Word count: 5700 (main text without Methods)
Tables: 2
Figures: 6

# Trait anxiety is associated with hidden state inference during aversive reversal learning

Ondrej Zika[1,2,3], Katja Wiech[3], Andrea Reinecke[4,5], Michael Browning[4,5] and Nicolas W. Schuck[1,2].

1 Max Planck Research Group NeuroCode, Max Planck Institute for Human Development, Berlin, Germany
2 Max Planck UCL Center for Computational Psychiatry and Aging Research, Berlin, Germany, and London, UK
3 Wellcome Center for Integrative Neuroimaging (WIN), Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
4 Department of Psychiatry, University of Oxford, Oxford, UK
5 Oxford Health NHS Trust, Warneford Hospital, Oxford, UK

Corresponding author: Ondrej Zika, zika@mpib-berlin.mpg.de, Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany

## Abstract

Updating beliefs in changing environments can be achieved either by gradually adapting expectations or by identifying a hidden structure composed of separate states, and inferring which state fits one's observations best. Previous studies have found that a state inference mechanism might be associated with relapse phenomena, such as return of fear, that commonly represent a major obstacle in clinical treatment of anxiety disorders. Here, we tested whether variability in trait anxiety among healthy individuals is associated with a tendency towards inferring a hidden structure of an aversive environment, as opposed to learning gradually from observations. In a Pavlovian probabilistic aversive learning paradigm, participants had to follow changes in cue-associated shock contingencies by providing probability ratings on each trial. In three sessions, the contingencies switched between high and a low levels of shock probability (60/40%, 75/25% or 90/10%). High trait anxiety was associated with steeper behavioral switches after contingency reversals, and more accurate probability ratings overall. To elucidate the computational mechanisms behind these behavioral patterns, we compared a "1-state" model, which reflects gradual updating, with a novel state-inference model ("n-state"). High trait anxiety was associated with improved fit of the state inference model (n-state) compared to the gradual model (1-state) in the session characterized by the largest shock contingency changes (90/10). This finding provides evidence that trait anxiety variations among health adults are associated with tendency to infer hidden causes that generate the observed aversive outcomes. This was particularly evident in environments with larger contingency changes and less outcome uncertainty. This association may contribute to relapse phenomena observed among high trait anxious individuals.

## Introduction

Aversive memories are notoriously difficult to forget, and often resist attempts to 'overwrite' them with new experiences. In exposure therapy, for instance, the feared situation or object is presented in the absence of an aversive outcome to achieve extinction of the fear response. While this procedure can lead to a decrease in fear responding, this reduction sometimes remains specific to the therapeutic context, and fails to generalize to the outside world (Craske et al., 2014; Vervliet et al., 2013). Such deficits in updating of aversive beliefs have been linked to anxiety disorders. Clinical anxiety has been associated with lowered discrimination between conditioned and unconditioned stimuli (Dibbets et al., 2015; Duits et al., 2015), decreased inhibition of responses to conditioned stimuli (Davis et al., 2000; Haaker et al., 2015) and heightened fear generalization (Dunsmoor & Paz, 2015). Some research has suggested that, even in healthy adults, heightened trait anxiety can lead to overly context specific unlearning, as indicated by lower success of cognitive behavioral therapy in high trait anxious individuals (Muris et al., 1998), suboptimal uncertainty adjustment of learning in volatile environments (Browning et al., 2015; Piray & Daw, 2021) and higher rates of fear relapse following treatment (Rodriguez et al., 1999; Staples-Bradley et al., 2018).

Here, we investigated the context-specificity of learning in aversive environments and its modulation by trait anxiety. Our main hypothesis was that trait anxiety is associated with a higher propensity to associate periods of relative safety and harm with distinct contexts, and we argue that this process might explain persistence and recurrence of unwanted experiences, akin to relapse phenomena observed in clinical populations.

We employed a probabilistic aversive learning paradigm where shock contingencies changed in semi-regular intervals. In such an environment, participants could either update their beliefs from trial-to-trial or they could infer that there were in fact two 'contexts' that corresponded to relative safety and threat. We reasoned that participants who are more prone to infer two distinct contexts switch beliefs about shock probabilities more abruptly when the actual shock rates change, and that they are less sensitive to oddball events. We tested our hypothesis by comparing the performance of two computational models that precisely captured the differences between gradual learning and context-dependent learning, and yielded detailed predictions about trial-to-trial changes in behavior. In addition to an effect of anxiety, we expected that decreased uncertainty in the learning environment will lead to more context-dependent learning. We therefore compared learning in three probabilistic environments which varied in the amount of outcome uncertainty.

One important aspect of our study was a focus on the process in which participants infer an unobservable (i.e., 'hidden') context, similar to the 'acquisition' and 'extinction' contexts described by Bouton, (2004). This differed from how the term 'context' is used to describe explicitly cued aspects of an experiment, such as different virtual rooms (e.g., Indovina et al., 2011). Our idea of a context relates to partially observable 'states' (also known as 'hidden states'), as commonly discussed in the reinforcement learning (RL) literature (Sutton & Barto, 2018). Previous theoretical (Wilson et al., 2014) and experimental (Bartolo & Averbeck, 2020; Schuck et al., 2016; Vaidya et al., 2021) work suggested that the brain uses such states to represent information and drive decisions (see Schuck et al., 2018, for a review). In the present paper, we used the term 'state' to refer to periods of high or low threat and distinguish between *objective states,* which refer to the objective high/low periods of threat, and *subjective states,* which are states inferred by the participant (e.g., participant may only notice that the probability of shock has dropped after several trials and so they will remain in a subjective high-threat state).

Learning in a state-dependent manner (Bouton, 2002; Gershman & Niv, 2012) stands in contrast to gradual learning as proposed by classical associative learning theories (e.g., Rescorla & Wagner, 1972). The key distinction between the two ways of learning is that, under gradual learning, the individual updates their expectation on a trial-by-trial basis, effectively overwriting their previous estimate with each update. On the other hand, an agent learning in a state-dependent manner creates classes of similar experiences (i.e., "states") and when predicting an outcome tries to infer which state they are currently in. Consequently, state-dependent learning often leads to abrupt jumps in an agent's predictions, called state switches, which reflect when a state change was detected. In addition, such a learning mechanism can also explain a persistence of previous experiences despite new learning, similar to relapse phenomena observed in clinical practice during which clinically extinguished fear fails to generalize to everyday life (Craske et al., 2014).

Experiments that aim to arbitrate between these perspectives typically use extinction learning designs, in which a neutral stimulus is first paired with an aversive unconditioned stimulus (US, typically a painful shock) during an acquisition phase, but is no longer followed by shocks during an extinction phase. The gradual learning perspective assumes that shock contingencies during the acquisition phase lead to gradual strengthening of a cue-outcome association, which is then gradually weakened, and hence forgotten, during the extinction phase (Pavlov, 1927; Rescorla &

Wagner, 1972). However, several experiments (Bouton & Bolles, 1979) and clinical observations (Craske et al., 2014) have challenged this view. Most importantly, seemingly extinguished memories can return unexpectedly, either after a sufficient period of time has elapsed (spontaneous recovery, Brooks & Bouton, 1993), an explicit change of context (renewal, Bouton & Bolles, 1979) or a presentation of the US on its own (reinstatement, Rescorla & Heth, 1975). In line with state-dependent accounts, these observations suggest that rather than forgetting or overwriting aversive associations, extinction involves the creation of a new memory together with the inhibition of the previous association (Bouton, 2002). One intriguing possibility which follows from this idea is that the individual may have learned that there are now two states that occur repeatedly ('acquisition'/'high-threat state' and 'extinction'/'low-threat state' in the example). The same stimulus can therefore have different associations with outcomes depending on the state, explaining why animals and humans can suddenly behave differently in response to the same observation in the case of spontaneous recovery, reinstatement or renewal (Dunsmoor et al., 2015). The notion of state-dependent learning is also in line with data showing that gradual, rather than abrupt, contingency changes lead to a decrease in the rates of spontaneous recovery, reinstatement and fear re-acquisition on a subsequent session (Gershman et al., 2013; Morís et al., 2017; Woods & Bouton, 2007), a phenomenon known as 'gradual extinction' or 'occasional reinforced extinction'. Using a model of state learning, Gershman & Hartley (2015) found that individuals who tend to represent the environment as multiple states have significantly higher rates of spontaneous recovery, as indexed by skin-conductance. This suggests a relationship between state inference and spontaneous recovery of aversive associations.

While much research supports the general existence of a state inference mechanism, the question of which factors influence the creation of internal states, and how trait anxiety might relate to it, has remained less clear. First, the role of trait anxiety (TA) in state inference has not been explicitly tested, although some studies suggest such a link (Gershman & Hartley, 2015; see also Norbury et al., 2021 for the same proposition in PTSD). High TA has been associated with an increased return of fear following phobia treatment (Rodriguez et al., 1999) and in renewal experiments (Staples-Bradley et al., 2018) as well as with heightened neural and physiological differentiation between cues associated with a shock (CS+) vs no shock (CS-) (Indovina et al., 2011; Sehlmeyer et al., 2011; Sjouwerman et al., 2020). Linking these findings to the theoretical work on state inference, it is conceivable that high TA individuals tend to learn in a more context-specific manner. If this context-specific learning is associated with a propensity to reactivate a previously experienced high-threat state, this could lead to repeated relapses as observed in patients. Here, we used computational models that take into account each individual's precise learning history to explore the relationship between trait anxiety and state inference. More specifically, we investigated whether state switching models provide a better account of learning from aversive outcomes, as a function of trait anxiety. Second, identifying the two states may depend on the amount of outcome uncertainty. Specifically, inferring two separate states may be easier when the objective "high" and "low" states are associated with 90% and 10% probability of shock versus when those contingencies are considerably lower (e.g., 60% and 40%). For this reason, relying on a gradual learning mechanism may be more suitable in some environments while relying on state inference may lead to more accurate predictions in others.

In this study, we investigated whether the level of outcome uncertainty (60/40%; 75/25% and 90/10%) impacts state inferences, and whether trait anxiety is associated with increased tendency towards inferring states.

## Results

Our main objective was to study how learning about changing aversive outcome associations is affected by variations in trait anxiety among healthy participants. Eighty-nine participants (44 female, mean age: 25.5 years) performed a probabilistic aversive reversal learning task during which they saw one of three possible cues and were then asked to rate the probability of receiving a shock (Fig 1a). The data set consisted of three experiments (N=30, N=22, N=37). Experiments I and II only involved one session (75/25, see below), Experiment III consisted of three sessions, with each session differing in outcome uncertainty (see below). In all three datasets, three different visual cues were each associated with a probability of receiving an electric shock. Two cues were consistently associated with either a high or a low shock probability throughout the session. We refer to these as stable-high and stable-low cues. The third cue started with a shock probability corresponding to either the stable-high or stable-low cue but reversed its probability a total of 6-10 times during each session (henceforth: *reversal cue*; reversals occurred randomly every on average 15.4 trials, see Fig 1b, as well as Materials and Methods for details).

The three sessions varied in the amount of outcome uncertainty. In the 90/10 session, the stable-high probability cue was followed by a shock on 90% of trials, while the shock appeared on only 10% of trials after the stable-low probability cue. In these sessions, the reversal cue switched between 90% and 10% objective states. The 75/25 and 60/40 sessions followed the same logic. The reversal cue probabilities switched between 75% and 25% and between 60% and 40%, respectively. Participants in Experiment 3 completed all three sessions, while participants in Experiments 1 and 2 only completed a 75/25 session. The session order in Exp. 3 was counterbalanced across participants (Fig 1c). Due to different conditions and session numbers in the three studies, the final number of participants for each condition differed when considering the full sample ($N_{60/40}$ = 36; $N_{75/25}$ = 88; $N_{90/10}$ = 37; see Methods and Supplementary Materials for a detailed breakdown).

