQ AA scores) and cognitive worry/anxiety (PSWQ; for completeness) against one RL-specific parameter (learning rate α), two WM-specific parameters (forgetting φ_{WM} , and neglect of negative feedback in WM η_{WM}), and two parameters indexing relative contribution of WM to the policy (WM confidence at low set sizes ρ_{low} , WM confidence at high set sizes ρ_{high}) from the winning model (see Methods).

Higher MASQ AA subscale scores were significantly related to lower RL learning rate ($\rho(162) = -0.224$, FWE-corrected $p=0.040$) and higher rate of working memory decay ($\rho(162) = 0.223$, FWE-corrected $p=0.040$; see [Fig 4](#)). Prior to statistical correction for multiple comparison, higher MASQ AA subscale scores were also associated with a greater bias against

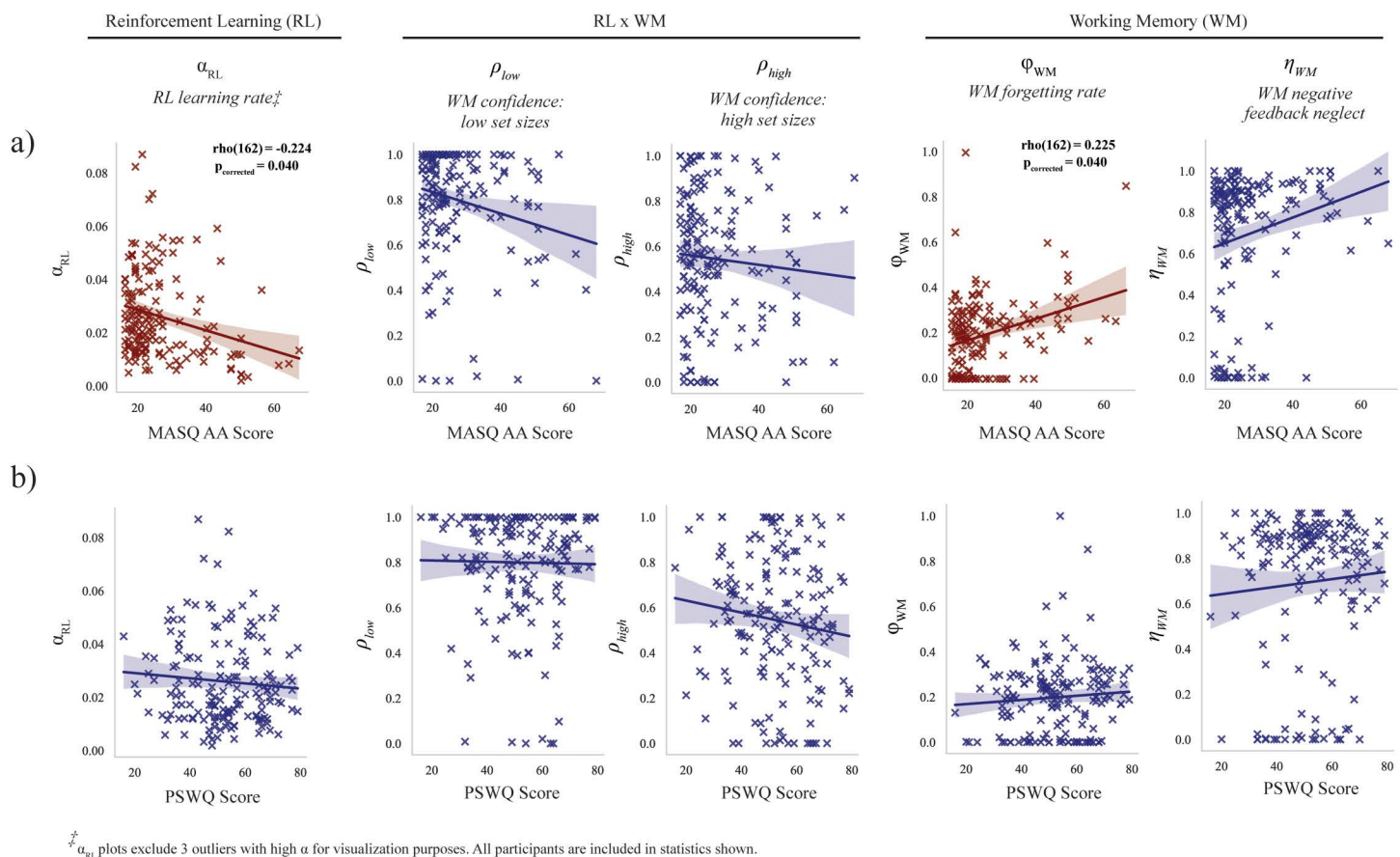


Fig 4. Relationship of model parameters with individual scores on anxiety measures. Relationship of learning and working memory model parameters from the winning model (Model #5) with individual scores on MASQ AA (a) and PSWQ (b).

<https://doi.org/10.1371/journal.pcbi.1012872.g004>

the use of negative feedback in working memory, but this did not survive correction for FWE ($\rho(162) = 0.199$, uncorrected $p = 0.010$, FWE-corrected $p = 0.104$). No significant associations were found between PSWQ scores and any of the specified parameter values.

From left to right, parameters tested: RL learning rate, WM forgetting rate, WM confidence weight for low and high set sizes, and WM bias against negative feedback updating.

RL models without WM module vary in characterization of role of anxiety

Modeling learning with joint RL and WM processes revealed that two factors contributed to differences in learning as a function of MASQ AA. We performed an exploratory analysis investigating how our finding would be interpreted should modeling fail to account for WM. To do so, we additionally fit our data with a series of 10 variants of canonical RL-only models that did not include WM modules (see Methods). This allowed us to examine what our computational model-based findings would be if the data were fit with typical reinforcement learning models which do not account for WM. Given the structure of the task, which features 5 different set sizes across blocks, we considered simple RL models with either one shared learning rate across all set sizes (RL_α models) or with a separate learning rate for each set size (RL_5α models). In each class, we considered a base RL model (with parameters comprising learning rate(s), softmax inverse temperature for testing phase β_{test} , and undirected choice noise ϵ), as well as four additional model variants which included various

combinations of parameters for choice perseveration, reinforcement learning decay (forgetting), and negative feedback neglect; see Methods for details.

None of the RL-only models fit the data as well as the winning RLWM model; however, the model fit as measured by total (sum of) group AIC was in the same range as the RLWM class, with the best-fitting RL-only model fitting the data better (based on AIC comparison) than the simplest RLWM model (Model #1) (see Fig 5).

Within each class of RL-only models, the model variant which included forgetting (φ_{RL}) and negative feedback neglect (η_{RL}), but no choice kernel, fit the data best based on a comparison of AIC (see Fig 5). Across all RL-only models, the best-fitting model was the set size specific learning rate (5-alpha) version of this variant (RL_5 α _ φ_{RL} _ η_{RL} , AIC = 144,170; see Fig 5).

Correlations between model parameters and scores on the MASQ AA varied somewhat across models depending upon specific parameterization; however, most models (8 out of 10) found a significant relationship between higher MASQ AA scores and higher undirected choice noise (ϵ) which was not present in the winning RLWM model, suggesting that RL-only models are likely to at least partially misattribute learning differences in anxiety to noise rather than deficits in learning or working memory (see Fig 6). The winning RL-only model (RL_5 α _ φ_{RL} _ η_{RL} ; see Fig 6) found that higher MASQ AA scores were significantly related to increased choice noise ϵ ; $\rho(162)=0.238$, $p_{\text{corrected}}=0.022$, increased forgetting in RL φ_{RL} ; $\rho(162)=0.279$, $p_{\text{corrected}}=0.003$, and increased neglect of negative feedback (η_{RL} ; $\rho(162)=0.353$, $p_{\text{corrected}}<0.001$). See Fig 6 for significant parameter relationships with MASQ AA for each RL-only model as compared with the winning RLWM model (Model #5).

