Modelling avoidance in mood and anxiety disorders using reinforcement learning of avoidance in mood and anxiety

Anahit Mkrtchian, Jessica Aylward, Peter Dayan, Jonathan P. Roiser, Oliver J. Robinson

PII: S0006-3223(17)30091-4
DOI: http://dx.doi.org/10.1016/j.biopsych.2017.01.017
Reference: BPS13111

To appear in: Biological Psychiatry

Cite this article as: Anahit Mkrtchian, Jessica Aylward, Peter Dayan, Jonathan P. Roiser and Oliver J. Robinson, Modelling avoidance in mood and anxiety disorders using reinforcement learning of avoidance in mood and anxiety, Biological Psychiatry, http://dx.doi.org/10.1016/j.biopsych.2017.01.017

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Modelling avoidance in mood and anxiety disorders using reinforcement-learning

Anahit Mkrtchian¹, Jessica Aylward¹, Peter Dayan², Jonathan P Roiser¹ and Oliver J Robinson¹*

¹Institute of Cognitive Neuroscience, University College London, London WC1N 3AZ
²Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG

*corresponding author (oliver.j.robinson@gmail.com)

Short title: Reinforcement learning of avoidance in mood and anxiety

Key words: Anxiety, avoidance, reinforcement learning, threat of shock, Pavlovian bias, diathesis-stress

Acknowledgements and Disclosures

OJR conceived the experiment. OJR and AM designed the study with critical input from JPR. AM programmed the task. AM and JA screened participants, collected the data and performed analyses under the supervision of OJR. OJR performed the computational modelling under the supervision of PD. OJR wrote the paper with critical input from JPR, PD, JA and AM. We thank Quentin Huys and Woo-Young Ahn for parameter fitting code and assistance. This research was funded by a Medical Research Council Career Development Award to OJR (MR/K024280/1) and a Medical Research Foundation Equipment Competition Grant (C0497, Principal Investigator OJR). PD is supported by the Gatsby Charitable Foundation. The authors report no biomedical financial interests or potential conflicts of interest.
Abstract

Background

Serious and debilitating symptoms of anxiety are the most common mental health problem worldwide, accounting for around 5% of all adult ‘years lived with disability’ in the developed world. Avoidance behaviour – avoiding social situations for fear of embarrassment, for instance – is a core feature of such anxiety. However, as for many other psychiatric symptoms, the biological mechanisms underlying avoidance remain unclear.

Methods

Reinforcement-learning models provide formal and testable characterizations of the mechanisms of decision-making; here, we examine avoidance in these terms. One hundred and one healthy and individuals with mood and anxiety disorders completed an approach-avoidance go/no-go task under stress induced by threat of unpredictable shock.

Results

We show an increased reliance in the mood and anxiety group on a parameter of our reinforcement-learning model that characterizes a prepotent (Pavlovian) bias to withhold responding in the face of negative outcomes. This was particularly the case when the mood and anxiety group was under stress.

Conclusions

This formal description of avoidance within the reinforcement-learning framework provides a new means of linking clinical symptoms with biophysically plausible models of neural circuitry and, as such, takes us closer to a mechanistic understanding of mood and anxiety disorders.
Introduction

Avoidance is a core feature of anxiety(1, 2) and plays a central role in psychological strategies for the treatment of anxiety(3), but its underlying neural and cognitive mechanisms are unknown. Avoidance can be adaptive: if an individual perceives a situation as stressful then it makes sense to avoid that stressor in the future. However, excessive avoidance can result in a pathological downward-spiral. The more one avoids a situation, the less opportunity there is to learn that the situation is not as bad as feared, and a vicious cycle of avoidance and impaired extinction learning emerges, which in turn promotes further anxiety(1). For example, an individual who fears social embarrassment might ultimately end up housebound, avoiding all social interaction.

The diathesis-stress model of mood and anxiety disorders(4) proposes that maladaptive avoidance should be greatest during periods of environmental stress in vulnerable individuals. This idea has clear face-validity, and is supported by clinical anecdote, but is largely derived from retrospective, subjective self-report. This is because quantifying avoidance under stress in an experimentally controlled yet ecologically valid manner in humans is methodologically challenging. In this study we address this challenge using: i) a translationally-validated (i.e. comparable behavioral responses can be elicited across human and animal models(5)) ‘threat of shock’ procedure to induce stress(6, 7); ii) a cognitive task that has been shown to reliably index avoidance behaviour in healthy individuals(1); and iii) a computationally precise method of defining of avoidance.