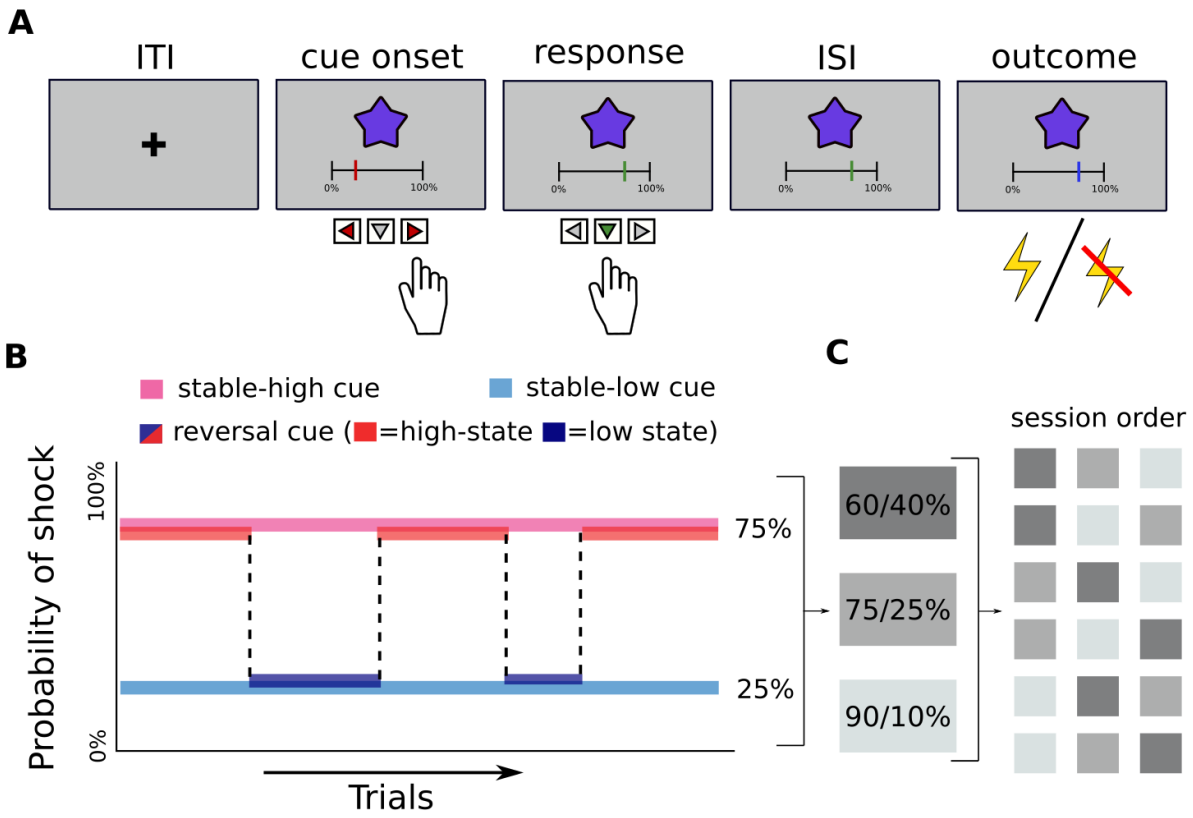


**Figure 1.** (**a**) Trial structure. Each trial started with a fixation cross (inter-trial interval; ITI) followed by one of the three cues (abstract fractals). Participants were asked to indicate the expected probability of receiving a shock on this trial by moving the red slider between 0% and 100% in increments of 1%. The final answer was submitted by pressing the downward arrow, after which the slider turned green to confirm the submission. After an inter-stimulus interval (ISI), a painful electrical stimulus (intensity: 8/10) was either delivered or omitted and the slider changed color to blue to indicate that the outcome had occurred. (**b**) Experimental design. The task was a continuous stream of trials. On each trial, one of the three cues was presented. While the shock contingency of the stable cues remained unchanged throughout the task (pink and light blue), the reversal cue changed every on average 15.4 trials between a low (dark blue) and a high (red) shock probability (i.e., 'state'). (**c**) In Experiment III participants completed the three sessions in a randomized order, in Studies I and II, only the 75/25 session was completed..

The shock intensity was individually calibrated at the beginning of each session to induce moderately high pain (rating of 8 on a numeric rating scale scale from 1 - 10, defined as 'painful but bearable considering the number of trials'). The calibrated stimulus intensity did not differ between studies. There was no significant relationship between shock intensity and probability ratings, or between pain intensity and trait anxiety ($p > .05$, see Methods). Session order and initial shock probability of the reversal cue (high or low) were also found to have no significant effect (see Methods). Experiments did not differ with respect to mean stimulus intensity applied or participants' trait anxiety level. However, because several details of the experimental protocol differed between experiment, we decided to include Experiment as a factor (i.e., random effect) in all analyses. All participants completed the STAI-TRAIT questionnaire (Spielberger, 2012), which was used to assess the individual trait anxiety (TA) scores (median score: 39; range 20-71). While TA was included as a continuous parametric variable in all relevant analyses, we report

and visualize mean ratings per anxiety 'group' (median split into high vs low trait anxious) in plots for illustration purposes.

### Differences in learning associated with trait anxiety

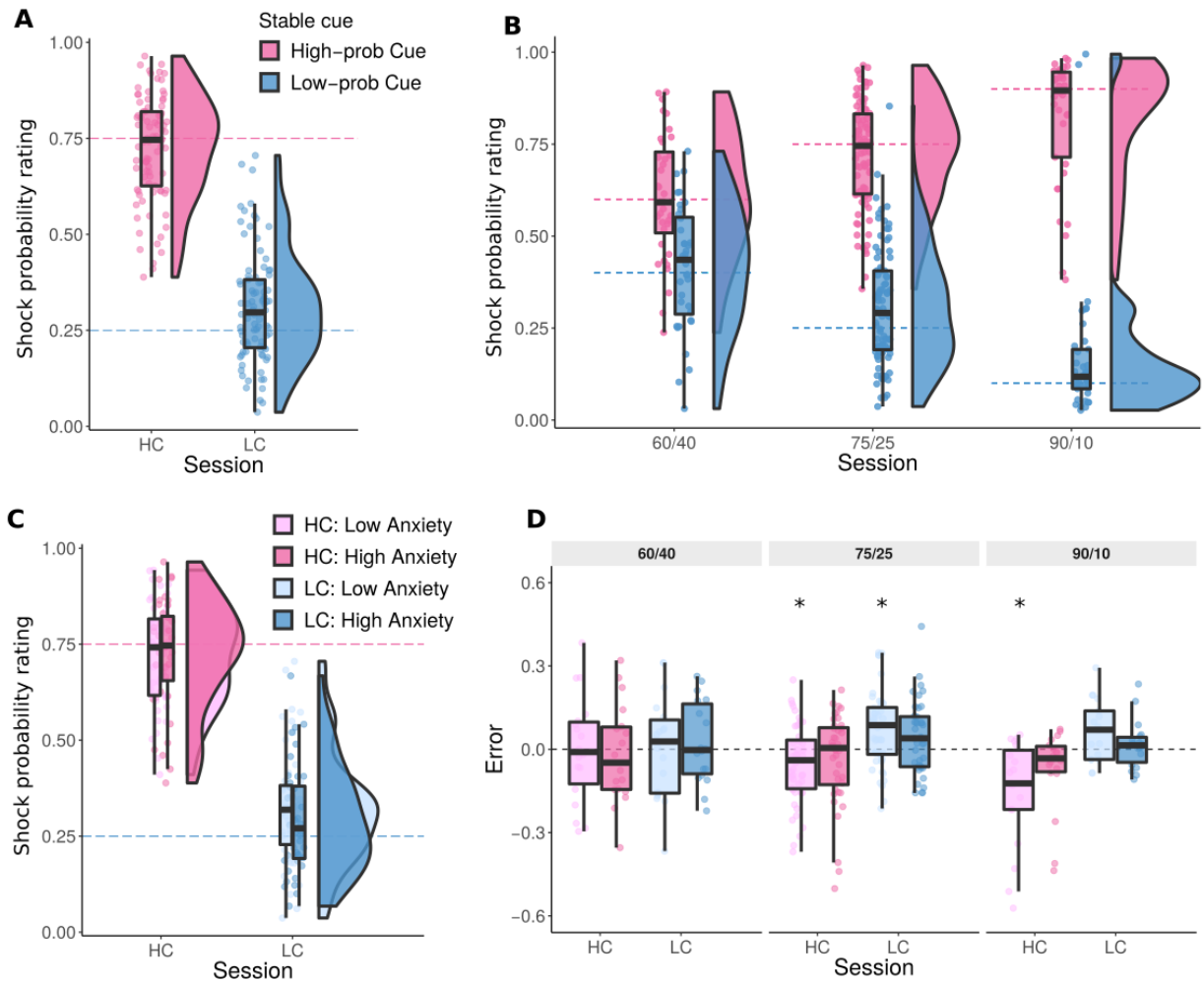


**Figure 2.** Mean probability ratings for the two stable cues across all sessions. Probability ratings were higher for the high-prob cue (pink) than the low-prob cue (light blue) across sessions (**a**) and (**b**) in each of the three sessions. (**c**) High trait anxiety was associated with lower shock expectation in the low-prob cue. (**d**). The divergence of ratings from the true reinforcement schedules (calculated as a cumulative mean) split by median TA, cue, and session. A positive error indicates an overestimation of shock probability, a negative error shows an underestimation. Asterisks indicate a significant difference from the true reinforcement level at p<.05 one-way t-test, p-value FDR-corrected. Dashed lines on all panels represent the true shock probability levels.

First analysis focused on participants' ability to track shock contingencies associated with the stable cues. A linear mixed effects model (LMM) with probability ratings as a dependent measure revealed a main effect of cue type (stable-high vs stable-low cue), $F(1,308)=435.8$, $p<.001$, with higher probability ratings for stable-high than stable-low cues (Fig. 2a). Although participants' ratings were relatively close to the true contingency levels, we found that participants slightly overestimated the shock probability for the stable-low cue (25% true vs. 30% estimated) and

slightly underestimated the probability for the stable-high cue (75% true versus 71.8% estimated probability). Participants' probability ratings also reflected the contingency differences between sessions, as shown by the significant interaction between cue type and session type (90/10, 75/25 and 60/40), $F(1,308)=33.83$, $p<.001$, see Fig 2b (also Supp. Mat. section results for details).

Next, we asked whether participants' shock probability ratings for the two stable cues differed depending on trait anxiety. We averaged shock probability ratings per cue and participant, and ran a LMM with trait anxiety and session as fixed effects. This analysis revealed that the difference in ratings between high- and low-prob cues increased as a function of trait anxiety, as indicated by an interaction of TA and cue type, $F(1,308)=6.91$, $p=.009$, see Fig 2c. There was a positive association with TA in stable-high cue, $\beta=.0024$, and a negative relationship in the stable-low cue, $\beta= -.0024$: high TA participants reported higher ratings in stable-high and lower ratings in stable-low cue. Direct contrast of the associations of TA and rating between high and low-prob cues showed significant difference, $t(242)=2.63$, $p=.009$. We also tested whether ratings differed significantly from the true contingency level using one-way t-tests, see Fig 2d. When judging the stable-high cue, less anxious participants significantly underpredicted the true reinforcement level in the 75/25, $t(47)=-2.62$, $p=.047$, and 90/10, $t(18)=-3.51$, $p=.015$, conditions. When judging the stable-low cue, less anxious participants overpredicted the probability in the 75/25 condition, $t(47)=3.58$ $p=.010$. More anxious participants, in contrast, did not show over- or underpredictions, all *p*s > .05.

