High trait anxious individuals represent aversive environment as multiple states: a computational mechanism behind reinstatement

Ondrej Zika (zika@mpib-berlin.mpg.de)[1,2]

Katja Wiech (katja.wiech@ndcn.ox.ac.uk)[2]

Nicolas W. Schuck (schuck@mpib-berlin.mpg.de)[1]

1 NeuroCode Research Group, Max Planck Institute for Human Development, Lentzealle 94, Berlin, 141 95, Germany
2 Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, OX39DU, Headley Way, Oxford

Abstract

Learning the likelihood of aversive events is achieved either by gradual learning or via inference of hidden states. We previously linked the tendency towards state switching to trait anxiety but the effect of environmental noise has not been investigated. In the present study we employ a Pavlovian probabilistic learning paradigm to test how environmental noise promotes either state switching or gradual lerning. Participants completed three sessions varying in shock contingency jumps (60/40%, 75/25% or 90/10%). As a signature of state-switching we analyzed steepness of post-reversal switch. In support of our hypothesis we found that steepest switches were present in the 90/10 environment. This effect was found to be driven by high trait anxiety. Trait anxiety also positively correlated with difference between acquisition and extinction. Next, we developed a state switching model and performed model comparison using cross-validation. Analysis of model parameters found positive correlation between trait anxiety and tendency to create more states. In summary, our behavioural and modelling result show that less noisy environments encourage state switching, and that anxious individual have an increased tendency to represent the environment as multiple states. This result highlights trait anxiety as vulnerability in successful extinction treatment.

Keywords: aversive learning; fear; extinction; trait anxiety; computational modelling

Introduction

Learning Strategies

Accurately predicting future threat is a key survival mechanism. In noisy environments an agent can either follow a gradual learning strategy (e.g. reinforcement learning; RL) or discover an underlying task structure and represent it as state space (structure learning). Gradual learning ensures that when faced with a large violation of expectation the agent does not change its belief too drastically which would lead to large errors in prediction. To accelerate adaptive behaviour learning rate can be increased when large prediction errors are generated (Pearce & Hall, 1980). This makes RL a suitable strategy when exploring a new environment. However, over longer periods of time gradual updating may become ineffective and computationally wasteful. As an alternative, an agent might learn that there is an underlying structure in the environment (e.g. shock reinforcement rate switches every 20 trials) which, when correctly learned, decreases the amount of energy used and increases prediction accuracy. This has been referred to as structure learning (Gershman, Blei, & Niv, 2010).

Hidden states are often discovered as a consequence of gradually sampling the environment, however, the exact mechanism of state discovery is not known. (Gershman, Jones, Norman, Monfils, & Niv, 2013) proposed that consistent large prediction errors will lead to the creation of a new state. An agent in a novel environment will initially employ a reinforcement learning strategy but when a consistent structure is discovered it will lead to representation of different clusters of features as states (Redish, Jensen, Johnson, & Kurth-Nelson, 2007). An example of such learning is the case of context where the agent learns that there is one time period in which shocks are frequent and another one in which they are infrequent. Once this becomes known learning essentially becomes a state classification process.

Aversive Learning and Trait Anxiety

In aversive learning, phases of high shock frequency (acquisition) and subsequent passages of low frequency of shock rate (extinction) have been reported to lead to asymmetric learning. In a recent work we showed that while acquisition of aversive associations is fast and lasting, extinction is characterized by slower and incomplete learning. Trait anxiety has previously been associated with difficulty to inhibit fear (Kindt & Soeter, 2014) and increased physiological and neural reactivity to fearful stimuli (Indovina, Robbins, Nez-Elizalde, Dunn, & Bishop, 2011). During aversive learning, this leads to lack of fear extinction. Anxiety has also been associated with increased likelihood to relapse post extinction (Lissek et al., 2005). In our recent data sets we showed that high anxious individuals have a tendency to represent the learning environment as multiple distinct states rather than to learn gradually.

Hypotheses

In this paper we investigate whether the magnitude of contingency changes influences the learning strategy employed by

participants and whether this is further modulated by trait anxiety. We hypothesize that state highly noisy environments will encourage gradual learning while environments low in noise and high in contingency difference will preferentially lead to multi-state representation of the environment. Further, we hypothesise that anxious individuals will have increased tendency to represent the environment as multiple states which could be a mechanism leading to higher rates of relapse.