Specifically, we operationalize avoidance as a behavioural bias towards withholding action ("no-go", i.e. inhibition) in the face of potentially negative outcomes. This powerful prepotent reflexive (or Pavlovian) bias has been observed consistently in humans and animals(8-11) and is so profound that it can disrupt instrumental goal-directed behaviour(8-11). This is known as
Pavlovian-Instrumental transfer (12), and we harness it here to measure the degree to which individuals rely on their prepotent avoidance biases. Given that both induced stress (13, 14) and pathological anxiety have been associated with increased inhibitory control, it seems plausible that a combination of stress and anxiety will increase reliance on Pavlovian inhibitory avoidance biases (15) (in contrast with depression alone which might plausibly be associated with reduced reliance on Pavlovian approach biases (16)).

Reinforcement-learning algorithms can provide parameterizations of avoidance behaviour that offer insight into both optimal behaviour when set correctly (17), and to dysfunction and pathology when set incorrectly (18). Critically, reinforcement-learning models enable us to parameterize the influence of Pavlovian avoidance biases on task performance in a formal manner. A large body of work has applied these models to healthy humans (8-10), and they form the basis of human-level artificial intelligence (17), but to date they have not been applied to individuals with mood and anxiety disorders.

We therefore tested individuals with mood and anxiety disorders and healthy individuals completing an approach- avoidance go/no-go task under stress, which was induced by threat of shock. Avoidance was defined and parameterised within a reinforcement-learning framework. We predicted that the disordered group would show high reliance on avoidance bias, and that this would be exacerbated by stress.
Materials and Methods

Participants

All data, task scripts and code to recreate the figures in this paper are freely available online*. A total of 101 participants were included in the study. Healthy participants (N=58 (originally N=62 but four excluded because they failed to follow task instructions); 36 male [62.1%]; age range: 18-57; mean (standard deviation) age=26.7 (7.1)) and unmedicated individuals suffering from pathological mood and anxiety symptoms (N=43; 27 male [62.8%]; age range 18-53; mean age=28.8 (8.8)) were recruited from online advertising and institutional subject databases. The primary difference between the groups in initial recruitment was that only the pathological group self-defined as experiencing distress from mood/anxiety symptoms. We recruited a mixed sample of anxiety and depression diagnoses because they are highly comorbid with overlapping symptoms and may not therefore represent truly distinct pathologies. Healthy participants responded to an advertisement asking for healthy individuals with no psychiatric symptoms. A phone screen confirmed no history of psychiatric, neurological or substance use disorders. The disordered group responded to an advertisement for individuals suffering with low mood, anxious or depressive symptoms. Following an initial phone screen, individuals who met criteria for mood or anxiety disorder symptomatology according to a face-to-face Mini International Neuropsychiatric Interview (MINI) (19) were included. According to the MINI, the majority of participants (N=27) met criteria for both GAD and MDD (N=9 with additional panic disorder), N=8 met criteria for GAD (N=3 with panic disorder, N=1 with agoraphobia), N=2 Panic disorder and MDD, and a further N=6 MDD alone (table S1). The average number of depressive episodes was 5 (standard deviation ±7). The average duration of episodes was 7 months

* https://figshare.com/articles/Avoidance_Anxiety_Materials/3860250
(standard deviation ±8; excluding one participant who reported continuous episode since adolescence). Further details are provided in the supplement.

**Manipulation**

State anxiety was induced via threat of unpredictable electric shocks delivered with two electrodes attached to the non-dominant wrist using a Digitimer DS5 Constant Current Stimulator (Digitimer Ltd, Welwyn Garden City, UK). A highly unpleasant (but not painful) subjective shock level was established using a shock work-up procedure prior to testing. No more than five (to avoid habituation) shocks with gradual increasing shock level were administered. Participants rated each shock on a scale from 1 (barely felt) to 5 (unbearable). Shock level was matched at a level of 4 across participants. The experimental task was programmed in Psychtoolbox (http://psychtoolbox.org) for MATLAB 2014 (The MathWorks Inc., Natick, MA), presented on a laptop and administered under alternating safe and threat blocks. During the safe block, the background colour was blue and proceeded by a 4000ms message stating: “YOU ARE NOW SAFE FROM SHOCK”. During the threat block, the background colour was red and the message: “WARNING! YOU ARE NOW AT RISK OF SHOCK” was presented for 4000ms. Participants were told that they might receive a shock only during the threat condition but that the shocks were not dependent on their performance. In practice, a single shock was delivered at a pseudorandom timepoint during one-third of threat blocks (a total of four shocks across 480 trials). Note that it is the anticipation of these shocks, not the shocks themselves that constitutes the manipulation (see supplemental analysis). At the end of each experimental task, participants retrospectively rated how anxious they felt during the safe and threat conditions on a scale from 1 (“not at all”) to 10 (“very much so”).
**Approach-Avoidance Task**