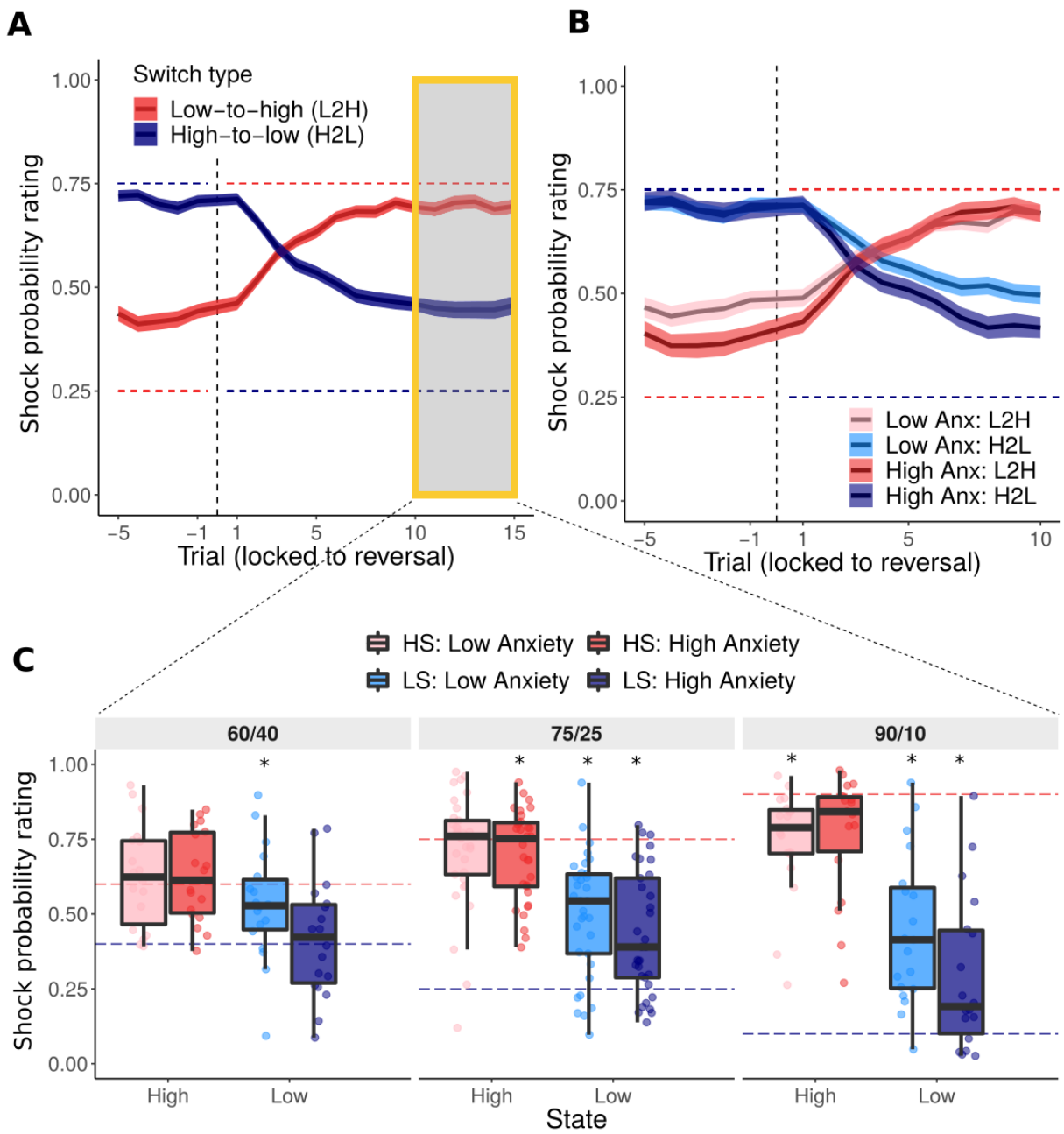


**Figure 3. (a)** Mean shock probability ratings locked to the reversal point for high-to-low (blue) and low-to-high (red) switches across the entire sample and (**b**) separately for high and low trait anxious participants. Note that negative trial numbers indicate trials prior to a reversal. (**c**). Mean shock probability ratings for trials 10 after current reversal until next reversal shown separately for each session, state, and median-split trait anxiety. This window was selected as the point from which shock ratings stabilized after the reversal. Asterisks indicate a significant deviation from the true reinforcement level at p<.05 one-sample t-test, p-value FDR-corrected. Dashed lines on all panels represent the true shock probabilities.

Next, we analyzed shock probability ratings for the reversal cue. This cue switched between states of high and low-probability of shock ("high-prob state" and "low-prob state"). After each reversal, ratings tended to change rapidly, settling within about 10 trials at a stable level thereafter

(see Fig 3a). We first focused on participants' probability estimates during the stable periods on trials 10 until the next reversal (orange box in Fig 3a), that is, after the initial learning period. Ratings provided during the high-prob state were higher than during the low-prob state across all three sessions (main effect of state: $F(1,207)=151.62$, $p<.001$; 63.5% vs 48.5% in 60/40, 70.6% vs 46.5% in 75/25 and 76.9% vs 38.3% in 90/10, high- vs low-prob states respectively, Fig 3a). Moreover, the model revealed an interaction between state and session, $F(2,207)=9.16$, $p<.001$. Post hoc tests found that this was driven by increased ratings of the high-prob state in the 90/10 relative to the 60/40 session, $t(205)=-3.40$, $p=.011$. No difference was found between sessions in the low state (See Results in Supp. Mat.).

This analysis also revealed a significant interaction between state and TA, $F(1,207)=15.33$, $p<.001$ (Fig 3b). Post-hoc tests found that this was driven by a significant negative relationship between TA and ratings in the low-prob state, $F(119)=10.82$, $p=.001$. More specifically, low trait anxious participants overestimated the shock probability in the low probability state by 25.3%, compared to 13.5% in high TA. No difference was found for ratings in the high-prob state, where low TA participants underpredicted the shock probability by 4.5% and high TA participants by 4.9%. This pattern was similar in all three sessions, although the association with TA was numerically most pronounced in the 90/10 condition (interaction with session: $p > .05$; Fig 3c, Supp. Table. 3).

We next focused on the learning immediately after a reversal, i.e., trials 1 to 10 (‘reversal period’). We characterized the speed of learning following a reversal by fitting a line to ratings on trials 1 to 10. This was done using a LMM with slope for each participant and state. As expected, slopes differed depending on the direction of the switch, i.e., low-to-high switches were positive (2.44%; read as ‘the shock probability rating increased by 2.44% per trial’) while slopes in high-to-low switches were negative (-2.41%, main effect of state, $F(1,621)=429.26$, $p<.001$, see Fig 4a). In both states, learning was steeper in the 90/10 condition compared to both 60/40 and 75/25 (low-to-high switch: $t_{60>90}(559)=-6.37$, $p<.001$, $t_{75>90}(619)=-4.979$, $p<.001$; low-to-high switch: $t_{60>90}(559)=4.97$, $p<.001$, $t_{75>90}(618)=3.74$, $p=.003$). Furthermore, there was a significant two-way interaction between anxiety and state, $F(1,621)=21.38$, $p<.001$, and three-way interaction between anxiety, state and session, $F(2,621)=8.56$, $p<.001$. Post-hoc tests revealed that faster learning was associated with higher TA in the 90/10 session both in low-to-high switches, $F(1,551)=12.52$, $p<.001$ (low-TA: 2.96% vs high-TA: 5.00%), and high-to-low switches, $F(1,551)=16.38$, $p<.001$ (low-TA: -2.39% vs. high-TA: -4.78%). Direct contrast of the TA*slope trends between high-to-low and low-to-high switches revealed that the three-way interaction was driven by a significantly stronger relationship in the high-to-low state, $t(204)=3.07$, $p=.029$, see Fig 4a.

Because averaging over variable time courses can lead to incorrect conclusions (e.g. Haider & Frensch, 2002), we estimated the steepness and ‘switch point’ of each individual reversal directly from the data, providing a measure that is more sensitive to individual time courses. The estimated steepness was closely linked to the slope measure reported in the previous paragraph, $r(88)=.775$, $p<.001$. It increased across sessions (-2.30, -2.10 and -1.4 on log scale), as reflected in a main effect of session, $F(2, 205)=20.80$, $p<.001$. This was driven by significantly higher steepness in the 90/10 condition compared to 60/40, $t(226)=-6.14$, $p<.001$, and to 75/25, $t(181)=-4.61$, $p<.001$. The steepness was also positively associated with trait anxiety TA, $F(1,94)=7.39$,

p=.008, indicating that more anxious participants adjusted their shock probability ratings faster than less anxious participants. Turning to analysis of the estimated switch point. The average switch occurred 2.91, 2.90 and 3.28 trials after the true reversal (60/40, 75/25, 90/10 respectively). There was no relationship between the estimated switch point, session and TA. This indicates that while high trait anxious participants performed more abrupt switches, these did not occur earlier or later compared to individuals lower in trait anxiety, and this was true for all sessions.

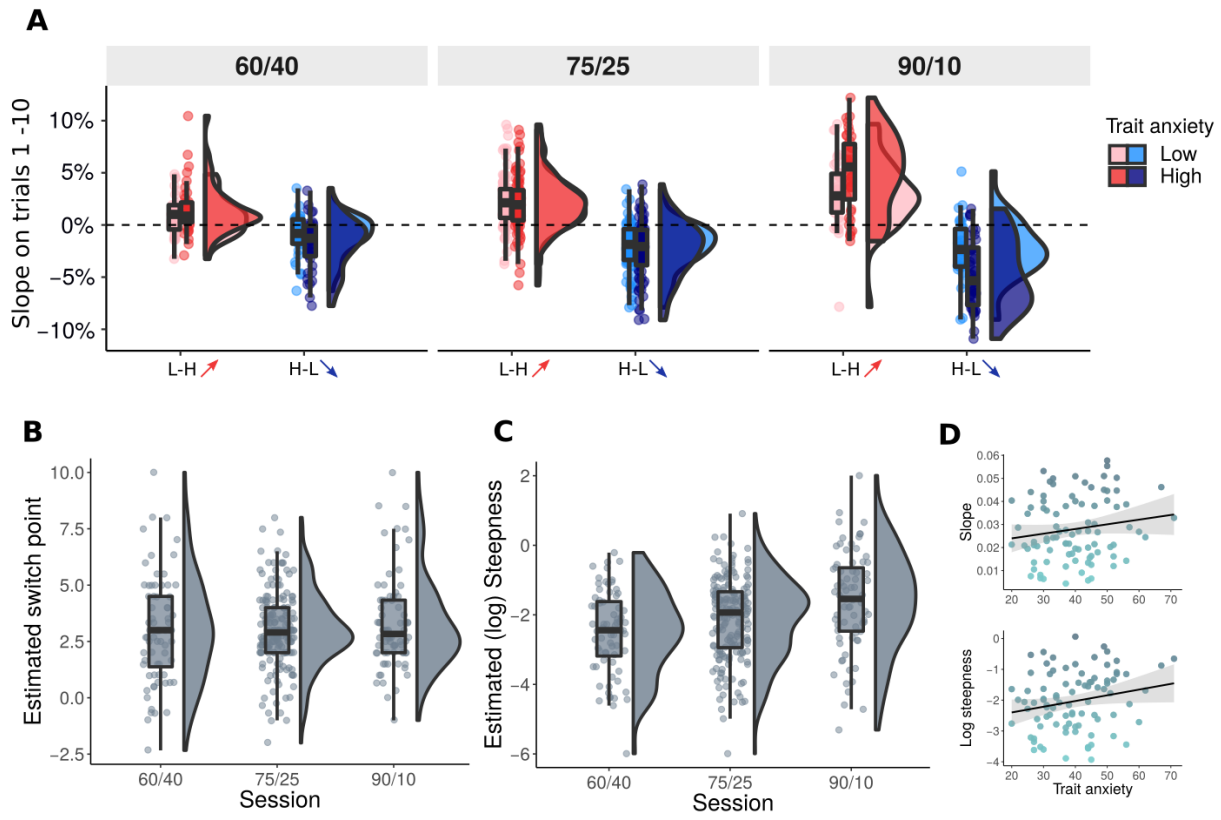


**Fig 4** (**a**) Slope of change of reported ratings on trials 1-10 following contingency reversal separately for each session and anxiety level. Positive values indicate an increase in shock probability ratings and negative values indicate a decrease. L-H denotes a switch from a low to high shock probability, H-L denotes a switch from high to low shock probability. (b) The distribution of switch-point and (c) steepness estimates separately for each session. (d) Both the slope on trials 1-10 (upper panel) and estimated switch steepness (lower panel) showed a positive relationship with TA across all sessions and switch directions.