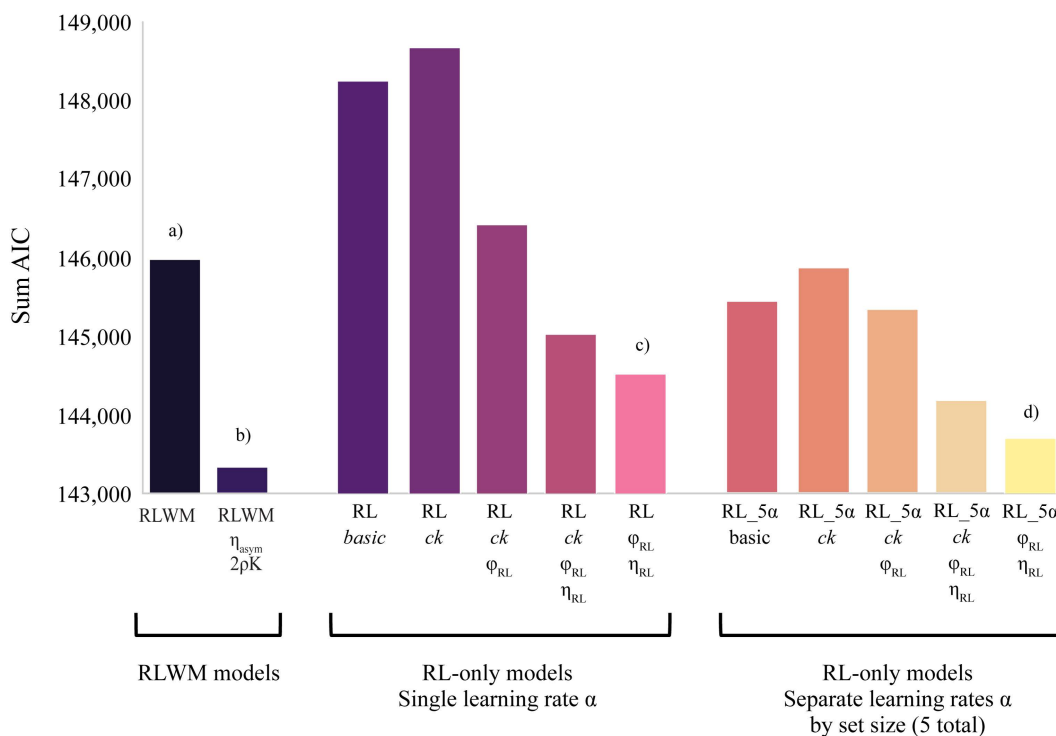


Fig 5. Model fit comparison (AIC) across base and winning RLWM model variants and 10 RL-only models tested. a) AIC for base RLWM variant. b) Winning RLWM model (Model #5) had best fit (lowest AIC) across all model variants tested. c) Best-fitting model variant in RL- α class. d) Best-fitting model variant in RL-5 α class was best fit across all RL-only variants tested.

<https://doi.org/10.1371/journal.pcbi.1012872.g005>

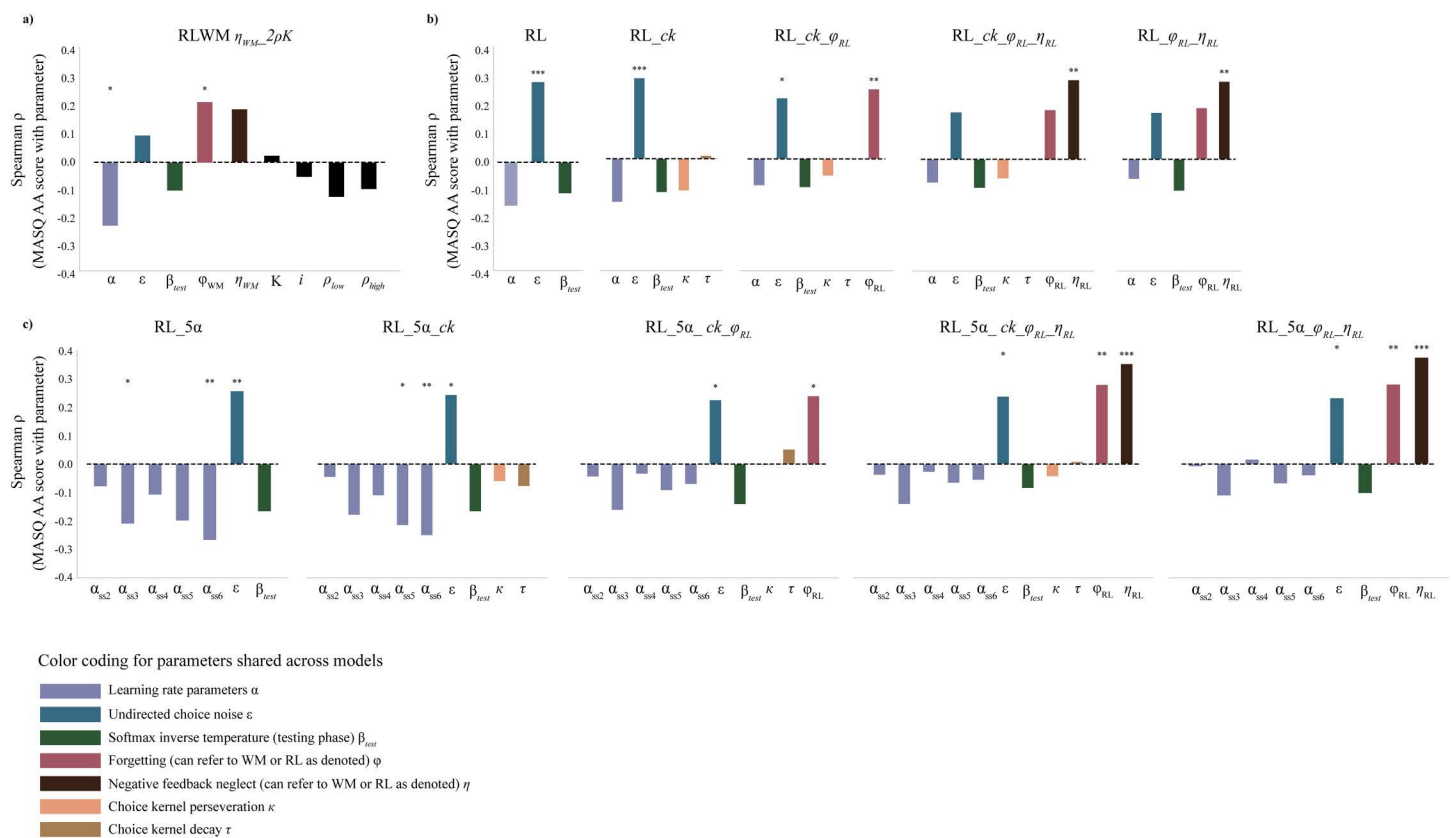


Fig 6. Comparison of relationship between MASQ AA scores and model parameters for winning RLWM model plus 10 RL-only model variants. Reinforcement learning models which do not specifically include the effect of working memory tend to attribute performance differences in high MASQ AA scoring participants to greater undirected choice noise, as well as to alternative mechanisms such as choice repetition and negative feedback neglect depending upon the specific model parameterization. a) Winning RLWM model from main analysis (Model #5) shows that higher MASQ AA scores are significantly related to reduced learning rate and increased working memory decay rate (forgetting). b) Results from 5 variants of RL-only models which feature a single learning rate parameter across all set sizes. Models considered include base model (learning rate (α), test phase inverse temperature (β_{test}), and undirected choice noise (ϵ)), as well as incrementally including the following parameters: choice kernel (ck), RL forgetting (ϕ_{RL}), and partial negative feedback neglect in RL (η_{RL}). c) Results from 5 parallel variants of RL-only models which feature a separate learning rate for each set size. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

<https://doi.org/10.1371/journal.pcbi.1012872.g006>

Discussion

Higher levels of anxiety have been separately associated with impairments in learning, working memory, and broader executive function in a number of studies [1,18,39,41,51,52]. However, the influence of working memory on reinforcement learning processes in anxiety-related psychopathology has not to our knowledge been computationally investigated. In the current analysis, we used an experimental paradigm specifically designed to manipulate the relative load on learning versus working memory systems [22]. We applied a computational model of reinforcement learning which accounts for the supportive role of working memory during learning to test whether relationships between anxiety and learning were attributable to learning processes, working memory processes, or both. We used two measures of anxiety-related psychopathology, one specifically characterizing physiological symptoms of anxiety and one characterizing cognitive symptoms of anxiety, to test any potential dimensional specificity in results.

Our findings indicate a complex picture of the relationship between anxiety, learning, and working memory. Firstly, we found no behavioral relationship between cognitive anxiety (as measured by PSWQ scores) and reduced performance

in learning or testing across the task (with the exception of reduced learning at set size = 4). This was consistent with the results of our computational modeling analysis: no parameters of the winning RLWM model were found to have significant relationships with increased levels of cognitive anxiety.