The task was based on the design of a previous probabilistic go/no-go reinforcement learning task (10, 20) modified to incorporate the threat manipulation. The prepotent Pavlovian bias to a win is a go response (approach) and the prepotent Pavlovian response to a loss is no-go (avoid). As such, the task comprised four experimental conditions where action (go/no-go) was crossed with valence (reward/punishment): 1) go to win reward (GW), 2) go to avoid losing (GA), 3) no-go to win reward (NGW), and 4) no-go to avoid losing (NGA). On each trial, participants were presented with one of four fractal cues per condition, followed by a target detection task, and subsequently by a probabilistic outcome (Figure 1; more task detail in supplement).

**Reinforcement-learning models**

Reinforcement-learning modelling proceeded in the same way as described in a prior paper (10). Briefly, we built seven parameterized reinforcement-learning models to fit to the behaviour of the subjects. All models were adapted Rescorla Wagner models. We use the term ‘Standard’ to denote the 6 parameter winning model from Guitart-Masip, et al. (2012) and either add or subtract parameters to test model fits for seven separate models (See Table 1 for a parameter specification summary).

**Learning models:** All the models assigned a probability to each action $a_t$ on trial $t$ based on an action weight and the current stimulus. The action weights were constructed according to a simple Rescorla–Wagner-like update equation with a learning rate. Reinforcements were coded as +1 for a reward, -1 for a punishment and 0 for no feedback. A sensitivity parameter determined the effective size of reinforcements for a subject. For the majority of models the sensitivity parameter could take on different values for the reward and punishment trials. For one model (‘Standard + 2 Approach-Avoid - 1 Sense’) there was only one sensitivity parameter.
Mkrtchian et al.

per subject, thus assuming that failure to obtain a reward was as aversive as obtaining a punishment. The initial value for the go action was set to zero and the action weight was modified to include a static general action bias parameter which denoted overall go tendency (with the exception of one model ‘Standard - action bias’, in which this was not included). The Pavlovian approach-avoid bias parameter (excluded for one model ‘Standard - Approach-Avoid’) inhibited the tendency to go in proportion to the negative value of the punishment stimulus, while it similarly promoted the tendency to go in proportion to the positive value of the reward stimulus. For the model with two Approach-Avoid parameters (‘Standard + 2 Approach-Avoid’), there were two parameters, updated separately for rewarded and punished trials. For the models with two learning rates (‘Standard + 2 Approach-Avoid + 2 learning rates’ / ‘Standard + 2 learning rates’), there were separate learning rates for rewarded and punished trials. In sum, for a given action \(a = \text{go} / \text{no-go}\), stimulus \(s = \text{GW} / \text{GA} / \text{NGW} / \text{NGA}\), reinforcement \(r = +1 / -1 / 0\) on each trial \(t\):

\[
Q_t(a,t,s,t) = Q_{t-1}(a,t,s,t) + \text{LearningRate} \cdot \left( (\text{Sensitivity} \cdot r_t) - Q_{t-1}(a,t,s,t) \right)
\]

\[
\text{Value}_t(s,t) = \text{Value}_{t-1}(s,t) + \text{LearningRate} \cdot \left( (\text{Sensitivity} \cdot r_t) - \text{Value}_{t-1}(s,t) \right)
\]

\[
\text{ActionWeight}_t(a,s) = \begin{cases} 
Q_t(a,s) + \text{ActionBias} + \text{AppAvoBias} \cdot \text{Value}_t(s) & a = \text{go} \\
Q_t(a,s) & a = \text{no-go}
\end{cases}
\]