Methods

Task and Design

Thirty-three (15 female, $\mu_{age} = 25.4, \sigma = 4.5$) participants completed three sessions of probabilistic aversive learning task while were recorded their shock expectancy ratings. On each trial a cue was presented and either followed by a shock or shock omission. Before the outcome was delivered participants were asked to submit an expectancy rating indicating how likely do they think that a shock will follow. In each session there were three cues, two stable and one reversal cue. One of the stable cues was associated with high probability of shock ("harmful cue") whilst the other had a low probability of being followed by a shock ("safe cue"). The "reversal cue" was irregularly (i.e. every $\mu = 35, \sigma = 5$ trials) switching between blocks of high and low probability of shock. In each session the high and low levels varied as follows: 40/60%, 25/75% 10/90%, where the first number represents shock probability in the low probability state and the second in the high state. The design is shown in Figure 1. Sessions were presented in pseudo-random order across participants.
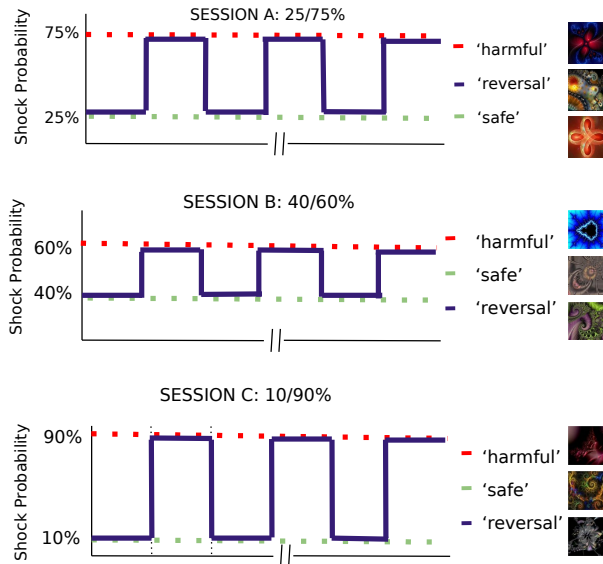

Figure 1: **Task Design** *consisted of three sessions with reinforcement contingencies set a t 40/60% [session A], 25/75% [session B], 10/90% [session C]*

Behavioural Measures

Shock Expectancy Ratings Shock expectancy ratings were collected on every trial using a continuous (0% to 100%).

Post-reversal Switch Point and Switch Steepness To estimate the steepest point of learning post reversal for each individual flip in contingencies, ten pre- and ten post-reversal trials were extracted and their mean subtracted from each data point. A cumulative sum was then calculated. Its extreme was then taken as the steepest learning point (Page, 1954) which we call "switch point" as it represents the point of steepest learning. Once the switch point was identified five pre and post trials were extracted and a sigmoidal function fitted to data segment. The steepness of the function was then used in the behavioural analysis as "switch steepness".

Computational Models

All models were fitted by minimising the negative log likelihood using the BADS algorithm (Acerbi & Ma, 2017). To dissociate between gradual learning and state switching we specified three learning models: Rescorla-Wagner (RW), Pearce-Hall (PH) and beta state switcher.

Rescorla-Wagner
A standard version of the RW algorithm was used. A probability P is updated on each trial t by the difference between current shock expectancy P_t and received outcome O_t (shock or noshock), weighted by the learning rate α.

$$P_{t+1} = P_t + \alpha(O_t - P_t) \tag{1}$$

Pearce-Hall
The PH model extends the RW by introducing a dynamic learning rate which is contingent on the magnitude of recent errors, a quantity known as associability. Consequently, learning is faster when larger errors are generated. PH uses the same equation as RW (Eq. 1) to update probability. Additionally, the learning rate for each trial is a combination the current associability η_t scaled by κ (Eq. 3). Associability a combination of the most recent unsigned prediction error and a previous associability value weighted by π (Eq. 2). i is an index for the two outcomes.

$$\eta_t = (1 - \pi_i)\eta_{(t-1)} + \pi_i|O_t - P_t| \tag{2}$$

$$\alpha_t = \kappa\eta_t \tag{3}$$

Beta State Learner
To capture cases where participants represent the environment as multiple states we developed a novel model based on the leaky beta model recently used by (Wise, Michely, Dayan, & Dolan, in submission). The value predicted by the agent on each trial is the mean of the beta distribution given the parameters $\alpha_{s,t}$ and $\beta_{s,t}$. The current expectancy and state uncertainty estimates are given mean and standard deviation by the beta distribution (Eqs. 4 and 5)

$$\mu_s = \frac{\alpha}{\alpha+\beta} \quad (4) \qquad \sigma_s^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad (5)$$