In contrast, we found significant behavioral impairment in task performance at every stimulus set size across both learning and testing phases for individuals with higher levels of physiological anxiety (as measured by MASQ AA subscale scores). There was no significant anxiety-related difference in performance drop between learning and testing phases, highlighting the complex interactions between RL and WM processes and the consequent inability of simple behavioral analyses to dissociate the underlying cause(s) of the performance impairment. This impairment was attributable to multiple parameterized mechanisms in computational modeling analysis. After correcting for multiple comparisons, MASQ AA subscale scores were related to a significantly lower learning rate and a significantly higher rate of working memory decay. This set of results suggests a multi-faceted relationship of anxiety with learning and its supportive processes, whereby both the learning process itself as well as the role of the supporting working memory systems are compromised in high levels of physiological anxiety. Here it is interesting to note that in a computational investigation of directed (information seeking) and undirected (random) exploration, it was also a measure of physiological anxiety that was linked to reduced estimation of relative uncertainty, reduced directed exploration and, to a lesser extent, reduced random exploration. Here, as in the current study, a measure of cognitive anxiety showed no significant influences on any of the parameters of interest [9]. Meanwhile, work using an aversive learning task reported a complex pattern of dissociations between cognitive and somatic indices of anxiety in their influence on task performance [8]. Our work adds to this literature and points to the need for further interrogation of the specific processes impacted by physiological versus cognitive subdimensions of anxiety.

Most studies of reinforcement learning within computational psychiatry have used paradigms and models which do not specifically account for working memory systems; indeed, our ability to model working memory contributions here is due to the experimental load manipulation and is thus not easily applied retrospectively to existing reinforcement learning datasets. An intriguing possibility is that inconsistencies in findings regarding humans, and indeed other species', ability to learn and update contingencies [53] might to some extent reflect a failure to take working memory mechanisms into account. Here we sought to assess what our findings for the current dataset would be when modeled without the inclusion of the working memory module. For comparison, we fit a series of more typical, RL-only models which did not account for contributions of WM to the dataset and examined the correlation between model parameters and MASQ AA scores for these model classes. Of note, 5 out of 6 RL-only models which included a parameter for forgetting (φ_{RL}) found a significant association between rate of forgetting and MASQ AA scores, in parallel to the main analysis (Model #5) finding that higher MASQ AA scores were tied to faster forgetting in working memory processes (φ_{WM}) (see Fig 6). Crucially, models which did not account computationally for WM contributions largely attributed the decreased learning performance for high MASQ AA scores to undirected choice noise (ϵ ; 9 out of 10 models) and increased negative feedback neglect (η_{RL} ; 4 out of 4 models which included this parameter). The specific characterization of impaired mechanisms in higher MASQ AA scores varied depending on the particular parameterization of the models. Only 2 out of 10 models found a significant relationship between higher MASQ AA scores and lower learning rates, and in these models the relationship was only consistently observed at the largest set size (Fig 6).

These RL-only model analyses do not provide an exhaustive nor necessarily direct comparison with the RLWM models considered. In particular, we note that the winning RLWM model reflected total negative feedback neglect for RL and parameterized negative feedback neglect for WM, making it different from any possible characterization of RL-only models which can only include a negative feedback parameter for RL (their only module). We also note that this analysis may not be directly comparable to other RL-only studies in the field, both given the structural difference of the current task's multiple set sizes across blocks when compared with typical static set size RL tasks, and in light of the purely deterministic (accurate) feedback used in the current study when compared with the (often) stochastic feedback structure of many RL

tasks [54,55]. Nonetheless, the comparison of attribution of anxiety-related learning task deficits between RLWM and RL models as applied to the current dataset points to the possibility that failure to model WM explicitly might lead to misleading interpretation of anxiety-related deficits in learning.

An important area for future study lies in the investigation of potential interactions between individual differences in the effectiveness of working memory system contributions to learning and the specific demands placed on working memory by RL tasks of varying complexity and structure. This would valuably be extended to work with other clinical populations; for instance, individuals with ADHD are known to have anxiety levels well above the population average, and responses to medication in reversal learning performance have been shown to relate to WM capacity [56]. Extension of the work conducted here might help to illuminate the mechanisms underlying such findings.

It will also be important to conduct similar investigations for other dimensions of internalizing psychopathology. In the current study, supplementary exploratory analyses of self-report measures of depressive affect (CES-D and BDI-II) revealed no significant relationships between these measures and any of the computational model parameters (see Supplemental Analyses in [S1 Text](#)). We include these findings for completeness but note that a larger scale study is required to provide a fully powered investigation of additional dimensions of psychopathology, including but not limited to depressive affect.

Our current study does not evaluate how these results might translate into more complex real-world behavior outside of a lab environment, and one interesting direction for future research would be the extension of the paradigm to more naturalistic settings [57]. An additional future direction of interest would be to more explicitly study shifts over time between WM and RL systems.

In summary, the study findings provide insight into the potentially complicated relationship of anxiety with interactive systems of WM and learning. By revealing the emergent learning rate effects of physiological anxiety when working memory is jointly modeled with the learning process, and pointing to the reduced effectiveness of working memory's contributions to learning due to quicker rates of forgetting in individuals with high levels of physiological anxiety, the current findings provides new insights into the problems anxious individuals have with learning while also highlighting the need, across the computational RL literature, to explicitly model cognitive processes such as, but not necessarily limited to, working memory when interpreting the behavior of healthy participants or investigating psychopathology-related alterations in task performance.

Methods

Participants

The study was conducted online with participants recruited via the UC Berkeley Research Participation Pool (RPP), which offers partial course credit to undergraduate students for participation in human subjects research. All participants completed online informed consent prior to participation. The study was approved by the UC Berkeley Committee for the Protection of Human Subjects (CPHS). An initial total of $n = 229$ students [143 = female and 86 = male; mean age = 21.2 ± 2.4] participated prior to exclusions.

Exclusions

Participants were eligible to participate in the online study if they confirmed that they were not currently taking antidepressant or anxiolytic medications and had not used cannabis within the preceding two weeks. Two attention checks were embedded in the self-report questionnaires (e.g., "Select 2 here to show that you are paying attention") as data validity checks. Additionally, following [58], once participants had completed the task they were informed that their participation credit was now guaranteed, and were asked to answer two questions honestly to ensure that the research would only use credible data: participants were asked to again confirm whether they had used cannabis in the last two weeks, and were

asked whether they felt they had paid sufficient attention during the task that their data should be used in our study. 45 participants were excluded based on responses to the end-of-task questions, and 6 additional participants were excluded for missing two or more attention checks during the questionnaires.

Participants were also excluded if they had greater than two standard deviations (2SD) above the mean number of trial timeouts (>11.9% of trials; 6 participants excluded). Additionally, we calculated mean set size = 2 asymptotic performance across the last 5 same-stimulus presentations for set size = 2 blocks; participants with mean accuracy lower than 2SD below the group mean (<64% accuracy; 8 participants) were excluded from further analysis. Pre-exclusion population distributions for number of timeouts and asymptotic performance on set size = 2 were highly skewed (see Fig G in [S1 Text](#)). Recent work exploring the effect of exclusions in highly skewed data on individual differences analyses has cautioned that overly aggressive exclusion criteria may introduce “shadow biases” into individual differences work by excluding participants disproportionately with respect to metrics of interest [59]. Simulation work examining various exclusion criteria approaches for reaction time distributions, which are also highly skewed, showed that a 2SD cutoff was among the least biased methods for tail exclusions in these distributions [60].

The number of participants included in the final analysis was $n = 164$ (104 = female; mean age 21.20 ± 2.43). 34 participants self-reported Hispanic or Latino ethnicity. Self-identified race distribution of participants was as follows: 94 Asian, 8 Black or African American, 34 White, 2 Native American or Pacific Islander, 12 who identified as “More than one race” and 14 who identified as “Unknown” race.

On a within-participant level, trials that timed out were excluded (mean number of timeout trials = 4.74), and subsequently any trial blocks for which a participant had fewer than 9 presentations of any stimulus remaining in a block after exclusions were also excluded (1 block excluded for each of 5 participants).