**Observation model:** For action selection, the probability of each action was passed through a squashed softmax function with the addition of an irreducible lapse parameter (referred to as ‘noise’ in earlier papers, but renamed lapse here to avoid confusion with temperature noise parameters), which was free to vary between 0 and 1.
Parameter Estimation

We used an hierarchical Type II ML expectation–maximization (EM) procedure to fit the parameters across all subjects and conditions. These procedures are identical to those used by Huys et al 2011(12). Briefly, the top level of the hierarchical model specified distributions over the parameters for the subjects (see below). At each iteration, the current top-level distributions were used as a prior for a Laplace approximation to the intermediate-level posterior distribution of the parameters for each subject (the E-phase). These intermediate-level distributions were then used to determine the next iteration of the top-level distributions (the M-phase). The algorithm was initialized with maximum likelihood values of all the parameters for the subjects; the Laplace approximation was based on the use of fminunc in MATLAB, using multiple random initial values at each iteration of optimization to help avoid local minima. Four different population distributions were tested (see Figure 3):

1) Four distributions: one for anxious individuals under threat, one for controls under threat, one for anxious individuals under safe, one for controls under safe. This is the most relaxed procedure and serves to pull all parameters apart.

2) Two distributions: one distribution for threat and one distribution for safe. This fitting procedure was blind to the existence of group.

3) A single distribution for all participants and conditions (i.e. each participant was included twice within the distribution; once for the safe, and once for threat conditions). This fitting procedure was blind to the existence of both group and threat condition, and serves to pull all parameters closer together.
4) Two distributions: one distribution for anxious individuals and one distribution for controls. This fitting procedure was blind to the existence of induced anxiety.

The fit of each model and distribution was compared using the integrated BIC (iBIC). The iBIC is the integral of the likelihood function over the individual parameters (for details, see [12]). Small iBIC values indicate a model that fits the data better after penalizing for the number of parameters. The parameter fitting procedure results in one iBIC per distribution. These are then summed together to provide a single iBIC to enable model comparison across distributions. The lowest overall iBIC denotes the ‘winning’ model and distribution combination (an approximate Bayes Factor of the comparison of iBIC scores can be calculated using $\exp(\Delta iBIC/2)$). Note that fitting the parameters of the winning model using a different, hierarchical Bayesian, approach recovered similar parameters (see supplement). During fitting, parameters are constrained to within meaningful ranges (see [12]). Exponential transforms are applied to ensure Approach-avoid and sensitivity parameters do not go below zero and sigmoid transform to ensure learning rate and action bias parameters are constrained between 0 and 1. These transformations mean that parameters are not normally distributed (and so fall outside conventional box and whisker plot in Figure 3f).

The parameters recovered from the winning model were then compared across groups and conditions using two-tailed permutation tests implemented R coin (http://tiny.cc/o6brdy IndependenceTest, oneway_test). The recovered p-values are comparable to those derived from standard t-tests, but do not require the assumption of normality (critical given the possibility of multimodal distributions recovered from the model fitting procedure).
Results

**Basic analysis of symptoms and behaviour**

As expected, the mood and anxiety group reported significantly higher symptoms of trait anxiety (F(1,96)=69.6, η\(_p^2\)=0.4, p<0.001; **Figure 2a**) and depressive symptoms (F(1,90)=50, η\(_p^2\)=0.4, p<0.001) relative to controls (for a breakdown by sub diagnosis see **table S1**; note that as is commonly observed these measures are highly correlated across the whole sample; r(96)=0.755, p<0.001). Participants retrospectively reported feeling greater anxiety during the stress manipulation relative to the matched safe condition (F(1,99)=166, η\(_p^2\)=0.6, p<0.001; **Figure 2b**), which was similar between groups (main effect of group: F(1,99)=2.0, η\(_p^2\)=0.02, p=0.16; group x condition interaction F(1,99)=0.007, η\(_p^2\)<0.001, p=0.9).