## Modeling state switching and gradual learning

The results above suggested that trait anxiety is linked to faster updating of expectations when contingencies change. Such behavior could either be based on faster gradual learning or reflect state switching. To distinguish between these two ideas more formally, we fitted models to participants' probability ratings in of the reversal cue that assumed either gradual updating of a single state ('1-state' model) or updating of, and switching between, multiple states ('n-state' model). In both models, a state was characterized by a beta distribution that reflected the learner's

current belief about shock probability. The 1-state model formalized gradual updating of a single beta distribution (see de Boer et al., 2017; Lamba et al., 2020; Wise & Dolan, 2020). The shape of the distribution was controlled by parameters α and β, which were adjusted following each outcome. Specifically, following a shock α was updated using a step size parameter $\tau^{+}$, whereas β was increased by $\tau^{-}$ when no-shock was received. To model forgetting, in all trials both parameters also decayed by a free parameter λ, so that $\beta_{t+1} = \lambda\beta_t$ and $\alpha_{t+1} = \lambda\alpha_t$. In essence, the model captured gradual learning in a manner resembling the Rescorla-Wagner model, although it should be noted that the beta updating rule also incorporated accelerated learning rates following reversals, akin to Pearce Hall model (see Methods and SI, for explicit comparison between 1-state learner and RW and PH). The n-state model started each session with a single beta distribution that was updated over time in a manner identically to the 1-state model. Crucially, however, the model kept track of how surprising recent observations were, given the current state. If the surprise exceeded a threshold (controlled by a free parameter η), the model created a new state that minimized the current surprise. If more than one state already existed, the model first polled for existing states, and switched to them if a suitable candidate was found, before creating a new state. Every time time a new state was created the threshold for creating a new state increased. This allowed the model to generate a meaningful number of states and effectively switch between them. Both models are described in detail in Materials and Methods, and parameter estimates are summarized in Tables 1a and 1b.

To assess which participants tended to learn gradually versus infer states, we calculated BIC scores for both models and subtracted them. Note that although the n-state model has one additional parameter (the threshold), it can behave almost identical to the 1-state model when the threshold is so large that the model never creates more than one state. Lower BIC scores can therefore be attributed to the necessity of inferring and switching states, i.e., that participants with improved model fit for the n-state over 1-state model are likely to rely on a state inference mechanism. Testing the model across all sessions, the n-state model fitted the data better (1-state BIC: -118; n-state BIC: -123). This was also true when comparing model fit for all three sessions individually, 60/40 (-84 vs -87), 75/25 (-133 vs -138) and 90/10 (-116 vs -126). However, the most pronounced difference was found in the 90/10 condition where the n-state model improved fit substantially (see Fig 5b). This suggests increased reliance on state inference in environments with larger switches in probabilities.

Improved model fit should also be reflected in behavioral signatures of state switching. In line with this assumption, data of participants better fit by the 1-state than the n-state model exhibited more gradual learning, while steeper post-reversal learning was found in participants with better relative fits of the n-state model (Fig 5a, see also Supp. Fig. 1). To quantify this impression, we assessed the relationship between behavioral signatures of state switching and the improvement in model fit. First, we regressed the differences in model fit against the fitted slopes from participants' shock ratings (Fig. 4). This revealed a significant positive association across all three sessions, $F(1,146)=4.08$, $p=.045$, correlation $r(88)=.398$, $p<.001$, indicating that improved fit of the n-state model related to the steepness of estimated switches. Second, we reasoned that those participants employing a state inference strategy should be better at dissociating *when* to learn from outcomes, i.e., they should show less learning from oddball events compared to learning

from trials just after reversal. To test this, we calculated model-free learning rates separately for 10 trials immediately after reversal (i.e. “meaningful learning”) and trials during the relatively stable periods between trial 10 and the next reversal (“oddball trials”, see Methods). Participants who were fitted better by the n-state learned more from outcomes occurring after reversal compared to oddballs ($\alpha_{meaningful-oddball} = 0.073$) while participants fitted better by the 1-state model had a smaller difference in learning rates ($\alpha_{meaningful-oddball} = 0.02$), t(56)=-2.14, p=.036.

We next examined the relationship between trait anxiety (TA) and state inference by constructing a LMM with model fit difference as the dependent variable and TA and session as fixed effects. This model identified a significant interaction between TA and session, F(2,101)=6.90, p=.002. Post-hoc analyses revealed that this was driven by a positive association between TA and fit improvement in the 90/10 condition, F(1, 153)=9.44, p=.003 (see Fig. 5c). The association was significantly stronger in the 90/10 condition compared to both 60/40, t(78)=-3.16, p=.006 and 75/25, t(111)=-3.33, p=.003. This result was confirmed by a permutation-tested correlation between TA and model fit improvement which was significant in the 90/10 condition, r(36)=.40, p<.05 (alpha = 0.05, two-tailed, corrected for multiple comparisons). These results suggest that individuals high in trait anxiety tend to rely more on state inference.

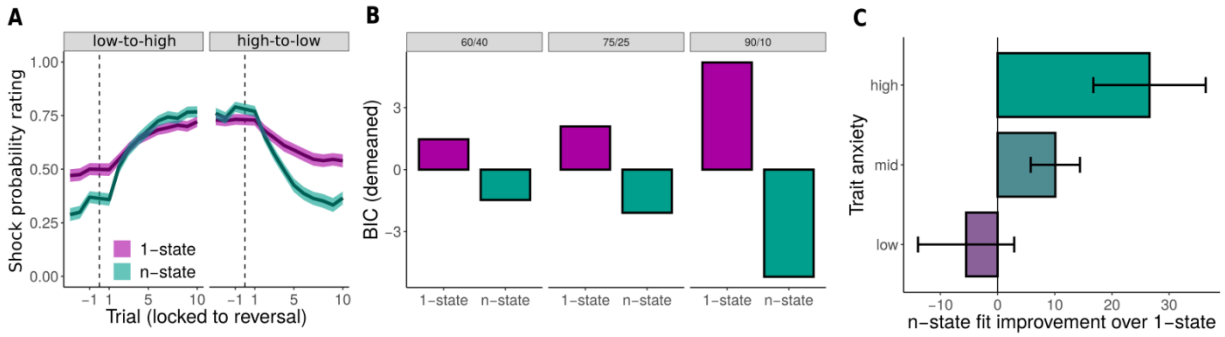


**Fig 5.** (**a**) Mean shock probability ratings separately for participants better fit by 1-state (purple) and n-state (green) models. (**b**) Mean BIC scores for the two models. BIC scores were demeaned to make sessions visually comparable. (**c**) Increased reliance on state inference with increasing anxiety in the 90/10 condition. TA was split into terciles for visualization.

To better understand the fitted n-state model, we explored it’s behavior and fitted parameters in more detail. First, we analyzed whether the number of states the fitted model created was related to TA, but this was not the case (mean number of states high TA: 1.68 vs. 1.66 in low TA on average; 90/10 condition: 1.83 vs. 1.68. Second, we investigated the timepoint when the model tended to switch states. A corresponding LMM found no effect of session or anxiety, consistent with our behavioral results reported above (see “estimated switch point”, see Fig 4b; model inferred switch points were 3.51, 3.15 and 3.52 trials after the true reversal, for the three sessions) Next, we analyzed the fitted step sizes for the positive and negative outcomes $\tau^+$and $\tau^-$. A LMM with parameter type ($\tau^+$/ $\tau^-$), TA and session as fixed effects found a significant main effect of parameter type, F(1,231)=24.26, p<.001, which reflected that shocks elicited larger updates than no-shocks $\tau^+$=1.08 vs. $\tau^-$=0.74). There was no main effect of TA, F(1,87)=.27, p>.05, or

interaction of outcome type and trait anxiety, F(1, 232)=.01, p>.05. Note that the same two parameters of the 1-state model, had a similar difference $\tau^{+}$=1.17 vs.$\tau^{-}$=0.85), suggesting that differential learning from shock and no-shock events alone was unable to explain our behavioral effects of TA. Lastly, no effects of session, TA or interaction where found when analyzing $\eta$ which controls the switch threshold $\eta$ (see Table 1b).

***Table 1a: Estimated parameters for 1-state model***

| | 60/40 | | | 75/25 | | | 90/10 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Stable-high (N = 36) | Stable-low (N = 36) | Reversal (N = 36) | Stable-high (N = 88) | Stable-low (N = 88) | Reversal (N = 88) | Stable-high (N = 37) | Stable-low (N = 37) | Reversal (N = 37) |
| **Tau shock** | 1.07 ± 0.71 | 0.91 ± 0.72 | 1.00 ± 0.69 | 1.21 ± 0.73 | 0.89 ± 0.69 | 1.15 ± 0.69 | 1.44 ± 0.70 | 0.69 ± 0.65 | 1.17 ± 0.76 |
| **Tau noshock** | 0.91 ± 0.73 | 1.04 ± 0.66 | 0.79 ± 0.57 | 0.72 ± 0.60 | 1.18 ± 0.75 | 0.74 ± 0.56 | 0.66 ± 0.77 | 1.47 ± 0.73 | 0.82 ± 0.68 |
| **Alpha0** | 5.07 ± 3.27 | 4.24 ± 3.19 | 5.00 ± 3.52 | 4.96 ± 3.37 | 4.53 ± 3.42 | 5.34 ± 3.22 | 4.70 ± 3.66 | 3.06 ± 2.52 | 5.55 ± 3.31 |
| **Beta0** | 4.93 ± 3.38 | 4.54 ± 3.24 | 4.59 ± 3.37 | 4.57 ± 3.10 | 5.21 ± 3.32 | 4.99 ± 3.35 | 3.37 ± 3.09 | 4.30 ± 3.02 | 3.85 ± 3.00 |
| **Lambda** | 0.87 ± 0.17 | 0.92 ± 0.10 | 0.86 ± 0.18 | 0.87 ± 0.19 | 0.88 ± 0.16 | 0.87 ± 0.13 | 0.89 ± 0.23 | 0.94 ± 0.07 | 0.85 ± 0.12 |

***Table 1b: Estimated parameters for n-state model***

| | 60/40 | | | 75/25 | | | 90/10 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Stable-high (N = 36) | Stable-low (N = 36) | Reversal (N = 36) | Stable-high (N = 88) | Stable-low (N = 88) | Reversal (N = 88) | Stable-high (N = 37) | Stable-low (N = 37) | Reversal (N = 37) |
| **Tau shock** | 1.24 ± 0.75 | 0.92 ± 0.66 | 0.96 ± 0.79 | 1.19 ± 0.76 | 0.77 ± 0.69 | 1.05 ± 0.71 | 1.31 ± 0.80 | 0.43 ± 0.61 | 1.10 ± 0.74 |
| **Tau noshock** | 0.77 ± 0.77 | 1.02 ± 0.69 | 0.83 ± 0.67 | 0.58 ± 0.61 | 1.14 ± 0.73 | 0.57 ± 0.53 | 0.67 ± 0.80 | 1.44 ± 0.73 | 0.70 ± 0.64 |
| **Alpha0** | 5.56 ± 3.28 | 4.85 ± 3.17 | 5.51 ± 3.48 | 5.32 ± 3.29 | 4.94 ± 3.47 | 5.16 ± 3.13 | 6.22 ± 3.60 | 5.63 ± 3.52 | 5.43 ± 3.15 |
| **Beta0** | 5.71 ± 3.15 | 4.95 ± 3.13 | 4.72 ± 3.10 | 5.24 ± 3.05 | 5.33 ± 3.16 | 5.11 ± 3.25 | 5.87 ± 3.17 | 6.09 ± 3.37 | 4.11 ± 2.89 |
| **Lambda** | 0.82 ± 0.31 | 0.92 ± 0.09 | 0.85 ± 0.25 | 0.88 ± 0.22 | 0.89 ± 0.19 | 0.88 ± 0.14 | 0.89 ± 0.25 | 0.94 ± 0.07 | 0.87 ± 0.14 |
| **Eta** | 4.88 ± 2.86 | 5.41 ± 3.19 | 5.25 ± 3.08 | 4.88 ± 3.29 | 5.25 ± 3.01 | 5.48 ± 2.85 | 3.34 ± 2.79 | 4.21 ± 3.38 | 5.03 ± 2.72 |

Finally, we checked whether the link between TA and model fit improvement would also be expressed in more differential learning between meaningful outcomes in the trials following a reversal and oddball outcomes.. While we found significantly higher learning on meaningful trials ($\alpha$=0.263) than on oddball ($\alpha$=0.218) trials in general, F(1, 208)=10.78, p=.001, we found no evidence for an association between meaningful/oddball learning and TA (see Fig 6).