Experimental paradigm

The behavioral task used was a variant of the classic RLWM task (Collins & Frank 2012). The main task comprised a learning (or “training”) phase, followed by an unrelated distractor task, and finally a surprise testing phase of the original task stimuli. During the learning phase, participants were presented with a series of stimuli (images) on screen, with one stimulus shown for each trial. There were three possible correct key presses associated with each stimulus; keyboard keys ‘j’, ‘k’, or ‘l’. Each stimulus within a block was associated with only one correct key press (but multiple stimuli within a block could be associated with the same correct key press). Participants had 1.5s to select an action for each stimulus presented; if no response was selected in the allowed time, a message “You did not make a selection in time!” was displayed for 500ms, and the task advanced to the next trial. If participants responded with a key press, the stimulus was removed from the screen and they were given accurate feedback (a green +1 for correct responses, or a red 0 for incorrect responses) presented for 500ms. A fixation cross was shown for 500ms between each learning trial with a 100ms blank screen immediately before and after the fixation cross. See [Fig 1](#).

The learning task consisted of 13 blocks of trials. Within each block, participants learned the correct key press action for each stimulus in the block through multiple trials and feedback. Each trial block included a stimulus set size (n_S) varying between 2, 3, 4, 5, or 6 stimuli. Stimuli for each block were drawn from a randomly selected image category (such as ‘nature’, ‘shapes’, ‘musical instruments’, etc.) without replacement, such that each set of unique stimuli within a block were of the same category and no stimuli or categories were repeated across blocks.

Participants were randomly assigned to one of 10 pre-generated learning task trial sequences. In each generated trial sequence, participants started and ended with a block of set size = 2 to mitigate the possible conflation of primacy and recency effects with working memory. Block 2 was always of set size = 3; the remaining intermediate blocks were shuffled so that each set size was presented once in blocks 3–7 and once in blocks 8–12. Each stimulus for a given block was shown 13 times within that block, so the total number of trials per block varied with stimulus set size. Stimuli sequences

within each block were generated pseudo-randomly controlling for a uniform distribution of delay between each two successive presentations of the same stimulus across every $[1:2 \cdot nS]$ trials per block.

Following the learning phase, participants were given an optional break of up to one minute, followed by a distractor task. The distractor task used in the current experiment was a short n-back task in which participants were shown a series of images (shapes which were not used in the learning task) and asked to press the left arrow key of the keyboard if the image on screen was the same as the image that had appeared “n” stimuli ago. Participants completed a 1-back, 2-back, and 3-back task (mean completion time = 18.8 minutes). Following the n-back task, participants were given an optional 30 second break.

Following this break, participants were informed that they would be tested on their learning of the actions associated with the stimuli they had seen during the earlier learning portion of the experiment. Participants were shown each stimulus from their learning phase a total of three times during the testing phase. In each testing trial, the stimulus appeared on screen for 1.5s, and the participant selected ‘j’, ‘k’, or ‘l’ based on their earlier learning for the stimulus. Importantly, participants were not given feedback on accuracy in the testing phase, so additional learning could not occur. If participants did not select an action quickly enough, a message “You did not make a selection in time!” was displayed for 500ms. A fixation cross was shown for 500ms between each testing trial with a 100ms blank screen immediately before and after the fixation cross.

Self-report measures of trait anxiety and depression

Prior to the behavioral task and following informed consent, participants were asked to complete a short series of questionnaires designed to measure levels of anxious and depressed symptomatology. The questionnaires used in this study comprised the following: the State-Trait Anxiety Inventory trait subscale (STAI-T; [61]); the Penn State Worry Questionnaire (PSWQ; 44); the Beck Depression Inventory (BDI-II; 46,47); and the Mood and Anxiety Symptom Questionnaire anxious arousal subscale (MASQ AA; 43). Items addressing suicidality were excluded. Final number of items included was $n = 126$.

Scores on the MASQ AA subscale had a mean of 27.25 (standard deviation = 10.86), with a minimum participant score of 17 (minimum possible scale score = 17) and a maximum participant score of 68 (maximum possible scale score = 85). Scores on the worry measure PSWQ had a mean of 52.12 (standard deviation = 13.81), with a minimum participant score of 16 (minimum possible scale score = 16) and a maximum participant score of 79 (maximum possible scale score = 80). These subscale scores differentiate between anxiety as characterized by somatic symptoms versus cognitive symptoms and were used in primary hypothesis testing (see Results). Complete score distributions for the participant group for each questionnaire measure, including the BDI and STAI-T scales used in exploratory post-hoc analyses, are shown in Fig A in [S1 Text](#).

Behavioral analysis

Previous investigations of learning under various set sizes have found a robust and highly replicable effect of set size on learning over time [20,22,36]. To confirm that our data replicated these established results, we first examined the mean learning curves across participants at each set size, measured as mean percent response accuracy by stimulus presentation. Additionally, we calculated overall learning accuracy as the mean accuracy across all presentations of all stimuli for each set size, and asymptotic learning accuracy as the mean of the last 5 presentations of each stimulus across blocks with the same set size. Overall testing accuracy was calculated as the mean accuracy across all stimuli presentations (each image was shown 3 times during testing) for each set size.

To quantify the influence of increases in set size on participant performance, we calculated a set size slope for each participant (following [36]) according to the following equation:

$$-1.5 \cdot perf_{setsize6} - 0.5 \cdot perf_{setsize5} + 0.5 \cdot perf_{setsize3} + 1.5 \cdot perf_{setsize2} \quad (1)$$

Where underlying data followed a normal distribution, we used Pearson correlations and t-tests as specified throughout the text. We used Spearman rank correlation, Wilcoxon, Kruskal-Wallis and Mann-Whitney U testing as specified throughout the text to test various correlations and distributional differences. Statistical tests were performed in Python and R.

Computational modeling

We modeled learning and testing performance using a series of dual-module reinforcement learning and working memory (RLWM) models of varying levels of complexity. This class of models has been shown to effectively tease apart the effects of working memory from those of reinforcement learning in RLWM tasks with a set size manipulation as described above [20–22,36]. Model variants were tested individually and in combination; see Model key for details. Model variants were primarily motivated by findings from previous literature and include one novel variant (which tests separate parameters for WM weight in set sizes above versus below WM capacity) as described below.

Baseline model: Reinforcement learning and working memory (RLWM). Our baseline model was a two-module reinforcement learning (RL) and working memory (WM) model in which separate RL and WM processes contribute collaboratively to learning and decision-making [20,49]. We note that while some previous studies investigating learning-phase data from the RLWM task have used models in which the RL and WM modules operate independently (e.g., [22]), these models are not capable of capturing participant testing phase performance (see Fig 4 of the current paper) and so are not considered here.

The WM module tracks weights for each possible action (a) given each stimulus (s) per block. Working memory weights (denoted as $W(s, a)$) are initialized to random chance ($1/nA$, where $nA=3$ is the number of possible actions per state) and are updated after each trial based on feedback assuming perfect information retention. Although working memory has high information retention in the short term, the stored weights are assumed to decay rapidly at a parameterized rate of φ between each trial update in order to reflect the short-term nature of WM.

The weight for the current action/stimulus (s_t, a_t) for each trial is updated based on a prediction error (equation 1 below) from the observed trial feedback r_t (1 = correct, 0 = incorrect) with complete retention (learning rate $\alpha_{WM} = 1$) such that:

$$\delta_t = r_t - W_t(s_t, a_t) \quad (2)$$

$$W_{t+1}(s_t, a_t) = W_t(s_t, a_t) + \alpha_{WM} \cdot \delta_t \quad (3)$$

WM weights for all stimuli and actions (s, a) decay back toward initial values after each trial according to:

$$W_t(s, a) = W_t(s, a) + \varphi \cdot ((1/nA) - W_t(s, a)) \quad (4)$$

The contribution of working memory to action selection in each block is weighted according to a prior WM confidence weight parameter $\rho \in [0,1]$ and WM capacity parameter $K \in [2,6]$ compared to the number of stimuli in the set size (ns) for the current block.

$$w_{WMblock} = \rho \cdot \min(1, K/ns) \quad (5)$$

Meanwhile, the reinforcement learning (RL) module tracks action values (denoted $Q_{RL}(s, a)$), also initialized to $1/nA$. Values are updated for the current action/stimulus at each trial using a cooperative approach whereby both RL values and WM weights contribute, with a reward prediction error (rpe) defined as follows:

$$rpe_t = r_t - [(w_{WMblock} \cdot i \cdot W(s, a)_t) + ((1 - w_{WMblock} \cdot i) \cdot Q_{RL}(s, a)_t)] \quad (6)$$

and parameter $i \in [0, 1]$ controls the strength of the information sharing between RL and WM modules, such that WM knowledge contributes to the expectation portion of the prediction error.