Analysis of overall performance accuracy revealed a main effect of action (F(1,99)=90, η\(_p^2\)=0.5, p<0.001), qualified by an action (go/no-go)-by-valence (reward/punishment) interaction (F(1,99)=94, η\(_p^2\)=0.5, p<0.001; **Figure 2c**). As expected, this was driven by worse relative performance in the conditions where Pavlovian biases had to be overcome in order to make the appropriate response (i.e. a loss-driven avoidance bias in GA; and a win-driven approach bias in NGW) as well as an overall bias towards making go responses (which means that NoGo performance is worse overall likely due to subjects’ prior belief that they should respond). There was a main effect of group (F(1,110)=15, η\(_p^2\)=0.1, p<0.001) driven by worse overall accuracy in the disordered group, but no other interactions with group or condition (all p>0.5). However, as apparent in **Figure 2d**, learning follows a complex time-course which differs by condition (and by individual). We therefore turned to a computational model-based analysis to integrate the results across conditions, and thereby examine these differences at a fine scale. In the supplement, we exploit this clearer understanding to show model-agnostic signatures of the model-based effects.
Reinforcement-learning model selection and validation

We fitted reinforcement-learning models to trial-by-trial choice behaviour using an hierarchical Type II maximum likelihood expectation–maximization approach (12). The most parsimonious model (‘Standard + 2 Approach-Avoid + 2 Learning Rates’; Table 1; Figure 3e; methods) is an adapted Rescorla-Wagner model (21) identical to the winning model in prior studies of healthy individuals (8, 10), with the exception that there are separate Pavlovian approach, avoid and learning rate parameters for the cases of rewards and punishments. In other words, this model included an approach bias parameter, an avoidance bias parameter, and accommodated separate speeds of learning about rewards and punishments.

The hierarchical model fitting procedure requires the specification of population level priors. This raises an important conceptual question when it comes to considering multiple groups. Should we consider disordered and healthy groups as being sampled from the same or different populations? We answered this question through the adoption of a population-level model comparison approach. We compared fits for models ranging from four separate prior distributions for each group and stress condition (Figure 3a) to a single distribution for all subjects and conditions (Figure 3c). The best fit for our winning model was achieved by fitting a single population distribution (Figure 3c), implying that we did not obtain sufficient evidence to suggest that anxious and healthy individuals were sampled from different populations. Box plots and means of the posterior parameter distribution across subjects (under the Type II empirical prior) are shown in Figure 3f; that all subjects share the same prior implies that the recovered parameters will be drawn closer together.

We next ran a posterior predictive model with parameters set to those from the winning model (i.e. having a computer make decisions as if it was each individual subject). Average parameters recovered from simulated data were close to those that were originally observed.
Pathological symptoms are associated with increased reliance on avoidance bias, especially under stress.

We finally performed permutation tests on the posterior parameters to assess the effects of group and threat condition. These revealed an increased reliance on the avoidance bias parameter in the disordered group (effect of group averaged across threat and safe): $p_{\text{(permutation)}}=0.042$ (Figure 4c) and a significantly greater increase in the avoidance parameter under threat vs safe conditions in the disordered group relative to controls ($p_{\text{(permutation)}}=0.015$; Figure 4c) driven by a significantly greater avoidance in the disordered group relative to controls under threat ($p_{\text{(permutation)}}=0.006$), but not safe ($p_{\text{(permutation)}}=0.17$) conditions (there was no significant condition effect within groups (disordered $p_{\text{(permutation)}}=0.36$; control $p_{\text{(permutation)}}=0.28$).

Discussion

Anxious individuals show strong avoidance behaviour that can be debilitating and self-perpetuating (1, 2). Here, using a computational approach, we provide evidence that mood and anxiety disorders are associated with increased reliance on an avoidance bias (a Pavlovian bias to withhold responding in the face of punishments) during reinforcement-learning. Moreover, consistent with the diathesis-stress hypothesis, this effect was exacerbated under stressful conditions in the disordered group only.

We provide a potential computational mechanism for this effect. We show that avoidance behaviour – which is currently measured by retrospective self-report - can emerge at the level of stimulus-action associations. Specifically, individuals with mood and anxiety disorders may show avoidance in the face of threats because they inhibit their action tendencies when faced with a perceived negative outcome. This is consistent with prior work demonstrating increased
behavioural inhibition under stress (13, 14), in pathological anxiety (15) and in high (non pathological) trait anxiety (22) (although see (23)). Over time, however, individuals may be ultimately able to learn to overcome this bias (i.e. promote instrumental override of Pavlovian bias parameters) if they are given the opportunity to experience outcomes (i.e., NGW go probability is lower at the end than GW here). However, in the real world, avoidance means that, by definition, predicted outcomes are rarely experienced and challenged, there is little opportunity to learn, and a persistent miscalibration can emerge.