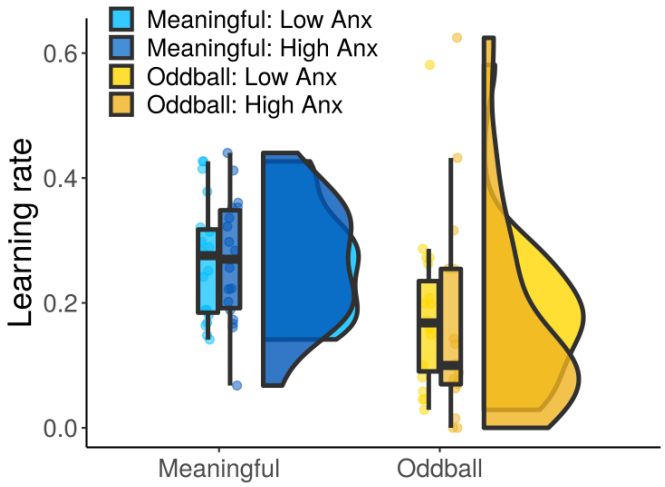


**Figure 6.** Learning rates in the 90/10 session split by 'meaningful learning' (blue) and 'oddball' (yellow) trials. Learning during meaningful periods (immediately after reversal) is generally higher than learning from oddball events. High TA participants tend to learn slower from oddballs, but this effect is not significant

## Discussion

We investigated how outcome uncertainty and trait anxiety influence the computational mechanisms of learning and reversal in a probabilistic aversive setting. We modelled participants' ratings of shock probabilities in a reversal learning task using two approaches: a gradual learning rule that updates a single cue-shock association (1-state model), or a model that could learn an appropriate number of states characterized by different cue-shock associations and switch between these states (n-state model). The two models captured the difference between gradual learning and context-dependent state inference, switching, and updating.

Our results showed that trait anxiety was associated with behavioral markers of state inference and switching, and improved fits of the n-state model. In particular, participants high in TA showed faster changes in probability ratings after a reversal, steeper switches, and better learning of the true outcome contingencies overall. These results suggest that high trait anxious individuals tend to represent the environment as consisting of distinct hidden states, which manifested as faster learning after a reversal. Hence, we provide direct evidence for a link between high trait anxiety and the tendency to infer hidden states, and to switch between them. Both TA and state inference have been independently linked to the return of fear. Trait anxiety is known to be associated with higher rates of fear relapse (Rodriguez et al., 1999; Staples-Bradley et al., 2018). Representing environments as multiple states (i.e. memories) leads to higher rates of spontaneous recovery (Hartley and Gershman, 2015). Our findings suggest that trait anxiety, as a time-invariant disposition (Usala & Hertzog, 1991), facilitates the parcellation of observations into different states that are characterized by different cue-outcome contingencies. In the clinical literature, the assumption of several independent states has been discussed in relation to prevention of updating of existing cue-outcome associations (i.e., overwriting previous memories) and thereby to hinderance of effective fear extinction (Craske et al., 2014). Instead of revising the current situation, the individual assumes an additional new state that reflects the altered contingencies.

We note that the two learning mechanisms (state inference versus gradual learning) are not mutually exclusive, but might rather reflect different *degrees* of state-dependent learning (as in Hartley and Gershman, 2015). Our results indicate that the propensity for state-dependent learning might depend on the amount of outcome uncertainty in the environment, since better fits of the n-state model were observed in sessions with more distinct high- and low-probability states (90% and 10%), as compared to sessions with less distinct states.

The results corroborate previous theoretical predictions. In particular, parcellation into separate states (or "contexts") was proposed to be associated with anxiety disorders and account for relapse phenomena (Bouton, 2002; Gershman & Hartley, 2015, see also Norbury et al., 2021). Here, it is assumed that 'extinction' involves learning that is separate and more complex from 'acquisition'. Therefore, the memory for acquisition needs to be inhibited during extinction while a new association needs to be learned (see Bouton, 2002). In our behavioural results, the effects of anxiety were driven by more accurate probability estimates in the stable-low cue and the low-state of the reversal cue, which both correspond to conditions of *relative* safety. This aligns with some previous reports. For example, in a gamified aversive learning paradigm Wise and Dolan (2020) reported a positive association between safety learning and state anxiety. Similarly, Raes et al., (2009) found that under a condition of higher cognitive load fear extinction (indexed by SCRs) was more successful in the high TA group. Interestingly, high anxiety (i.e., factor loading high on trait anxiety) was associated with increased engagement of cognitive control in a go/no-go paradigm (Scholz et al., unpublished), suggesting that non-clinical TA might be associated with better use of cognitive resources. The broader literature on the relationship between trait anxiety and fear yields mixed results. While some studies report increased discrimination of CS+ and CS- (Sjouwerman et al., 2020) and comparable fear inhibition during extinction in high vs. low TA individuals (Kindt & Soeter, 2014; Torrents-Rodas et al., 2013), others report deficits in inhibitory processing (Ansari & Derakshan, 2011; Haaker et al., 2015; Myers & Davis, 2007) and safety learning (Gazendam et al., 2013; Indovina et al., 2011). This diversity of findings has not been reconciled, however, a possible explanation in terms of methodological differences (e.g., modality, aversiveness, outcome uncertainty) has been suggested (Torrents-Rodas et al., 2013).

How and whether states are inferred depends on uncertainty. In our data, state inference was most favored in environments where objective changes in shock contingencies were largest (90%/10%), i.e., when outcome uncertainty was low. However, in changing environments, inferring whether the objective state has changed also depends on higher order uncertainty such as volatility (Behrens et al., 2007). When receiving a surprising outcome, one must consider both outcome uncertainty and volatility to determine whether it reflects change in state or an oddball event (Piray & Daw, 2021; Yu et al., 2021)**.** Interestingly, high TA has previously been associated with the inability to adjust learning to environmental volatility, as reflected in a high learning rate despite stable contingencies (Browning et al., 2015). Piray and Daw (2021) demonstrated how mis-estimation of outcome uncertainty (stochasticity) rather than volatility can drive learning and cause fast, jump-like learning from rare events due to misestimation of stochasticity. To test whether such mis-estimation could explain our findings (as opposed to state inference), we compared learning rates during the period just after a reversal ('meaningful learning') with learning rates in relatively stable periods, where unexpected outcome did not signal a reversal but rather an exception that was to be ignored ('oddball learning'), see Fig 6. We argue that bigger difference

in learning rates for meaningful and oddball events reflects state awareness. While we indeed found higher learning rates in the period just after the switch, this was not modulated by trait anxiety. Importantly, high TA individuals did also not learn relatively more from oddball events. As mentioned above, our results diverge from previous findings that reported lack of fear inhibition during extinction (e.g., Kindt & Soeter, 2014) and deficits in safety learning (Gazendam et al. 2013) in high trait anxiety. A number of methodological differences between our study could account for these differences. Our study aimed to investigate a temporally extended learning process, which contrasts with the fear extinction paradigm used in the above-named studies. In particular, we used a task in which both 'acquisition' (high state) and 'extinction' (low state) are probabilistic. Notably, in the low probability, i.e. extinction, state, probabilities ranged between 10% and 40% whereas in other studies, this range was used during acquisition (e.g., 33% in Phelps et al., 2004 or Schiller et al., 2008). Our choice was in part motivated by the importance of keeping the degree of outcome uncertainty identical in both states within each session. As Ojala & Bach, (2020) pointed out, uncertainty is a confound in studies where acquisition is probabilistic (e.g., 50% shock) but extinction is deterministic (0%) (see also Discussion in Torrents-Rodas et al. 2013). Another important factor is that in our design, states of high and low shock probability occur repeatedly as each participant experienced at least six contingency switches. This decision was again motivated by real-world conditions where aversive stimuli often reoccur (e.g., periods of back pain, exam stress). Our focus was to understand how individuals with varying degrees of trait anxiety intrinsically learn and represent the structure of an aversive environment which sets the study apart from classical studies on acquisition and extinction. However, future research should systematically investigate the role of trait anxiety under different relative conditions of threat, including the difference in probabilistic versus deterministic environments.

A noteworthy aspect of our work is the novel model that captures state inference and updating. It combines single-state updating models under a beta distribution (de Boer et al., 2017; Lamba et al., 2020; Wise & Dolan, 2020) with state inference models proposed previously (Costa et al., 2015; Gershman & Niv, 2012; Redish et al., 2007). The key feature of the model is that it can translate binary outcomes into probabilistic states, quantifying the current expectation and its uncertainty in the process (see Materials and Methods). We showed that the n-state model was able to estimate the appropriate number of states and that model-estimated switches occurred in the same period as in the behavioral data (see Methods and Supp. Materials). Most importantly, there was a clear behavioral distinction between participants better fitted by the 1- versus n-state models (see Fig. 5a and Supp. Fig. 2).

Despite the model performing well for our purpose, it might require adjustments in other paradigms depending on the task and data. For example, in our version the mechanisms by which the model switches between states as opposed to creating new states are codependent and only differ in the difficulty parameter (i.e., more surprise is needed to create a new state with increasing number of states). Future implementations of this model could include entirely separate thresholds for state switching versus state inference, for example, to study whether some groups tend to create too many states but never switch to a pre-existing state.

### Conclusion

Taken together, our results suggest that more trait anxious individuals have a tendency to represent aversive environments involving high and low threat contexts as distinct states and to switch between them. We suggest that this parcellation of the environment into states may explain previously observed fear relapse phenomena associated with trait anxiety.

## Materials and Methods

We presented a pooled analysis of three studies that used a probabilistic aversive learning task (see *Task*). Experiment I is an fMRI experiment consisting of a short screening session and a main session 1 – 3 days later. Only data from the main session were included in this analysis. Experiment II is a three-visits (visit 1: baseline, visit 2: drug administration; visit 3: follow-up) drug study investigating the role of the angiotensin-II inhibitor drug losartan in aversive learning. Only the placebo group from the second visit was included in this analysis as the task on visits 1 and 3 was shorter. This ensures that participants in all three studies had the most similar experience. For a detailed overview of the three studies see Supplementary Table I.

Data were collected in the 3 Tesla MRI scanner of the Wellcome Center for Integrative Neuroimaging (Experiment I), in behavioral testing laboratories of the Nuffield Department of Clinical Neurosciences (Experiment III) (John Radcliffe Hospital, Oxford) and the Department of Psychiatry, Warneford Hospital (Experiment II), Oxford. The factor ‘experiment’ was included as a random effect in all main analyses. All three studies were approved by the Central University Research Ethics Committee (CUREC) of Oxford University (R44738/RE001, R29583/RE004, R52892/RE001).