Values for RL are then updated using a learning rate parameter $\alpha_{RL} \in [0, 1]$ governing the rate at which feedback is incorporated into the estimate.

$$Q_{RL}(s, a)_{t+1} = Q_{RL}(s, a)_t + \alpha_{RL} \cdot rpe_t \quad (7)$$

To model the probability that an agent selects a given action at a given time, we use a weighted mixture softmax choice policy to evaluate the relative value of each action during the decision process. The probability of selecting action a following stimulus s is denoted as follows:

$$p(a|s) = (1 - w_{WM}) \cdot \frac{\exp(\beta \cdot Q_{RL}(s, a))}{\sum_i \exp(\beta \cdot Q_{RL}(s, a_i))} + w_{WM} \cdot \frac{\exp(\beta \cdot W(s, a))}{\sum_i \exp(\beta \cdot W(s, a_i))} \quad (8)$$

where the parameter $\beta \in [0, 100]$, referred to as the softmax inverse temperature, controls the extent to which relative action values (as opposed to stochastic choice noise) are used in decision choice. During learning, the softmax inverse temperature was set to a fixed value of $\beta = 50$ to improve parameter reliability [20] and choice noise was captured via an undirected noise parameter $\varepsilon \in [0, 1]$, such that:

$$p(a|s)_{noisy} = \frac{\varepsilon}{nA} + (1 - \varepsilon) \cdot p(a|s) \quad (9)$$

Modeling of choice selection during the testing phase used the final learning phase Q_{RL} values for the test stimuli, and assumed an epsilon-noisy softmax action selection with no contribution of working memory to choice. The parameter $\beta_{test} \in [0, 100]$ was fit in the testing phase softmax action selection policy (reflecting potential weakening of RL-values' influence on test), while ε was fit jointly between the learning and testing phases (reflecting a shared tendency for lapses or inattention).

$$p(a|s)_{test} = \frac{\varepsilon}{nA} + (1 - \varepsilon) \cdot \frac{\exp(\beta_{test} \cdot Q_{RL}(s, a))}{\sum_i \exp(\beta_{test} \cdot Q_{RL}(s, a_i))} \quad (10)$$

Negative feedback neglect variant (“_bias”). Previous studies [20] have found that individuals often update action values more slowly following negative feedback than following positive feedback during learning. In *_bias* variant models, a parameter $\eta \in [0, 1]$ is introduced which reduces the learning rate for both RL and WM modules following negative feedback.

We parameterize this by reducing the learning rate in both RL and WM modules following negative feedback according to the rule:

$$\left. \begin{array}{l} \alpha_{RL} = \alpha_{RL_base} \\ \alpha_{WM} = 1 \end{array} \right\} \quad \text{for } r = 1 \quad (11a,b)$$

$$\left. \begin{array}{l} \alpha_{RL} = \alpha_{RL_base} * (1 - \eta) \\ \alpha_{WM} = 1 * (1 - \eta) \end{array} \right\} \quad \text{for } r = 0 \quad (12a,b)$$

Asymmetric negative feedback neglect variant (“_asymbias”). Previous studies have shown that learning in smaller set sizes is more influenced by negative feedback than learning in larger set sizes [62]. In the asymmetric negative feedback neglect model variant, we tested the hypothesis that individuals entirely neglect negative feedback in the RL module while still allowing for biased neglect in the WM module. In this model variant the learning rates after unrewarded trials are not linked between RL and WM modules.

We parameterize this by reducing the learning rate asymmetrically for RL and WM modules following negative feedback according to the rule:

$$\left. \begin{array}{l} \alpha_{RL} = \alpha_{RL_base} \\ \alpha_{WM} = 1 \end{array} \right\} \quad \text{for } r = 1 \quad (13a,b)$$

$$\left. \begin{array}{l} \alpha_{RL} = 0 \\ \alpha_{WM} = 1 * (1 - \eta) \end{array} \right\} \quad \text{for } r = 0 \quad (14a,b)$$

Split WM confidence variant (“_2r”). In the split WM confidence model variant, we expanded the WM prior confidence parameter ρ into two parameters, ρ_{above_K} and ρ_{below_K} . This modification allowed the weight placed on working memory during choice selection to vary based on whether the current block set size exceeded an individual’s working memory capacity K .

The working memory weights for this model variant were then calculated according to:

$$w_{WMblock} = \rho_{below_K} \cdot \min\left(1, \frac{K}{ns}\right) = \rho_{below_K} \quad \text{for } ns < K \quad (15a)$$

$$w_{WMblock} = \rho_{above_K} \cdot \min\left(1, \frac{K}{ns}\right) = \rho_{above_K} \cdot \frac{K}{ns} \quad \text{for } ns > K \quad (15b)$$

Choice kernel variant (“_ck”). When making choices in learning tasks, participants may have a propensity to repeat previous actions regardless of value (a phenomenon sometimes referred to as perseveration or sticky choice). A choice kernel allows for multiple previous motor actions to influence the current choice, with a decaying influence of choices that are further in the past. In some (but not all) data sets, models which do not include action perseveration over multiple previous actions may induce a bias in parameter fitting, allocating variance inappropriately to asymmetric value updating following negative feedback [50,63].

We therefore tested a model variant which included the influence of past actions on current action choice during the learning phase of the task. The choice kernel (Ck) acts as a weighted trace of previous actions (i.e., pressing ‘J’, ‘K’, or ‘L’), and is initialized to 0 for each action.

Following each trial, the choice kernel is updated for the selected action and the stored action values decay at rate τ according to:

$$Ck_{t+1} = Ck_t + \tau \cdot (action - Ck_t) \quad (16a)$$

where *action* is an array of length 3 with value = 1 for the chosen action and value = 0 for each of the two unchosen actions on the current trial.

Each action decision is directly influenced by the choice kernel according to perseveration parameter κ :

$$p(a|s)_{ck} = (1 - \kappa) \cdot p(a|s)_{noisy} + \kappa \cdot Ck \quad (16b)$$

Model fitting procedure

Model parameters were fit using maximum likelihood estimation in Matlab with the function `fmincon` to minimize negative log likelihood. Individual maximum likelihood estimates were performed using 20 independent randomly selected starting points to improve selection of global maxima. All parameters were constrained to [0,1], except the softmax inverse temperature for the testing phase, which was scaled to be constrained on the interval [0,100], and the working memory capacity parameter K , which was constrained to the continuous interval [2,6].

Model comparison

We evaluated the fit of variants of the RLWM model with combinations of each modification described above. The final model space included 6 models as shown in [Table 1](#).

Model fit was assessed by comparing the Akaike information criterion (AIC; [64]) for each model in aggregate. The AIC measures overall model fit while penalizing for complexity, and we verify that it supports adequate model identification within the RLWM framework [36]. The winning model with the lowest total AIC was Model #5 (RLWM_asymbias_2r), with the following 9 parameters: RL learning rate, test phase softmax inverse temperature, epsilon undirected choice noise, WM confidence for low set sizes, WM confidence for high set sizes, WM forgetting, WM neglect of negative feedback, WM capacity, and an interaction parameter for cooperative updating between RL and WM systems.

There were substantial individual differences in best fit model, with Model #4 best fitting a greater percentage of participants based on individual AIC (see [Fig 7](#)). Model #4 was nested within the model with the overall lowest AIC (Model #5), indicating that the additional parameters in Model #5 were not necessary to fit some participants but were necessary to best fit others. Of note is that, because some models are nested, the more complex model could yield the exact same maximum likelihood as the simpler model for individuals who did not require the additional parameters by finding null-effect parameter values as optima for those individual parameters, so any improvement in AIC between Models #4 and #5 for a given individual would be due only to the penalty for modeling more parameters than necessary for that individual in the more complex model.

We compared individual differences in model best fit for high versus low anxiety participants using a median split on each of PSWQ and MASQ AA score, which illustrated that individual differences across model fit were qualitatively related to scores on metrics of anxiety (see [Fig 7](#)). Given that Model #5 showed the overall lowest AIC, best fit a subset of individuals, and contained Model #4 as a nested component, we proceeded with Model #5 as the winning model for purposes of hypothesis testing. We performed post-hoc confirmatory analysis of the significant findings in the analogous parameters in Model #4 to verify that results were robust across the two models.