The growing field of computational psychiatry (18) seeks to use theory-driven approaches to explain psychiatric phenomena. Testable theories are a prerequisite to a clear mechanistic understanding: here, we have outlined a precise and formalised computational theory about how avoidance emerges in anxiety under stress. This approach has at least two further advantages. Firstly, it allows us to reduce a highly dimensional dataset (here, choices over time) into small number of parameters that respect the temporal variability of the data (unlike responses averaged over time). Secondly, we can directly integrate this model into biophysically plausible models of underlying neural activity (24). Indeed, performance of this task in healthy individuals has been linked neurocognitively to striatal and midbrain regions associated with network models of action (9, 10) as well as dopaminergic modulation of this circuitry (25). Striatal regions of this circuitry are also modulated by the threat of shock technique used here (26), providing a link between these substrates and stress. This computational approach therefore holds promise as a means of unifying complex psychiatric phenomena, such as avoidance, with their underlying neural circuitry.

Such a mechanistic link is critical if we wish to develop improved treatments. Without mechanistic understanding, treatment development has to be targeted at downstream symptoms – e.g. self-reported avoidance. The problem with this approach can be illustrated by the symptom of cough (27). Lung cancer, allergies, bronchitis or tuberculosis all result in a cough
Mkrtchian et al.

through fundamentally different mechanisms, but the treatment for one will be ineffective for the others (and indeed may even cause harm through side effects). Targeting clearly defined mechanisms, not symptoms, should ultimately improve the effectiveness of interventions. For example, extensive work in the development of psychological interventions for mood and anxiety disorders has suggested that exposure therapy should be paired with behavioural training to overcome avoidance in order to be effective(1), but the mechanism is unclear. The present findings suggest that this may be because such training encourages an instrumental override of Pavlovian bias during action selection. One avenue for future exploration, therefore, is whether training to overcome bias on GA trials on tasks like the present could promote instrumental override (cf.(28) but also(29)). If proven effective, such speculative task-based interventions (completed via smartphones, for example) could have enormous potential value for public health.

Limitations

While our model may provide a mechanism by which avoidance behaviour occurs in anxiety and depression, it does not provide a means of disentangling its relationship with specific constructs under the broad category of distress(30). Indeed, symptoms of anxiety and depression are highly co-morbid (mixed MDD and GAD is the most common diagnosis in our sample and our self-report measures of anxiety and depression are highly correlated), so future work is needed to delineate how, if at all, avoidance processes map separately onto feelings of anxiety or depression. In this study we did not find a reliable relationship between the avoidance parameter and self-reported anxiety symptoms using a dimensional approach (see supplement). One potential explanation is that our self-report measures are not optimal for capturing the symptoms measured by our task. Self-reported avoidance behaviour might, for instance, show a stronger relationship with task performance.
It is also worth highlighting that there is a difference between ‘passive avoidance’ and ‘active avoidance’, the latter being where an individual performs an action to avoid harm (i.e. GA). There are clear individual differences in avoidance learning strategies(31), so reliance on active vs passive avoidance may differ across subgroups of anxious individuals. For instance, active avoidance may be especially prominent in PTSD(32), so an interesting question for future work is whether PTSD may be associated with corresponding improved GA performance and hence improved task performance.

Another important limitation is that, while it is possible to see evidence of the influence of the avoidance parameter when performance averages are divided into separate time bins (see supplement), our non-modelling analysis is inherently less sensitive to the avoidance effects because focusing on means reduces our sensitivity to detect effects that evolve over trials.

Finally, it should be noted that we use a Bayesian framework for evaluating model fit and then use a frequentist approach to compare output parameters. This approach asks whether parameters, which were fitted under a single distribution, actually come from separate distributions. This is highly conservative and will require large effects in order for differences to be detected. A better approach would be to test the effect of varying the population priors at the parameter level. In light of the present data, we would predict that avoidance bias would be best fit using multiple distributions, while all other parameters will be best fit under a single distribution. This would enable inference about group differences in parameters to be fully confined within the model comparison framework. We are actively developing tools that will enable this approach in the future. Relatedly, this is the first study using this task to report results for a model that includes separate avoidance and approach parameters. To the best of our knowledge this model has not previously been reported, and it is possible that it would also offer the most parsimonious account of other samples. However, it is also plausible that the
addition of an extra parameter is only warranted in a sample in which this captures additional variance (as is the case here, being the only parameter that differs across groups).