### Participants

Participants for all three studies were recruited using local advertisement and the SONA recruitment system managed by the Department of Experimental Psychology, University of Oxford. Inclusion criteria varied slightly between studies (due to MRI data collection in Experiment I and drug administration in Experiment II). A comprehensive list of criteria can be found in the Supplementary Materials (section ‘Inclusion and exclusion criteria’). All studies included right-handed healthy adults aged between 18 and 40 years without a history of psychiatric illness and not taking any psychoactive medication (including recreational drugs) at least 3 months prior to the experimental session. In line with recent recommendations for exclusion criteria in aversive learning studies (Lonsdorf et al., 2019), data of all participants, including non-learners, were included in the analyses.

In total, 116 participants took part in the three studies. Four participants were excluded due to missing behavioral data (presentation computer or shock administration stopped working properly), two because of missing anxiety scores (both participants had to leave the lab before completing the questionnaire) and one for misunderstanding the task. Twenty-two participants who received the drug in Experiment II were not included because questionnaire and physiological measures necessary to control for the effect of the drug were not available in the other two studies. In two participants, data of one of three sessions was missing due to script or stimulation failure. In this case, we included data of the remaining sessions in the analyses. Note, however, that therefore the degrees of freedom vary between sessions. The final number of participants included in the analyses was 89 (87 without any missing sessions).

### Aversive stimuli

Electrical stimuli were applied using a commercial electric stimulation device (Constant Current Stimulator, model DS7A; Digitimer, Hertfordshire, UK), delivering a 2 ms monopolar square waveform pulse via a concentric silver chloride electrode attached to the back of the left hand.

The stimuli were calibrated individually at the beginning of the task and during any pauses (Experiment I – every 13 to 18 min; Experiment II – every 12 to 15 min; Experiment III – just once at the beginning of each session – every 20 to 25 min). The target intensity was 8 on a scale ranging from 0 (= ‘not painful’) to 10 (= ‘too painful to take part’) scale. The 8/10 pain level was defined as a sensation that is painful but tolerable for a given number of trials (study-specific number corresponding to 50% of trials). Three qualitative anchor points were defined to help standardize the calibration across participants and studies: 1/10 which was defined as the intensity at which the sensation starts to be moderately painful (pain threshold); 8/10 is a sensation that is clearly painful but tolerable; and 10/10 which would be the level of pain which is too strong to be tolerated. The calibration followed the Method of Limits (see e.g., Ploner et al., 2010). The stimulus intensity started at the pre-calibrated 1/10 level and changed after each rating in an increasing trend (individual stimuli could however get stronger or weaker). Upon each stimulus delivery, participants were asked to report how painful the sensation was on a rating scale ranging from 1 to 10. When a rating was higher than 8, the next stimulus was always lower. The calibration terminated once three out of the last five stimuli were rated as exactly 8. To ensure that the intensity remains at a subjective 8/10 level, regular re-calibrations took place.

### Task

The goal of the study was to investigate how participants learn to predict the probability of an aversive event and how they update their expectations on a trial-to-trial basis. To this end, we used a Pavlovian probabilistic learning task in which participants learned to associate three visual cues (abstract fractals, selected randomly for each participant from a pool of 20 possible fractals) with the delivery or omission of a painful electrical stimulus (shock). On each trial, participants were presented with one of the cues which could be followed by the electrical stimulation and asked to submit an expectancy rating. Throughout the experiment, one of the cues was followed by a shock on a high proportion of trials (60% to 90%) while no stimulus was applied in the remaining trials (‘stable high-prob cue’). For the second cue, contingencies were reversed, i.e.,

the electrical stimulus was applied in a low proportion (10% to 40%) of trials ('stable low-prob cue'). For the third cue, shock contingency switched between the low and high probability in semi-regular intervals, mean 15.3 trials. Since our primary analysis goal was to study how people learn about changes in contingencies, we designed the task in a way that the reversal cue appeared more often than the stable cues in all three studies (see Supp. Table I).

*Standard Trial Structure*

Each trial started with the presentation of a fixation cross (inter-trial-interval, ITI; Experiment I: 3 – 5 s; Experiment II: 2 s; Experiment III: 1 – 2 s). Next, the cue for this trial was presented and the participant had 4 s to submit a response. The fractal was shown in the middle of the screen while the slider used to provide the rating was positioned below. Using the left and right arrow keys (MRI button box in Experiment I), participants could move the slider on a scale from 0% to 100% (in increments of 1%). Once the desired position was reached, they could confirm and submit their rating by pressing the down arrow key (middle button on the MRI button box). Participants had up to 4 s to submit the rating. If rating was not submitted on time a warning message appeared and the trial was restarted. Once the rating had been submitted, the slider changed color to green. After an inter-stimulus interval (ISI; Experiment I: 2 – 4 s; Experiment II: 1 s; Experiment III: 1 – 2 s) the outcome was delivered (i.e., shock delivery or omission). The outcome was accompanied by a change in the color of the slider to blue (to make timing of outcome equally clear to participants for both shock and no-shock trials) in Experiment III (in Experiments I and II the slider did not change color). The cue remained on the screen for additional 2 s (Experiment III: 1.5 s) and disappeared with the onset of the next ITI. See Fig 2. Three fractals out of a pool of twenty were assigned randomly to the three cue conditions (stable-low, stable-high, reversal). The background color was gray (rgb=[0.71, 071, 0.71]) and this stimulus occupied 9 degrees of visual angle. The rating scale was shown just below the fractal. Only the two ends of the 0% to 100% expectancy rating scale were labeled by ticks. The slider was initiated at a random position on each trial.

*Bonus trials*

In Experiments I and II, participants were occasionally presented with two of the cues and asked to select the one with either lower or higher probability of shock. Unbeknown to the participants, there was always one cue with currently low (i.e., stable-low or reversal in low-prob state) and one with high (i.e., stable-high or reversal in high-prob state) probability. In Experiment III, on a similarly small proportion of trials, participants could wager an amount between £0 and £5 to avoid a single shock on the next trial. Both tasks were introduced to keep participants engaged and to obtain an additional measure of value. Due to the different nature of the ratings, analysis of this data was not included in the present work.

*Task structure*

The task was characterized by changes in the contingency of the reversal cue ('switches') which

occurred in irregular intervals (see Supp. Table I). We use the term ‘state’ to refer to a section of the task during which the shock probability of the reversal cue was constant. Each participant experienced between 5 to 9 switches which results in 6 to 10 states per participant. States where the probability of the reversal cue was low are referred to as ‘low-prob state’ while states with high probability are called ‘high-prob state’. The number of trials and the dispersion of the switch point was slightly different in the three studies: Experiment I (M=30, +/- 2 trials), Experiment II (M=30, +/- 5 trials), Experiment III (M=35, +/- 10 trials). The mean refers to the total number of trials across all three cues. The proportion of stable-low, stable-high and reversal trials was as follows: Experiment I: 30%-30%-40%; Experiment II: 25%-25%-50%; Experiment III: 20%-20%-60%. This means that for example in Experiment III it was on average 0.6 * 35 = 21 reversal trials / state, although with higher variability (minimum 15 trials). Individual trials were presented in pseudorandomized orders. The schedules were generated as follows. First, the number of trials for a given state was determined (by ‘state’ we mean the period during which the reversal cue did not switch). Next, it was ensured that the contingency of each of the cues was within +/- 5% of the target. This means that if there were for example 40 trials in total, out of which 10 were stable-low, 10 were stable-high and 20 were the reversal cue, the target ‘objective’ probabilities were 25%, 75% and 75% (assuming the 75/25 condition and the reversal cue being in the high-state) it was ensured that the objective shock rates delivered for each cue were within +/- 5% of these contingencies, e.g., for the reversal cue there were between 14 (70%) and 16 (80%) shock trials within this ‘mini block’. Additionally, it was ensured that within each state a given cue was not presented on more than three subsequent trials. For the reversal cue, on the first five trials after reversal, at least three outcomes were in the new direction (if switch from high-to-low state just happened, at least three out of the first five trials ended with no-shock). Furthermore, once each mini-block passed the above criteria, the mini-blocks were assembled into a schedule. There was a slight difference between studies. While in studies I and II a change in state occurred with 75% probability (i.e. sometimes it didn’t happen), in Experiment III switch always happened. Lastly, a second +/- 5% contingency check was performed, this time across the entire trial schedule separately for each cue.

### Instructions

To minimize any influence of the experimenter, the information about the task was presented in writing. Only if the participant required further explanation, instructions were clarified verbally according to protocolled answers. Participants were presented with minimal information regarding the number of cues, task duration, cue frequency and switches. They were told that ‘each cue is associated with a certain probability of receiving a painful stimulus’ and to ‘pay attention to all three cues as any of them may or may not change their probability signaling the painful stimulation at any point’. For details on the instructions see Supplementary Materials (Instructions section).

### Questionnaires

Trait anxiety was assessed using the STAI-TRAIT (Spielberger, 2012). Additional study-specific personality measures were collected (e.g., pain-related fears and attitudes in Experiment I). For the complete list of questionnaires see Supplementary Materials, section ‘Questionnaires’. In Experiments I and III, questionnaires were completed using a computerized interface based on the LimeSurvey software. In Experiment II, pen & paper versions of the questionnaires were used.

## Data analyses

Data were analyzed using custom MATLAB (MATLAB 2019a, The MathWorks, Natick, 2019) and R 3.6.0. scripts (for a complete list of packages and versions see associated repository). All analysis scripts, data and reproducibility instructions are stored and documented in the associated github and GIN repositories.

### *Statistical and visualization approach*

Statistical analyses were performed using Linear Mixed Models (LMMs, as implemented in lmer 1.1-25 R package Bates et al., 2015) with study and participant included as an effect with a random intercept. For each analysis we included fixed effect of interest also as a random slope, and we performed a model comparison between the two version. Adding random slope didn't result in improved model fit in any of the analyses, so we didn't include them. Following ANOVA analysis of LMM results, post-hoc tests are reported using corrected p-value (Holm). Where variables were continuous (e.g., trait anxiety) they were included as such in the statistical models. To visualize data, we include raw data, summary statistics (mean or median), information about variance (standard error or interquartile range) and density (raincloud plots; Allen et al., 2019). In time-series plots (e.g., Fig 3a) we plot mean per condition and SEM (standard error of the mean).

### *Computational modeling*

All models were fitted to the trial-by-trial shock expectancy data using Bayesian Adaptive Direct Search (BADS; Acerbi & Ma, 2017) by minimizing the negative log likelihood of the data given a model. Our 1- and n-state models naturally use beta likelihood. To assess model fit across all trials, BIC (Schwarz, 1978) scores were calculated. To prevent convergence to local extremes, fitting was performed 45 times for each participant and cue, ensuring that computational resources were identical across all models.

## Measures

### *Behavioral measures*

All main analyses including model fitting are based on shock probability ratings (0% to 100%) provided by the participant on each trial. Additionally, each study contained either cue preference ratings or a shock wagering task to provide additional measure of shock expectancy. However, because the measures varied across studies, they are not included. Lastly, at the end of the task we collected visual and general liking ratings for each of the images used in the task. Participants were presented with the three fractals and asked to rate their 'visual appeal' and 'general liking' on a scale from 1 to 10.

### *Slope after reversal*

To calculate the average speed of updating after reversal, we fitted the shock probability ratings data on trials 1 to 10 (period of change) using a linear mixed effect model with estimated slope for each participant, session, half (early/late) and switch type (high-to-low, low-to-high). Due to convergence issues, such model was fitted separately for each session, half and switch type. The

estimates slopes for each participant/condition were then extracted from the models and analyzed separately using another LMM.

*Switch point and steepness*

To isolate subject-specific learning metrics, we used the trial-by-trial expectancy ratings to estimate switch point and switch steepness for each individual change in probability. Only trials from the reversal cue were included in this analysis. For each switch, we extracted 5 trials prior and 10 trials post reversal and demeaned the time-series. Using this 'chunk', we determined the point of steepest change ('switch point') by calculating the cumulative sum and identifying the point of highest (high-to-low switch) or lowest (low-to-high switch) value in the series.'' Next, a smaller chunk of 10 trials (5 preceding and 5 following) around the identified switch point was extracted. A sigmoidal curve (Eq. 1) with a free parameter for steepness was fitted to this smaller chunk. ***x*** corresponds to the time series of 10 trials, ***b*** is the inflection point which was fixed to ***b****=5.5* (midpoint) and ***a*** represents the steepness.

$$\text{Eq. 1 } \mathrm{f(x,a,c)} = \frac{1}{1+\mathrm{e}^{-\mathrm{a(x-b)}}}$$

The fitted value for ***a*** was recorded and log transformed before being used as an estimate for switch steepness in the main analysis.