Model validation

Model validation was performed by simulating data from a generative version of the winning computational model (Model #5, above). Learning curves by set size, difference in performance between low and high set size during learning, and difference in performance between low and high set size during testing were replicated by the model (see [Fig 4](#)).

Parameter recovery was performed for the winning model to test the identifiability of all model parameters. Participant data was simulated using the fit parameter values from the final model, and this simulated data with known underlying parameters was then fit using the optimization process described above. Recovered (fit) parameter values were then compared with the generative (known) parameter values. Parameters generally recovered well, with both significant

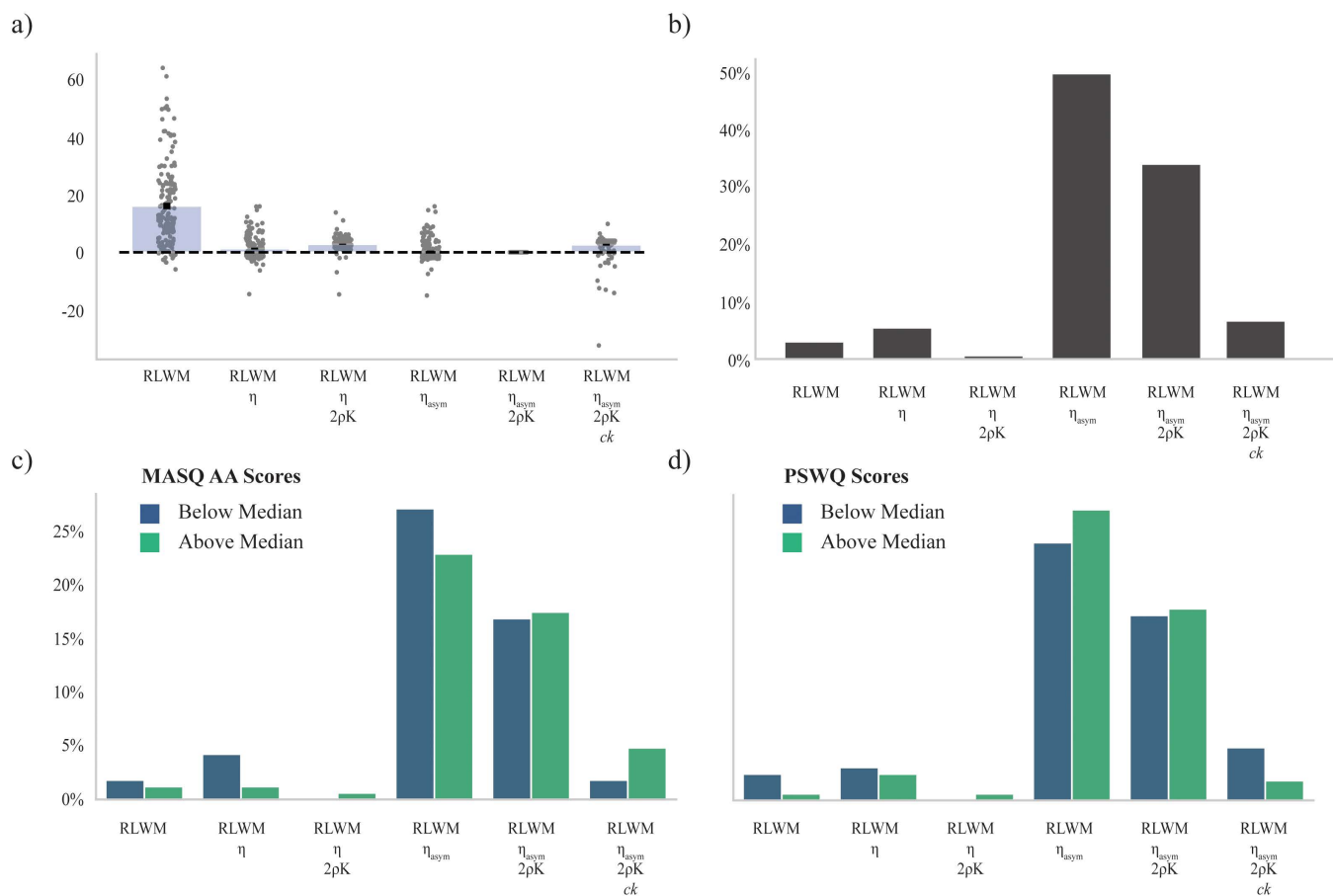


Fig 7. Model fit comparison using Akaike Information Criterion (AIC). a) Distribution of individual differences between model fit (AIC) for each model relative to individual AIC for the group winning model (Model #5). b) Percentage of overall participants best fit by each model. c) Percentage of participants best fit by each model by median split on MASQ AA subscale scores. d) Percentage of participants best fit by each model by median split on PSWQ scores.

<https://doi.org/10.1371/journal.pcbi.1012872.g007>

parameters (learning rate and working memory decay) having correlations between generative and fit parameters greater than or equal to 0.80. See Fig C in [S1 Text](#) for parameter recovery analysis for each model parameter.

Model identifiability was confirmed via a model-recovery analysis. Three sets of participant data, representing $n = 492$ simulated participants, were simulated from each of the following models: Model #2 (RLWM_bias), Model #4 (RLWM_asymbias), and the best-fit Model #5 (RLWM_asymbias_2r). Each simulated participant was then fit by each of the three selected models using AIC, and the resulting best-fit model was compared to the generative model for that data. We performed an additional model recovery analysis between the winning RLWM model from the main analysis (Model #5) and the two winning RL-only model variants (the winning single learning rate RL-only model and the winning 5 learning rate RL-only model) using the same procedure described above. Models accurately recovered their generated data using AIC; see Fig D in [S1 Text](#) for model recovery analysis. Previous research has shown that Bayesian model selection criteria such as the Bayesian Information Criteria (BIC) tend to over-penalize models in the RLWM class [22]. To confirm this in the current data and support our use of AIC as a measure of model fit, we performed a parallel model recovery analysis for the selected RLWM models using BIC. The confusion matrix for this analysis, shown in Fig D of [S1 Text](#) panel (c),

confirms that data generated from more complex underlying processes tends to be (incorrectly) best-fit by simpler models when BIC is used.

Model-based hypothesis testing (main analysis)

We performed initial hypothesis testing for dimensions of anxiety characterized by physiological arousal (MASQ AA scores) and cognitive anxiety/worry (PSWQ). We tested participant scores on each measure against one RL-specific parameter (learning rate) and four WM-related parameters (forgetting, WM confidence at low set sizes, WM confidence at high set sizes, and neglect of negative feedback in WM) from the winning model. We performed nonparametric correlation analyses due to the non-normality of the underlying data. We report p-values corrected using the Bonferroni statistical correction for 10 simultaneous comparisons (2 trait scores across 5 parameters each) to account for potential family-wise error across multiple tests (and additionally report uncorrected p-values for comparison where specified).

Additional modeling analysis (RL-only models)

For comparison and illustration (see Results), we fit 8 additional RL-only models which did not include WM modules to our data. Since the structure of our task included 5 different set sizes of stimuli across blocks, we considered one class of RL-only models with a single learning rate across set sizes (RL_α models) and one class of RL-only models with a learning rate for each set size (RL_5α models).

Four RL-only models were fit for each class; a base variant, and 3 additional variants which incrementally (cumulatively) added parameters describing the following mechanisms: stickiness of choice (s), RL forgetting (φ_{RL}), and negative feedback neglect (η_{RL}), formulated as outlined in the equations below.