Updating occurs by adding 1 if $\alpha_{s,t}$ if shock has occurred or to $\beta_{s,t}$ if shock was omitted. To account for differential learning from shock and no-shock outcomes, an additional constant is added for each outcome type. I define this constant as $\tau^+ \in [1, 1]$ for shock and $\tau^- \in [1, 1]$ for no-shock. These can be loosely interpreted as attention weights and they provide the model with the ability to dissociate between shock and no-shock updating. Due to the nature of beta distribution the more outcomes the agent has experienced, the more certain it is about the probability of an outcome on the next trial. In a changing and noisy environment, this model with no free parameters converges to the mean. To make the model sensitive to recent experience a decay parameter $\lambda \in [0, 1]$ was introduced. Equations 6 and 7 show the full updating scheme.

$$\beta_{s,t+1} = \begin{cases} \lambda\beta_{s,t}, \\ \lambda(\beta_{s,t}+1+\tau^-), \end{cases} \quad \alpha_{s,t+1} = \begin{cases} \lambda(\alpha_{s,t}+1+\tau^+), & \text{if shock} \\ \lambda\alpha_{s,t}, & \text{if no-shock} \end{cases}$$

(6) (7)

The model also keeps track of current level of surprise I, similarly to PH (Eq. 8) α_{surp} corresponds to surprise learning rate and λ_{surp} corresponds to surprise decay.

$$I_t = I_{t-1} + \alpha_{surp}(O_t - P_t) - \lambda_{surp}I_{t-1} \qquad (8)$$

Finally, after each outcome is delivered the agent decides whether to stay in the current state or whether to create a new state / switch to an existing state by the decision rule in Eq. 9.

$$I_t > \sigma_{s,t} * \eta * max(S) \qquad (9)$$

where $\sigma_{s,t}$ is the uncertainty of the current state, η is an individual threshold for switching/creating states and S is the existing number of states.

If the level of surprise exceeded the threshold in Eq. 9 and there is a suitable state to switch to then a switch is performed. This is done by adding surprise (I_t) to the current shock expectancy the current state ($P_{S_{current}}$), calculating the expected next value $\hat{P}$ and testing whether it lies in any state's expected range $\eta\sigma_{S_s,t}$. If no candidate state matched a new state is created using the the current state expected value, and initialized with $\sigma_{S_{new}} = 0.29$ which is the standard deviation of the beta distribution with parameters $\alpha = 1, \beta = 1$.

Results

Mean Ratings

Participants were able to on average learn the underlying contingencies. While in the safe and harmful cues the mean ratings were relatively accurate, there was a slight overall overprediction in the reversal cue (see Figure 2). Splitting the reversal cue into acquisition and extinction, it was apparent that

this was driven by lack of extinction in all three sessions. Interestingly, there was no difference between mean probabilities in extinction even between the most extreme conditions (40/60% and 90/10%).
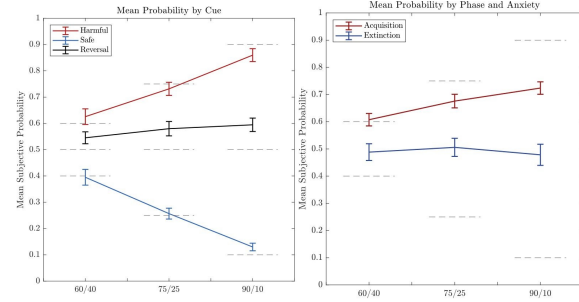

Figure 2: **Mean Probabilities** *by cue (left) and phase (right*

Investigating the role of trait anxiety (TA), the difference between acquisition and extinction positively correlated with TA $r(32) = .439, p = 0.12$, supporting our previous finding that high TA individuals are better at dissociating between the phases. This effect was strongest in the 90/10 condition and weakest in the 60/40 condition.
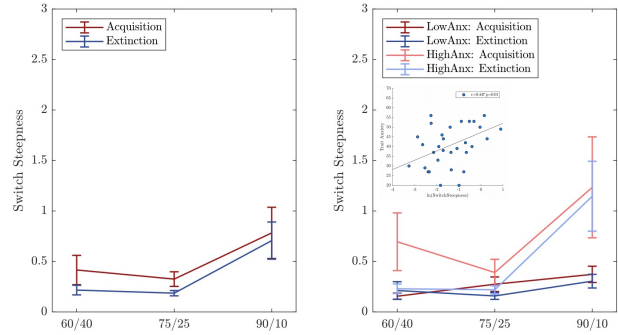

Figure 3: **Switch steepness** *left plot shows switch steepness by visit, right plot shows that the effect is driven by high trait anxiety group.*

Switch Steepness

Switch steepness was significantly higher in the 90/10 condition than in any of the other two. Importantly, there was no difference between 60/40 and 75/25 suggesting that state switching only starts occuring at contingency differences bigger than 50%.