*Error from true reinforcement*

To evaluate how much the individual learning time courses deviated from the delivered rates of shock, we calculated the running mean reinforcement rate (mean over shocks=1 and noshock=0 outcomes) for each state condition separately. This measure serves as an estimate of the true shock probability under the assumption that the agent knows which state they are in. To obtain a directional measure of error, the true reinforcement rate was subtracted from the expectancy ratings.

*Model-free learning rates in "meaningful" and "oddball" trials*

To obtain trial-wise learning rates, we rearranged the Rescorla-Wagner (Eq. 3) learning rule and calculated the trial-specific learning rate α (Eq. 4).

$$\text{Eq. 3 } P_{t+1} = P_t + \alpha_t\left(O_{i,t} - P_t\right)$$

$$\text{Eq. 4 } \alpha_t = \frac{P_{t+1}-P_t}{O_t-P_t}$$

$$\text{where } 0 \leq \alpha_t \leq 1$$

In some cases, such calculated learning rates became negative, for example, when the participant received a shock, but they lowered their expectation. In this instance, ratings were excluded from the analysis (assigned NaN values).

To distinguish between learning immediately after reversal, when learning rates should be relatively higher ("meaningful" learning), and later in stable periods of each state, when learning from surprising events should be relatively slower ("oddball" learning), we split model free learning rates at tenth trial after reversal. For example, if shock occurred in the first ten trials after low-to-

high switch then it was considered “meaningful” to learn from, while if it occurred after tenth trial of high-to-low switch it was considered an “oddball”.

## Computational models

Our primary goal was to provide a set of two models which used the same updating mechanisms and distributional assumptions (beta distribution), and that differed only in the ability to infer states. To model gradual learning and switching, we used a framework based on the beta distribution, similarly to de Boer et al., (2017) and Wise & Dolan, (2020). Our goal was to model the current shock probability estimate (ranging between 0 and 1) based on the received binary outcomes using the beta distribution. This approach is well-suited to model probabilities because beta distribution is bounded by 0 and 1. Additionally, it implicitly quantifies the amount of uncertainty about the current state. Lastly, this probability distribution naturally arises from binary outcomes. This provides a logical link between the outcomes delivered in the task (shock/no-shock) and the data reported by participants (probability estimates) but stands in contrast to more commonly used Normal distribution, which offers no straightforward mapping between binary outcomes and probability density.

*1-state model*

Each state was characterized by a beta distribution with parameters and α and β (Eqn. 5).

$$\text{Eq. 5} \qquad Beta_{PDF}(\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$$

Given this distribution, we assumed that the reported subjective probability of a shock reflected the mode (Eq. 6a) of the probability density function provided above, while state uncertainty was defined as standard deviation of the same distribution (Eq. 6b).

$$\text{Eq. 6a} \quad \widehat{\mathrm{P}} = \frac{\alpha-1}{\alpha+\beta-2}$$

$$\text{Eq. 6b} \quad \sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$$

Parameters $\alpha$ and $\beta$ can be thought of as proportional to the number of shocks and no-shocks received up until this point. As the sum of $\alpha$ and $\beta$ increases, the variance (and therefore state uncertainty) of the distribution decreases. In other words, the more evidence is available to the model, the more certain it is about its probability estimate. The starting values of $\alpha$ and $\beta$ are estimated as free parameters ($\alpha_0, \beta_0$, both $\in [1,10]$, values smaller than 1 were not included because in this case the distributions become bimodal, and Eqn. 6b). On each trial, the two parameters are updated by the amount equal to shock and no-shock attention weights $\tau^+$ or $\tau^-$(both $\in [0,2]$) depending on whether the shock was received (+) or omitted (-). Specifically, if

the cue was followed by a shock, then $\alpha$ is updated by the amount of $\tau^+$ and if no-shock occurred then $\beta$ is updated by $\tau^-$. Additionally, on each trial both $\alpha$ and $\beta$ are subject to decay $\lambda \in (0,1)$ (estimated in log space) which results in an increase in state uncertainty. This is the conceptual equivalent to forgetting. Lastly, the uncertainty of a given state is kept within realistic boundaries so that the sum of $\alpha$ and $\beta$ does not exceed 30. This is done to ensure numerical stability. See Eqs. 7 – 11.

*If outcome is shock ($O_t = 1$):*

$$\text{Eq. 7} \quad \alpha_{(t+1,s)} = \lambda\left(\alpha_{(t,s)} + \tau^+\right)$$

$$\text{Eq. 8} \quad \beta_{(t+1,s)} = \lambda\beta_{(t,s)}$$

*If outcome is not a shock ($O_t = 0$)*

$$\text{Eq. 9} \quad \beta_{(t+1,s)} = \lambda\left(\beta_{(t,s)} + \tau^-\right)$$

$$\text{Eq. 10} \quad \alpha_{(t+1,s)} = \lambda\alpha_{(t,s)}$$

*Both parameters of all non-active states decay*

$$\text{Eq. 11} \quad \alpha_{(t+1,s')} = \lambda\alpha_{(t,s')}$$

$$\beta_{(t+1,s')} = \lambda\beta_{(t,s')}$$

The 1-state model can behave very similarly to the more commonly used associative learning models such as the Pearce-Hall model (Li et al., 2011; Roesch et al., 2012), see Supp. Mat. For a comparison.

*Beta state inference model ('n-state')*

The 1-state model described above assumes a single state. However, alternatively we can let the model infer states from the data and allow for the possibility of switching between them. Specifically, our goal was for such "switching" model (a) to infer state switches from binary outcomes without any context cues, (b) to infer a state switch at a rate similar to humans, ideally in less than 10 trials and (c) to have a tendency to only create a handful of states to allow for meaningful generalization. The last aspect reflects the fact that every additional state (e.g., fear memory) must be maintained in parallel, but it also needs to be distinct from the already existing states. We note that a number of state switching models have been proposed previously (Blanco & Moris, 2018; Costa et al., 2015; Gershman et al., 2010; Redish et al., 2007), but none of these meets the goals set above.

As in the 1-state model, each state is characterized by a beta distribution that is updated as described above. In addition to updating each state, the model keeps track of the running average of surprise, ***S*** (see Eq. 12).

$$\text{Eq. 12} \quad S_{(t,s)} = (1-\pi)S_{(t-1,s)} + \pi\left|O_t - \widehat{P_{t,s}}\right|$$

In order to distinguish between inferring new states and state switching (it might be optimal to infer just two states but to switch between them multiple times, i.e., every time the contingencies change) the model uses two decision thresholds to guide behavior. The 'basic threshold' is defined by the uncertainty of the current state $\sigma$ times the threshold parameter $\eta$. Exceeding this threshold triggers a polling mechanism during which all existing states are compared against an expected value which is simply the mode $\hat{P}$ of the current state +/- the current running surprise ***S.*** Following this procedure, the most likely next state is switched to, or the current state is kept active. If the surprise ***S*** exceeds the compound threshold (Eq. 14), the model first checks for any existing states in the range around the expected value ( $\widehat{\mathrm{P_{t,s}}} +/- \mathrm{S_{t,s}}$) and if none exist it creates a new state. The compound threshold is additionally controlled by the parameter ***q*** which represents the difficulty of creating a new state. In order to allow new states to be created but to prevent the model from creating too many states, ***q*** follows a Chinese Restaurant Process distribution with parameters $\theta = 0.25$ and $\alpha = 1$ under which the creation of each next state becomes progressively more difficult.

Eq. 13 ('basic threshold') $S_{(t,s)} > \sigma_{(t,s)}\eta$

Eq. 14 ('compound threshold') $S_{(t,s)} > \sigma_{(t,s)}\eta q$

When a new state is being created it is initialized with mean at the current expected value ($P_{(t,s)} \pm S_{(t,s)}$) and standard deviation calculated using Eq. 6b from the estimated parameters $\alpha_0$ and $\beta_0$ (i.e., all states will have the same starting uncertainty).

Both models were found to recover well (see the 'Model recovery' section below for full description), providing support for a unique identifiability of the state switching strategy.
Models were also fitted to artificial data containing either one or two reinforcement levels, mimicking the stable and reversal cues from the actual task and the three contingency levels (Supp. Fig. 6). In stable environments (columns 1 and 3) the models were able to fit the data almost exactly. In environments with two reinforcement levels the 1-state model updated appropriately following contingency changes. The n-state model on the other hand was able to approximate high and low state and effectively switch between them

**Model recovery**

The 1-state and n-state models were included in a model recovery procedure. First, we fitted all models to the data of the participants. Second, we used the mean and standard deviation of the fitted parameter values to generate synthetic data. Third, the data generated using each model were fitted by each of the candidate models. The fitting procedure was identical to the one used to fit real participant data (45 runs, separate fit for each cue). Last, model comparison was performed for each artificial data set using the mean BIC as the quantitative criterion. A model was considered to recover well if the winning model matched the model used to generate the data. All investigated models recovered uniquely (see Supp. Fig. 7).

**Data quality and checks**

*Shock intensity and perception*

A linear mixed effects model (LMM) was estimated to check for the differences in shock intensity between studies and its relationship with trait anxiety. The mean shock intensity did not differ between studies, $F(2,79)=2.92$, *n.s.* nor was there a statistically significant interaction with trait anxiety $F(2,80.8)=0.02$, *n.s.* Kolmogorov-Smirnov tests found no difference in shock intensity values between full dataset and dataset after exclusions. Lastly, there was no association between shock intensity and reported probabilities in either low or high state indicating that probability ratings did not differ due to the participants' general sensitivity to electrical stimuli. We also tested whether the perceived shock unpleasantness and pain intensity correlated with true shock intensity or trait anxiety. If the calibration procedure had been successful, the objective shock intensity should not relate to the subjective ratings. Employing a LMM, we found no evidence for an association between the subjective painfulness/unpleasantness and the objective current or trait anxiety.

*Trait anxiety*

We tested whether anxiety scores differed between the three studies. While there was no statistically significant difference, $F(2,157)=1.35$, *n.s.*, the median TA in Experiment II was 35 compared to 42 in 1 and 41 in Experiment III. We therefore decided to include 'experiment' as a random effect in all linear mixed effect models. On occasions, anxiety results are shown as median split for convenience. Where possible, such plots are accompanied by parametric visualization. All statistics are performed using a full range of trait anxiety scores. Kolmogorov-Smirnov tests found no difference in anxiety scores between the full data set and the data set after exclusions.

*Cue appeal*

Although fractals were randomly allocated to the different conditions across participants, there was a possibility that participants would rate a specific fractal more favorably due to its visual appeal. To check whether this was the case, the visual appeal ratings collected at the end of the task were included as dependent variables in a LMM with cue and contingency condition as fixed effects. LMM found no significant effect of cue or contingency on visual appeal of the presented cues.

*Initial bias*

To test whether the first rating differed from the unbiased estimate of 0.5 indicating a pre-existing bias in shock expectancy, each participant's first rating of the first session was entered into a one-way t-tests (separately for each experiment). These analyses did not reveal any significant effect. Since there was a degree of variability around the mean we next tested for an association between trait anxiety and the first rating, but no relationship between the variables was found.

*Session order*

In Experiment III, the three contingency conditions were presented in a random order. To verify that our findings are not a result of an order effect, we used a LMM to test whether the session order had an influence on mean ratings separately for the high and low state of the reversal cue.

The model found no significant effect of the order in which the sessions were delivered on the probability ratings.