The basic formulation of the RL-only models was directly analogous to the basic RLWM model (Model #1) with the WM components removed. For the RL_α base model, the reinforcement learning (RL) module tracks action values (denoted $Q_{RL}(s, a)$), initialized to $1/nA$. Values are updated for the current action/stimulus at each trial according to a reward prediction error (rpe) defined as:

$$rpe_t = r_t - Q_{RL}(s, a)_t \quad (17)$$

Values are then updated using a learning rate parameter $\alpha_{RL} \in [0, 1]$ governing the rate at which feedback is incorporated into the estimate.

$$Q_{RL}(s, a)_{t+1} = Q_{RL}(s, a)_t + \alpha_{RL} \cdot rpe_t \quad (18)$$

To model the probability that an agent selects a given action at a given time, we use a softmax choice policy to evaluate the relative value of each action during the decision process. The probability of selecting action a following stimulus s is denoted as follows:

$$p(a|s) = \frac{\exp(\beta \cdot Q_{RL}(s, a))}{\sum_i \exp(\beta \cdot Q_{RL}(s, a_i))} \quad (19)$$

where the parameter $\beta \in [0, 100]$, referred to as the softmax inverse temperature, controls the extent to which relative action values (as opposed to stochastic choice noise) are used in decision choice. During learning, the softmax inverse temperature was set to a fixed value of $\beta = 50$ to improve parameter reliability [20] and choice noise was captured via an undirected noise parameter $\epsilon \in [0, 1]$, such that:

$$p(a|s)_{noisy} = \frac{\varepsilon}{nA} + (1 - \varepsilon) \cdot p(a|s) \quad (20)$$

Modeling of choice selection during the testing phase used the final learning phase Q_{RL} values for the test stimuli and assumed an epsilon-noisy softmax action selection. The parameter $\beta_{test} \in [0, 100]$ was fit in the testing phase softmax action selection policy (reflecting potential weakening of RL-values' influence on test), while ε was fit jointly between the learning and testing phases (reflecting a shared tendency for lapses or inattention).

$$p(a|s)_{test} = \frac{\varepsilon}{nA} + (1 - \varepsilon) \cdot \frac{\exp(\beta_{test} \cdot Q_{RL}(s, a))}{\sum_i \exp(\beta_{test} \cdot Q_{RL}(s, a_i))} \quad (21)$$

The RL_5 α base model followed the same equations as the RL_ α base model shown above, but included 5 separate learning rate parameters $\alpha_{ss2}, \alpha_{ss3}, \alpha_{ss4}, \alpha_{ss5}, \alpha_{ss6} \in [0, 1]$ which applied only to blocks of the appropriate set size.

In addition to the base models, 3 additional variants were included for each RL-only model class.

Choice kernel variant. First, a choice kernel was included to reflect the decaying influence of previous choices on current choice during the learning phase of the task. The choice kernel (Ck) acts as a weighted trace of previous actions (i.e., pressing 'J', 'K', or 'L'), and is initialized to 0 for each action.

Following each trial, the choice kernel is updated for the selected action and the stored action values decay at rate τ according to:

$$Ck_{t+1} = Ck_t + \tau \cdot (action - Ck_t) \quad (22a)$$

where *action* is an array of length 3 with value = 1 for the chosen action and value = 0 for each of the two unchosen actions on the current trial.

Each action decision is directly influenced by the choice kernel according to perseveration parameter κ :

$$p(a|s)_{ck} = (1 - \kappa) \cdot p(a|s)_{noisy} + \kappa \cdot Ck \quad (22b)$$

Forgetting variant. Next, a forgetting parameter (denoted as φ_{RL} to distinguish it from the working memory decay parameter φ in the RLWM models) was added to the model, which parameterized decay of RL values back to their initial (random) values according to the following:

$$Q_t(s, a) = Q_t(s, a) + \varphi_{RL} \cdot ((1/nA) - Q_t(s, a)) \quad (23)$$

Negative feedback neglect variant. Finally, a parameter for negative feedback neglect (η_{RL}) was included to allow for reduced incorporation of non-rewarded trials according to:

$$\alpha_{RL} = \alpha_{RL} \quad \text{for } r = 1 \quad (24a)$$

$$\alpha_{RL} = \alpha_{RL} * (1 - \eta) \quad \text{for } r = 0 \quad (24b)$$

Parameter correlation methods for additional analysis

We investigated the relationship of scores on the MASQ AA with each parameter of these 8 RL-only model variants to illustrate the potential attribution of variance by these models relating to the effects of interest from the main analysis. We

used Spearman correlations corrected for FWE across all comparisons (MASQ AA x number of parameters) within each model and compared these results to the findings from the winning RLWM model (Model #5). We note that the significant effects from the winning RLWM model (Model #5) do not change in significance whether corrected for FWE across 10 comparisons (as in the main analysis) or across 9 comparisons (which would reflect correction across MASQ AA for all 9 of the model parameters). See [Fig 6](#) for comparative results of additional analysis.

Supporting information

S1 Text. Supplemental Figs A–G and supplemental analyses.
(PDF)

Author contributions

Conceptualization: Jennifer D. Senta, Anne G.E. Collins.

Data curation: Jennifer D. Senta.

Formal analysis: Jennifer D. Senta.

Investigation: Jennifer D. Senta.

Methodology: Jennifer D. Senta, Sonia J. Bishop, Anne G.E. Collins.

Project administration: Jennifer D. Senta.

Supervision: Sonia J. Bishop, Anne G.E. Collins.

Validation: Jennifer D. Senta.

Visualization: Jennifer D. Senta.

Writing – original draft: Jennifer D. Senta.

Writing – review & editing: Jennifer D. Senta, Sonia J. Bishop, Anne G.E. Collins.

References

1. LaFreniere LS, Newman MG. Probabilistic Learning by Positive and Negative Reinforcement in Generalized Anxiety Disorder. *Clin Psychol Sci*. 2019;7(3):502–15. <https://doi.org/10.1177/2167702618809366> PMID: [31448183](#)
2. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, Mass: MIT Press; 1998. p. 322. (Adaptive computation and machine learning).
3. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012;16(1):72–80. <https://doi.org/10.1016/j.tics.2011.11.018> PMID: [22177032](#)
4. Ting C-C, Palminteri S, Lebreton M, Engelmann JB. The elusive effects of incidental anxiety on reinforcement-learning. *J Exp Psychol Learn Mem Cogn*. 2022;48(5):619–42. <https://doi.org/10.1037/xlm0001033> PMID: [34516205](#)
5. Gagne C, Zika O, Dayan P, Bishop SJ. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *Elife*. 2020;9:e61387. <https://doi.org/10.7554/eLife.61387> PMID: [33350387](#)
6. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND. Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(52):20941–6.
7. Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA. From fear to safety and back: reversal of fear in the human brain. *J Neurosci Off J Soc Neurosci*. 2008;28(45):11517–25.
8. Wise T, Dolan RJ. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nat Commun*. 2020;11(1):4179. <https://doi.org/10.1038/s41467-020-17977-w> PMID: [32826918](#)
9. Fan H, Gershman SJ, Phelps EA. Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty. *Nat Hum Behav*. 2023;7(1):102–13. <https://doi.org/10.1038/s41562-022-01455-y> PMID: [36192493](#)
10. Raio CM, Hartley CA, Oederu TA, Li J, Phelps EA. Stress attenuates the flexible updating of aversive value. *Proc Natl Acad Sci U S A*. 2017;114(42):11241–6. <https://doi.org/10.1073/pnas.1702565114> PMID: [28973957](#)