References

1. Krypotos A-M, Effting M, Kindt M, Beckers T (2015): Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*. 9:189.
2. LeDoux JE, Moscarello J, Sears R, Campese V (2016): The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. *Molecular psychiatry*.
3. Baum M (1970): Extinction of avoidance responding through response prevention (flooding). *Psychological Bulletin*. 74:276.
4. Monroe SM, Simons AD (1991): Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. *Psychological bulletin*. 110:406.
5. Davis M, Walker DL, Miles L, Grillon C (2010): Phasic vs Sustained Fear in Rats and Humans: Role of the Extended Amygdala in Fear vs Anxiety. *Neuropsychopharmacology*. 35:105-135.
6. Robinson OJ, VytaI K, Cornwell BR, Grillon C (2013): THE IMPACT OF ANXIETY UPON COGNITION: PERSPECTIVES FROM HUMAN THREAT OF SHOCK STUDIES. *Frontiers in Human Neuroscience*. 7.
7. Schmitz A, Grillon C (2012): Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the NPU-threat test). *Nat Protocols*. 7:527-532.
8. Guitart-Masip M, Economides M, Huys QJ, Frank MJ, Chowdhury R, Duzel E, et al. (2014): Differential, but not opponent, effects of L-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology*. 231:955-966.
Mkrtchian et al.

9. Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJ, Dayan P, Dolan RJ, et al. (2011): Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *The Journal of Neuroscience*. 31:7867-7875.

10. Guitart-Masip M, Huys QJ, Fuentemilla L, Dayan P, Duzel E, Dolan RJ (2012): Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage*. 62:154-166.

11. Hershberger WA (1986): An approach through the looking-glass. *Animal Learning & Behavior*. 14:443-451.

12. Huys QJ, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, et al. (2011): Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol*. 7:e1002028.

13. Grillon C, Robinson OJ, Krimsky M, O’Connell K, Alvarez G, Ernst M (2016): Anxiety-Mediated Facilitation of Behavioral Inhibition: Threat Processing and Defensive Reactivity During a Go/No-Go Task.

14. Robinson OJ, Krimpsky M, Grillon C (2013): The impact of anxiety on response inhibition. *Front Hum Neurosci*.

15. Grillon C, Robinson O, O’Connell K, Davis A, Alvarez G, Pine D, et al. (2016): Clinical anxiety promotes excessive response inhibition. *Psychological Medicine*. 1-11.

16. Henriques JB, Glowacki JM, Davidson RJ (1994): Reward fails to alter response bias in depression. *Journal of abnormal psychology*. 103:460.

17. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. (2015): Human-level control through deep reinforcement learning. *Nature*. 518:529-533.

18. Huys QJM, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 19:404-413.
19. Sheehan D, Lecrubier Y, Sheehan KH, Janavs J, Weiller E, Keskiner A, et al. (1997): The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. European Psychiatry. 12:232-241.

20. Crockett MJ, Clark L, Robbins TW (2009): Reconciling the Role of Serotonin in Behavioral Inhibition and Aversion: Acute Tryptophan Depletion Abolishes Punishment-Induced Inhibition in Humans. The Journal of Neuroscience. 29:11993-11999.

21. Rescorla RA, Wagner AR (1972): A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory. 2:64-99.

22. Bach DR (2015): Anxiety-Like Behavioural Inhibition Is Normative under Environmental Threat-Reward Correlations. PLOS Computational Biology. 11:e1004646.

23. Forster S, Elizalde AON, Castle E, Bishop SJ (2013): Unraveling the anxious mind: Anxiety, worry, and frontal engagement in sustained attention versus off-task processing. Cerebral Cortex.bht248.

24. Wiecki TV, Frank MJ (2013): A computational model of inhibitory control in frontal cortex and basal ganglia. Psychological Review. 120:329.

25. Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ (2012): Action controls dopaminergic enhancement of reward representations. Proceedings of the National Academy of Sciences. 109:7511-7516.

26. Robinson OJ, Overstreet C, Charney DS, Vytal K, Grillon C (2013): Stress increases aversive prediction-error signal in the ventral striatum. Proc Natl Acad Sci U S A.