*Starting contingency of the reversal cue*

Next, we assessed whether ratings later in the task were influenced by the starting contingency (high vs low) of the reversal cue. To perform this analysis, we removed the first half of each time course (there would of course be an effect in the early ratings, here we are checking for any lasting anchoring bias) and fitted a LMM with state, contingency, experiment and starting contingency as fixed effects. There was no significant main effect or interaction of starting contingency. By adding trait anxiety to the model, we further checked whether there was any interaction with TA but found no effect of starting contingency on trait anxiety.

### Data and code

Data and code are published in the associated Github and GIN repositories.

## Acknowledgements

We would like to thank Judith Appel and Lorika Shkreli for help with data collection for Experiment II, and to Sam Hall-McMaster and Fabian Renz for comments on the manuscript. Funding: Medical Research Council, UK (scholarship to O.Z. and New Investigator grant MR/L011719/1 to K.W.), The Engineering and Physical Sciences Research, Council, UK (scholarship O.Z.), Max Planck Research Group Grant (awarded to N.W.S., supporting O.Z.), European Research Council Starting Grant (awarded to N.W.S.), MQ: Mental Health Research (awarded to A.R.). K.W. was supported by an MRC UK New Investigator grant (MR/L011719/1). M.B. was supported by the Oxford Health NIHR Biomedical Research Centre and Oxford Health Clinical Research Facility. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This research was funded in whole, or in part, by the Wellcome Trust (Grant number 203139/Z/16/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## Competing interests

MB has received travel expenses from Lundbeck for attending conferences, and has acted as a consultant for J&J, Novartis and CHDR.

## References

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 1836–1846). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/df0aab058ce179e4f7ab135ed4e641a9-Paper.pdf

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Ansari, T. L., & Derakshan, N. (2011). The neural correlates of impaired inhibitory control in anxiety. *Neuropsychologia*, *49*(5), 1146–1153. https://doi.org/10.1016/j.neuropsychologia.2011.01.019

Bartolo, R., & Averbeck, B. B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron*, *106*(6), 1044-1054.e4. https://doi.org/10.1016/j.neuron.2020.03.024

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Blanco, F., & Moris, J. (2018). Bayesian methods for addressing long-standing problems in associative learning: The case of PREE. *Quarterly Journal of Experimental Psychology*, *71*(9), 1844–1859. https://doi.org/10.1080/17470218.2017.1358292

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. https://doi.org/10.1016/S0006-3223(02)01546-9

Bouton, M. E. (2004). Context and Behavioral Processes in Extinction: Table 1. *Learning & Memory*, *11*(5), 485–494. https://doi.org/10.1101/lm.78804

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*(4), 445–466. https://doi.org/10.1016/0023-9690(79)90057-2

Brooks, D. C., & Bouton, M. E. (1993). A retrieval cue for extinction attenuates spontaneous recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*(1), 77–89. https://doi.org/10.1037/0097-7403.19.1.77

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal Learning and Dopamine: A Bayesian Perspective. *Journal of Neuroscience*, *35*(6), 2407–2416. https://doi.org/10.1523/JNEUROSCI.1989-14.2015

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Davis, M., Falls, W. A., & Gewirtz, J. (2000). Neural Systems Involved in Fear Inhibition: Extinction and Conditioned Inhibition. In M. S. Myslobodsky & I. Weiner (Eds.), *Contemporary Issues in Modeling Psychopathology* (Vol. 1, pp. 113–141). Springer US. https://doi.org/10.1007/978-1-4757-4860-4_8

de Boer, L., Axelsson, J., Riklund, K., Nyberg, L., Dayan, P., Bäckman, L., & Guitart-Masip, M. (2017). Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age. *ELife*, *6*, e26424. https://doi.org/10.7554/eLife.26424

Dibbets, P., van den Broek, A., & Evers, E. A. T. (2015). Fear conditioning and extinction in anxiety- and depression-prone persons. *Memory*, *23*(3), 350–364. https://doi.org/10.1080/09658211.2014.886704

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. P. (2015). UPDATED META-ANALYSIS OF CLASSICAL FEAR CONDITIONING IN THE ANXIETY DISORDERS: Review: Updated Meta-Analysis of Fear Conditioning in Anxiety Disorders. *Depression and Anxiety*, *32*(4), 239–253. https://doi.org/10.1002/da.22353

Dunsmoor, J. E., Niv, Y., Daw, N., & Phelps, E. A. (2015). Rethinking Extinction. *Neuron*, *88*(1), 47–63. https://doi.org/10.1016/j.neuron.2015.09.028

Dunsmoor, J. E., & Paz, R. (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biological Psychiatry*, *78*(5), 336–343. https://doi.org/10.1016/j.biopsych.2015.04.010

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biological Psychology*, *92*(2), 342–352. https://doi.org/10.1016/j.biopsycho.2012.11.006

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197–209. https://doi.org/10.1037/a0017808

Gershman, S. J., & Hartley, C. A. (2015). Individual differences in learning predict the return of fear. *Learning & Behavior*, *43*(3), 243–250. https://doi.org/10.3758/s13420-015-0176-z

Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: Implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00164

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & Behavior*, *40*(3), 255–268. https://doi.org/10.3758/s13420-012-0080-8

Haaker, J., Lonsdorf, T. B., Schümann, D., Menz, M., Brassen, S., Bunzeck, N., Gamer, M., & Kalisch, R. (2015). Deficient inhibitory processing in trait anxiety: Evidence from context-dependent fear learning, extinction recall and renewal. *Biological Psychology*, *111*, 65–72. https://doi.org/10.1016/j.biopsycho.2015.07.010

Indovina, I., Robbins, T. W., Núñez-Elizalde, A. O., Dunn, B. D., & Bishop, S. J. (2011). Fear-Conditioning Mechanisms Associated with Trait Vulnerability to Anxiety in Humans. *Neuron*, *69*(3), 563–571. https://doi.org/10.1016/j.neuron.2010.12.034

Kindt, M., & Soeter, M. (2014). Fear Inhibition in High Trait Anxiety. *PLoS ONE*, *9*(1), e86462. https://doi.org/10.1371/journal.pone.0086462

Lamba, A., Frank, M. J., & FeldmanHall, O. (2020). Anxiety Impedes Adaptive Social Learning Under Uncertainty. *Psychological Science*, *31*(5), 592–603. https://doi.org/10.1177/0956797620910993

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10), 1250–1252. https://doi.org/10.1038/nn.2904

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V. L., Meir Drexler, S., Mertens, G., Richter, J., Sjouwerman, R., Wendt, J., & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning research. *ELife*, *8*, e52465. https://doi.org/10.7554/eLife.52465

Morís, J., Barberia, I., Vadillo, M. A., Andrades, A., & López, F. J. (2017). Slower reacquisition after partial extinction in human contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 81–93. https://doi.org/10.1037/xlm0000282

Muris, P., Mayer, B., & Merckelbach, H. (1998). TRAIT ANXIETY AS A PREDICTOR OF BEHAVIOUR THERAPY OUTCOME IN SPIDER PHOBIA. *Behavioural and Cognitive Psychotherapy*, *26*(1), 87–91. https://doi.org/10.1017/S1352465898000095

Myers, K. M., & Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry*, *12*(2), 120–150. https://doi.org/10.1038/sj.mp.4001939

Norbury, A., Brinkman, H., Kowalchyk, M., Monti, E., Pietrzak, R. H., Schiller, D., & Feder, A. (2021). Latent cause inference during extinction learning in trauma-exposed individuals with and without PTSD. *Psychological Medicine*, 1–12. https://doi.org/10.1017/S0033291721000647

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neuroscience & Biobehavioral Reviews*, *114*, 96–112. https://doi.org/10.1016/j.neubiorev.2020.04.019

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex.* (pp. xv, 430). Oxford Univ. Press.

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction Learning in Humans. *Neuron*, *43*(6), 897–905. https://doi.org/10.1016/j.neuron.2004.08.042

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587. https://doi.org/10.1038/s41467-021-26731-9

Ploner, M., Lee, M. C., Wiech, K., Bingel, U., & Tracey, I. (2010). Prestimulus functional connectivity determines pain perception in humans. *Proceedings of the National Academy of Sciences*, *107*(1), 355–360. https://doi.org/10.1073/pnas.0906186106

Raes, A. K., De Raedt, R., Verschuere, B., & De Houwer, J. (2009). Failure to loose fear: The impact of cognitive load and trait anxiety on extinction. *Behaviour Research and Therapy*, *47*(12), 1096–1101. https://doi.org/10.1016/j.brat.2009.08.002

Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review*, *114*(3), 784–805. https://doi.org/10.1037/0033-295X.114.3.784

Rescorla, R. A., & Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*(1), 88–96. https://doi.org/10.1037/0097-7403.1.1.88

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory* (pp. 64–99). Appleton-Century-Crofts, New York, NY.

Rodriguez, B. I., Craske, M. G., Mineka, S., & Hladek, D. (1999). Context-specificity of relapse: Effects of therapist and environmental context on return of fear. *Behaviour Research and Therapy*, *37*(9), 845–862. https://doi.org/10.1016/S0005-7967(98)00106-5

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain: Neural

correlates of RW and PH. *European Journal of Neuroscience*, *35*(7), 1190–1200. https://doi.org/10.1111/j.1460-9568.2011.07986.x

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., & Phelps, E. A. (2008). From Fear to Safety and Back: Reversal of Fear in the Human Brain. *Journal of Neuroscience*, *28*(45), 11517–11525. https://doi.org/10.1523/JNEUROSCI.2265-08.2008

Scholz, V., Algermissen, J., Rostami Kandroodi, M., & den Ouden, H. (2021). *Too much, too little or just right? The clinical relevance of cognitive control over motivational biases in decision-making.* Symposium on Biology and Decision-Making. https://www.youtube.com/watch?v=rnI4oqYVlno

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., Zwitserlood, P., Schiffbauer, H., Heindel, W., Arolt, V., & Konrad, C. (2011). Neural correlates of trait anxiety in fear extinction. *Psychological Medicine*, *41*(4), 789–798. https://doi.org/10.1017/S0033291710001248

Sjouwerman, R., Scharfenort, R., & Lonsdorf, T. B. (2020). Individual differences in fear acquisition: Multivariate analyses of different emotional negativity scales, physiological responding, subjective measures, and neural activation. *Scientific Reports*, *10*(1), 15283. https://doi.org/10.1038/s41598-020-72007-5

Spielberger, C. D. (2012). *State-Trait Anxiety Inventory for Adults* [Data set]. American Psychological Association. https://doi.org/10.1037/t06496-000

Staples-Bradley, L. K., Treanor, M., & Craske, M. G. (2018). Discrimination between safe and unsafe stimuli mediates the relationship between trait anxiety and return of fear. *Cognition and Emotion*, *32*(1), 167–173. https://doi.org/10.1080/02699931.2016.1265485

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Caseras, X., Andión, O., & Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biological Psychology*, *92*(2), 185–190. https://doi.org/10.1016/j.biopsycho.2012.10.006

Usala, P. D., & Hertzog, C. (1991). Evidence of differential stability of state and trait anxiety in adults. *Journal of Personality and Social Psychology*, *60*(3), 471–479. https://doi.org/10.1037/0022-3514.60.3.471

Vaidya, A. R., Jones, H. M., Castillo, J., & Badre, D. (2021). Neural representation of abstract task structure during generalization. *ELife*, *10*, e63226. https://doi.org/10.7554/eLife.63226

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology*, *9*(1), 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Wise, T., & Dolan, R. J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications*, *11*(1), 4179. https://doi.org/10.1038/s41467-020-17977-w

Woods, A. M., & Bouton, M. E. (2007). Occasional reinforced responses during extinction can slow the rate of reacquisition of an operant response. *Learning and Motivation*, *38*(1), 56–74. https://doi.org/10.1016/j.lmot.2006.07.003

Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience & Biobehavioral Reviews*, *128*, 270–281. https://doi.org/10.1016/j.neubiorev.2021.06.024