11. Bijsterbosch J, Smith S, Bishop SJ. Functional Connectivity under Anticipation of Shock: Correlates of Trait Anxious Affect versus Induced Anxiety. *J Cogn Neurosci*. 2015;27(9):1840–53.
12. Bruckner R, Heekeren HR, Nassar MR. Understanding learning through uncertainty and bias. *Commun Psychol*. 2025;3(1):1–13.
13. Jepma M, Schaaf JV, Visser I, Huizenga HM. Uncertainty-driven regulation of learning and exploration in adolescents: A computational account. *PLoS Comput Biol*. 2020;16(9):e1008276. <https://doi.org/10.1371/journal.pcbi.1008276> PMID: [32997659](#)
14. Lee JK, Rouault M, Wyart V. Adaptive tuning of human learning and choice variability to unexpected uncertainty. *Sci Adv*. 2023;9(13):eadd0501. <https://doi.org/10.1126/sciadv.add0501> PMID: [36989365](#)
15. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. *Nat Neurosci*. 2007;10(9):1214–21. <https://doi.org/10.1038/nn1954> PMID: [17676057](#)
16. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Associative learning of social value. *Nature*. 2008;456(7219):245–9. <https://doi.org/10.1038/nature07538> PMID: [19005555](#)
17. Blain B, Rutledge RB. Momentary subjective well-being depends on learning and not reward. *Elife*. 2020;9:e57977. <https://doi.org/10.7554/eLife.57977> PMID: [33200989](#)
18. Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci*. 2015;18(4):590–6. <https://doi.org/10.1038/nn.3961> PMID: [25730669](#)
19. Xia L, Xu P, Yang Z, Gu R, Zhang D. Impaired probabilistic reversal learning in anxiety: Evidence from behavioral and ERP findings. *Neuroimage Clin*. 2021;31:102751. <https://doi.org/10.1016/j.nicl.2021.102751> PMID: [34242887](#)
20. Collins AGE. The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *J Cogn Neurosci*. 2018;30(10):1422–32. https://doi.org/10.1162/jocn_a_01238 PMID: [29346018](#)
21. Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ. Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. *J Neurosci*. 2014;34(41):13747–56.
22. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur J Neurosci*. 2012;35(7):1024–35. <https://doi.org/10.1111/j.1460-9568.2011.07980.x> PMID: [22487033](#)
23. Frank MJ, Doll BB, Oas-Terpstra J, Moreno F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci*. 2009;12(8):1062–8. <https://doi.org/10.1038/nn.2342> PMID: [19620978](#)
24. Frank MJ, Badre D. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cereb Cortex*. 2012;22(3).
25. Courtney SM, Ungerleider LG, Keil K, Haxby JV. Transient and sustained activity in a distributed neural system for human working memory. *Nature*. 1997;386(6625):608–11. <https://doi.org/10.1038/386608a0> PMID: [9121584](#)
26. Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*. 1989;61(2):331–49. <https://doi.org/10.1152/jn.1989.61.2.331> PMID: [2918358](#)
27. Fuster JM, Alexander GE. Neuron activity related to short-term memory. *Science*. 1971;173(3997):652–4. <https://doi.org/10.1126/science.173.3997.652> PMID: [4998337](#)
28. O'Reilly RC, Frank MJ. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput*. 2006;18(2):283–328. <https://doi.org/10.1162/089976606775093909> PMID: [16378516](#)
29. Rmuis M, McDougall SD, Collins AG. The role of executive function in shaping reinforcement learning. *Curr Opin Behav Sci*. 2021;38:66–73.
30. Yoo AH, Collins AGE. How working memory and reinforcement learning are intertwined: A cognitive, neural, and computational perspective. *J Cogn Neurosci*. 2022;34(4).
31. Jocham G, Klein TA, Ullsperger M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci*. 2011;31(5):1606–13.
32. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*. 2007;104(41):16311–6.
33. Wimmer GE, Li JK, Gorgolewski KJ, Poldrack RA. Reward learning over weeks versus minutes increases the neural representation of value in the human brain. *J Neurosci*. 2018;38(35):7649–66.
34. Collins AGE, Ciullo B, Frank MJ, Badre D. Working Memory Load Strengthens Reward Prediction Errors. *J Neurosci*. 2017;37(16):4332–42.
35. Collins AGE, Frank MJ. Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proc Natl Acad Sci*. 2018;115(10):2502–7.
36. Master SL, Eckstein MK, Gottlieb N, Dahl R, Wilbrecht L, Collins AGE. Distangling the systems contributing to changes in learning during adolescence [Internet]. *Neuroscience*; 2019 Apr [cited 2023 Aug 14]. Available from: <http://biorxiv.org/lookup/doi/10.1101/622860>
37. Rmuis M, He M, Baribault B, Walsh EG, Festa EK, Collins AG. Age-related differences in prefrontal glutamate are associated with increased working memory decay that gives the appearance of learning deficits. *eLife*. 2023;12:e85243.
38. Bishop S, Duncan J, Brett M, Lawrence AD. Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nat Neurosci*. 2004;7(2):184–8. <https://doi.org/10.1038/nn1173> PMID: [14703573](#)

39. Forster S, Nunez Elizalde AO, Castle E, Bishop SJ. Unraveling the Anxious Mind: Anxiety, Worry, and Frontal Engagement in Sustained Attention Versus Off-Task Processing. *Cereb Cortex*. 2015;25(3):609–18.
40. Moran TP. Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychol Bull*. 2016;142(8):831–64. <https://doi.org/10.1037/bul0000051> PMID: 26963369
41. Balderston NL, Vytal KE, O'Connell K, Torrisi S, Letkiewicz A, Ernst M. Anxiety patients show reduced working memory related dlPFC activation during safety and threat: research article: anxiety patients show reduced dlPFC activity. *Depress Anxiety*. 2017;34(1):25–36.
42. Glasgow S, Imbriano G, Jin J, Zhang X, Mohanty A. Threat and uncertainty in the face of perceptual decision-making in anxiety. *J Psychopathol Clin Sci*. 2022;131(3):265–77. <https://doi.org/10.1037/abn0000729> PMID: 35357845
43. Watson D, Clark LA. The Mood and Anxiety Symptom Questionnaire. Iowa City: Univ Iowa Dep Psychol; 1991.
44. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther*. 1990;28(6):487–95. [https://doi.org/10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6) PMID: 2076086
45. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*. 1977;1(3):385–401.
46. Beck AT. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4(6):561.
47. Beck AT, Steer RA, Brown G. Beck Depression Inventory–II [Internet]. 1996 [cited 2025 Feb 3]. Available from: <https://doi.apa.org/doi/10.1037/t00742-000>
48. Zou AR, Muñoz Lopez DE, Johnson SL, Collins AGE. Impulsivity Relates to Multi-Trial Choice Strategy in Probabilistic Reversal Learning. *Front Psychiatry*. 2022;13:800290. <https://doi.org/10.3389/fpsy.2022.800290>
49. Rac-Lubashevsky R, Cremer A, Collins AGE, Frank MJ, Schwabe L. Neural index of reinforcement learning predicts improved stimulus–response retention under high working memory load. *J Neurosci*. 2023;43(17):3131–43.
50. Sugawara M, Katahira K. Dissociation between asymmetric value updating and perseverance in human reinforcement learning. *Sci Rep*. 2021;11(1):3574. <https://doi.org/10.1038/s41598-020-80593-7> PMID: 33574424
51. Pittig A, Treanor M, LeBeau RT, Craske MG. The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neurosci Biobehav Rev*. 2018;88:117–40. <https://doi.org/10.1016/j.neubiorev.2018.03.015> PMID: 29550209
52. Bishop SJ, Gagne C. Anxiety, depression, and decision making: a computational perspective. *Annu Rev Neurosci*. 2018;41(1):371–88.
53. Farashahi S, Donahue CH, Hayden BY, Lee D, Soltani A. Flexible combination of reward information across primates. *Nat Hum Behav*. 2019;3(11):1215–24. <https://doi.org/10.1038/s41562-019-0714-3> PMID: 31501543
54. Eckstein MK, Master SL, Xia L, Dahl RE, Wilbrecht L, Collins AG. The interpretation of computational model parameters depends on the context. Hartley C, Behrens TE, Radulescu A, editors. *eLife*. 2022;11:e75474. <https://doi.org/10.7554/eLife.75474>
55. Eckstein MK, Wilbrecht L, Collins AG. What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. *Curr Opin Behav Sci*. 2021;41:128–37.
56. Clatworthy PL, Lewis SJG, Brichard L, Hong YT, Izquierdo D, Clark L. Dopamine Release in Dissociable Striatal Subregions Predicts the Different Effects of Oral Methylphenidate on Reversal Learning and Spatial Working Memory. *J Neurosci*. 2009;29(15):4690–6.
57. Wise T, Emery K, Radulescu A. Naturalistic reinforcement learning. *Trends in Cognitive Sciences*. 2024;28(2):144–58.
58. Meade AW, Craig SB. Identifying careless responses in survey data. *Psychol Methods*. 2012;17(3):437–55. <https://doi.org/10.1037/a0028085> PMID: 22506584
59. Siritzky EM, Cox PH, Nadler SM, Grady JN, Kravitz DJ, Mitroff SR. Standard experimental paradigm designs and data exclusion practices in cognitive psychology can inadvertently introduce systematic “shadow” biases in participant samples. *Cogn Res Princ Implic*. 2023;8(1):66.
60. Berger A, Kiefer M. Comparison of different response time outlier exclusion methods: a simulation study. *Front Psychol*. 2021;12:675558. <https://doi.org/10.3389/fpsyg.2021.675558>
61. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
62. Collins A. RL or not RL? Parsing the processes that support human reward-based learning. [Internet]. 2024 [cited 2024 Sep 12]. Available from: <https://osf.io/he3pm>
63. Katahira K. The statistical structures of reinforcement learning with asymmetric value updates. *J Math Psychol*. 2018;87:31–45.
64. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716–23.