27. Roiser J (2015): What has neuroscience ever done for us. The Psychologist. 28:284-287.

28. Amir N, Kuckertz JM, Najmi S (2013): The Effect of Modifying Automatic Action Tendencies on Overt Avoidance Behaviors. Emotion (Washington, DC). 13:478-484.
Figure Legends and Table

**Figure 1 Experimental paradigm:** The trial sequence for each trial-type condition under threat (red) and safe (blue) conditions. There were equal numbers of Go to Win (GW), Go to Avoid (GA), No-Go to Win (NGW) and No-Go to Avoid (NGA) trials within each safe and threat block, and these were randomly ordered within each block (note that safe sequence proceeds in the same way as threat but is curtailed here for brevity). The prepotent Pavlovian bias to a win is a go response (approach) and the prepotent Pavlovian response to a loss is no-go (avoid); hence in GW and NGA, the bias and task instructions are aligned; but in GA and NGW, participants have to learn to overcome their avoidance and approach biases respectively. The safe and threat blocks were presented in alternating order, counterbalanced across participants. A different set of fractal cues was used for the safe and threat blocks, counterbalanced across participants. At feedback, a face (happy +10 points, fear -10 points) was shown 80% of the time, and no points (i.e. a yellow bar – not shown in the figure) was shown 20% of the time.
Figure 2: Self-report anxiety and task performance. Between groups, a) our disordered sample reported significantly higher trait anxiety scores (data missing for two participants in the healthy and one in the disordered group), while b) the whole sample reported increased (induced) anxiety, rated retrospectively, under threat relative to safe conditions (violin plots; each point represents a subject, background shading represents estimated distribution). c) Collapsed mean accuracy differs as a function of trial type, but this ignores that d) performance on the task changed over time such that the probability of making a response (P(go); as distinct from accuracy in c) differed as a function of trial type, condition, group and time (shading represents standard error of the mean). (HC=healthy control=green; ANX=mood and anxiety group=grey; Saf =safe; Thr= threat; Avo=Avoid)

Figure 3: Model Fitting and Comparison Four different population distributions were tested separated by a) group and threat condition (4 distributions); b) by threat condition alone (2 distributions); c) blind to group and threat condition (1 distribution); and d) by group alone (2 distributions). Comparison of models and distributions using integrated Bayesian Information Criteria (iBIC) scores (colours match distributions throughout figure) revealed a winning model of ‘Standard + 2 Approach-Avoid + 2 Learning Rates’, fit across a single prior distribution (inset zoomed in on the distribution comparison for this model). Box and whisker plots of the recovered parameters from the wining model/distribution are presented in f) separated by group and condition (red triangles denote means, lines denote medians; based on individual parameter estimates). Log scales are used for the sensitivity and approach-avoidance parameters to aid visualisation of these exponentially transformed parameters (HC=healthy control; ANX = mood and anxiety group; Rew = reward; Pun = punishment; Stand= standard; Ap-Av = approach avoid; Sense = sensitivity; LR=learning rate; Avoid=Avoidance Bias; Approach = Approach Bias)
Mkrtchian et al.

**Figure 4: Posterior Predictive Model** Running the estimated parameters for each subject through a posterior predictive model recovered both a) average go probabilities for each trial type (sensitivity plots: each marker represents one subject under one condition so there are twice as many markers as subjects), and b) group-averaged trial-by-trial performance (compare to real data in Figure 2c). Comparing parameters across group and condition revealed c) a significantly higher avoidance bias parameter in pathological anxiety across conditions as well as greater threat-potentiated avoidance in pathological anxiety (error bars represent standard error of the mean; HC=healthy control=green; ANX=mood and anxiety group=grey; Saf =safe; Thr= threat; Avo=Avoid)
Table 1: Model specification *(NP = number of parameters)*

| Model Name                  | NP | Parameters                  | Parameters                  |
|-----------------------------|----|-----------------------------|-----------------------------|
| Standard - Action Bias      | 5  | Reward sensitivity          | Punishment sensitivity      |
| Standard - Approach-Avoid   | 5  | Reward sensitivity          | Punishment sensitivity      |
| Standard + 2 Approach-Avoid | 6  | Sensitivity                 | Learning rate               |
| Standard + 2 Approach-Avoid | 7  | Reward sensitivity          | Punishment sensitivity      |
| Standard + 2 Learning Rates | 7  | Reward Learning rate        | Punishment Learning rate    |
| Standard + 2 Approach-Avoid | 8  | Reward sensitivity          | Punishment sensitivity      |