Including anxiety, increase in switch steepness in the 90/10 condition was driven by the high TA group. Trait anxiety significantly correlated with switch steepness overall, $r(32) = 0.447, p = 0.01$, post-hoc test revealing that this was driven by the 90/10 condition, $r(32) = 0.494, p_{corr} = 0.012$ (Figure 3).

Modelling Results

Model Comparison Model comparsion found that the beta state learner fitted the data best in the 60/40 and the 90/10 condition while in the 75/25 condition the Pearce-Hall fitted best. These results suggest that state learning occurs in the 90/10 condition but not as often in the 75/25 condition. The BSL also fitted well in the 60/40 condition as it captures RW-like gradual learning in one state.
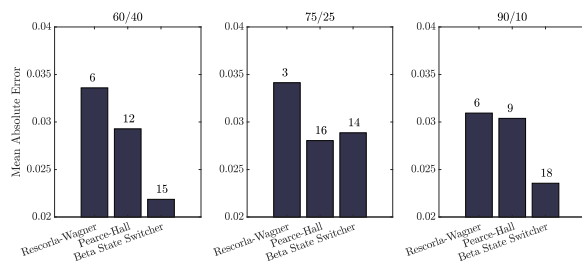

Figure 4: **Model Comparison** *data were fitted to initial 60-80% data points and model used to predict the remainder. The mean absolute error corresponds to average sum of least squares errors over twenty runs. Numbers above bars represent number of participants best fitted by the given model.*

Winning Model Parameters Analysis of the parameters of the winning model found a significant correlation between number of states estimated for each participants and the behavioural measure of switch steepness $r(32) = 0.47, p = 0.006$. Furthermore, there was a negative correlation in the 90/10 condition between the η parameter and trait anxiety, $r(32) = -0.4, p = 0.032$. This shows that trait anxiety is associated with lower threshold for state switch/state creation.

Discussion

In a behavioural and computational analysis of human aversive learning we found that less noisy environment with high contingency difference lead to steeper switches between them. There was no progressive increase in switch steepness from high though mid to low noise conditions, which suggests that this is not due to mere contingency difference but that there is something fundamentally different in the 90/10 condition. We propose that this is due to state rather than gradual learning in the 90/10 condition. Our modelling results show that the state switching strategy dominates in the 90/10 condition, further supporting the notion that low noise environment encourages state rather than gradual learning.

In relation to anxiety, our data show that high trait anxious individuals can better distinguish between acquisition and extinction and that it is the high anxiety group that drives the tendency for state learning in the 90/10 condition, as shown by group ANOVA and significant correlation between switch steepness and trait anxiety. Furthermore, the critical parameter of the beta state learner η that controls the tendency to

switch or create states correlated negatively with trait anxiety, providing i) further evidence for state learning in high TA and b) proposing a mechanism driving the effect. In summary, our data show that low noise environments lead to state learning and that high trait anxious individuals have increased tendency for multi-state represetnation of the world. This finding has important implication for anxiety disorders, proposing a potential mechanism for increased tendency for fear relapse in anxiety.

References Instructions

References

Acerbi, L., & Ma, W., Ji. (2017). Practical bayesian optimization for model fitting with bayesian adaptive direct search. In I. Guyon et al. (Eds.), *Advances in neural information processing systems 30* (pp. 1836–1846). Curran Associates, Inc.

Gershman, S., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological review*, *117*(1), 197–209.

Gershman, S., Jones, C., Norman, K., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, *7*, 164.

Indovina, I., Robbins, T. W., Nez-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.

Kindt, M., & Soeter, M. (2014). Fear inhibition in high trait anxiety. *PLOS ONE*, *9*(1), 1-6.

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391 - 1424.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

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), 784805.

Wise, T., Michely, J., Dayan, P., & Dolan, R. J. (in submission). A computational account of threat-related attentional bias.