JNeurosci
THE JOURNAL OF NEUROSCIENCE

*Research Articles: Behavioral/Cognitive*

# Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals

**Payam Piray[1], Verena Ly[2], Karin Roelofs[1], Roshan Cools[1] and Ivan Toni[1]**

[1]*Donders Institute, Radboud University, the Netherlands*
[2]*Department of Clinical Psychology; Leiden Institute for Brain and Cognition, Leiden University, the Netherlands*

https://doi.org/10.1523/JNEUROSCI.1394-18.2018

Received: 1 June 2018

Revised: 19 November 2018

Accepted: 11 December 2018

Published: 17 December 2018

**Author contributions:** P.P., V.L., K.R., R.C., and I.T. designed research; P.P. and V.L. performed research; P.P. contributed unpublished reagents/analytic tools; P.P. analyzed data; P.P., R.C., and I.T. wrote the paper; V.L. and K.R. edited the paper.

**Conflict of Interest:** The authors declare no competing financial interests.

The authors would like to thank Nathaniel Daw for helpful advice. K.R. was supported by a starting grant from the European Research Council (ERC_StG2012_313749) and a VICI grant (#453-12-001) from the Netherlands Organization for Scientific Research (NWO). R.C. was supported by a James McDonnell Scholar Award (grant number 220020328).

Corresponding author, current address: Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540, Email: ppiray@princeton.edu

**Cite as:** J. Neurosci 2018; 10.1523/JNEUROSCI.1394-18.2018

# Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals

Payam Piray[1,*], Verena Ly[2], Karin Roelofs[1], Roshan Cools[1,+] and Ivan Toni[1,+]

[1] Donders Institute, Radboud University, the Netherlands

[2] Department of Clinical Psychology; Leiden Institute for Brain and Cognition, Leiden University, the Netherlands

*Corresponding author, current address: Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540, Email: ppiray@princeton.edu

[+] These authors contributed equally to this work.

Conflict of Interest: the authors declare no conflict of interest.

Acknowledgments: The authors would like to thank Nathaniel Daw for helpful advice. K.R. was supported by a starting grant from the European Research Council (ERC_StG2012_313749) and a VICI grant (#453-12-001) from the Netherlands Organization for Scientific Research (NWO). R.C. was supported by a James McDonnell Scholar Award (grant number 220020328).

## Abstract

Learning and decision-making are modulated by socio-emotional processing and such modulation is implicated in clinically-relevant personality traits of social anxiety. The present study elucidates the computational and neural mechanisms by which emotionally aversive cues disrupt learning in socially anxious human individuals. Healthy volunteers with low or high trait social anxiety performed a reversal learning task requiring learning actions in response to angry or happy face cues. Choice data were best captured by a computational model in which learning rate was adjusted according to the history of surprises. High trait socially anxious individuals employed a less dynamic strategy for adjusting their learning rate in trials started with angry face cues and unlike the low social anxiety group, their dorsal anterior cingulate cortex (dACC) activity did not covary with the learning rate. Our results demonstrate that trait social anxiety is accompanied by disruption of optimal learning and dACC activity in threatening situations.


## Significance statement

Social anxiety is known to influence a broad range of cognitive functions. This study tests whether and how social anxiety affects human value-based learning as a function of uncertainty in the learning environment. The findings indicate that, in a threatening context evoked by an angry face, socially anxious individuals fail to benefit from a stable learning environment with highly predictable stimulus-response-outcome associations. Under those circumstances, socially anxious individuals failed to use their dorsal anterior cingulate cortex, a region known to adjust learning rate to environmental uncertainty. These findings open the way to modify neurobiological mechanisms of maladaptive learning in anxiety and depressive disorders.

## Introduction

Economics, psychology, and neuroscience have often assumed that emotions compete with reason during decision-making (Cohen, 2005; Kahneman, 2011). Recent theories challenge this notion, suggesting that in fact emotions are deeply embedded within decision-making computations (Phelps et al., 2014; Lerner et al., 2015). For instance, recent work has shown that trait-anxiety and stress sensitivity influence learning rate, a quantity reflecting the rate at which decision values are updated by new information (Browning et al., 2015; de Berker et al., 2016). These observations are in line with older descriptive studies suggesting that emotions modulate cognitive flexibility (Dreisbach and Goschke, 2004; van Steenbergen et al., 2010). Although recent studies have revealed neural correlates of dynamic learning rate (Behrens et al., 2007, 2008; Li et al., 2011), particularly in the dACC (Behrens et al., 2007, 2008), the computational and neural mechanisms by which emotional cues and emotion-related traits modulate learning rate are unknown.

Psychological models of conditioning, such as Rescorla-Wagner (Rescorla et al., 1972), suggest that animals learn by computing prediction errors. Such errors are positive when an outcome (reward or punishment) is better than expected and negative when the outcome is worse than expected. According to these models, animals learn by updating their expectation in proportion to the prediction error multiplied by a learning rate. In Rescorla-Wagner models, the learning rate is assumed to be a constant parameter between zero and one. Models of associative learning, such as Pearce-Hall (Pearce and Hall, 1980), however, suggest that animals learn stimulus-outcome associations by tracking associability, a quantity reflecting the extent to which each cue has previously been accompanied by surprise (unsigned prediction errors). This quantity guides animals' attention towards cues with large associability. According to these models, the associability signal gates the amount of future learning about the cue on the basis of whether it has been a reliable or poor

predictor of reinforcement in the past. Bayesian or temporal difference models proposed for learning in uncertain environments essentially combine the key features of both accounts, in which error-driven learning depends on a dynamic learning rate closely resembling the notion of associability (Behrens et al., 2007, 2008; Li et al., 2011; Iglesias et al., 2013). These models indicate that when the environment is highly surprising, the learning rate should be higher allowing expectations to get updated quickly. This causal inference about changes in the environment might be particularly disrupted in anxiety and depressive disorders, which are associated with self-blame symptoms. As noted by Beck (Beck, 1967), self-blame in a depressed patient "expresses a patient's notion of causality". In other words, in an uncertain environment, these patients might attribute negative outcomes to their own actions instead of the stochasticity of the environment and change their decisions frequently. This view is consistent with theories suggesting that emotion-related traits modulate associability tracking in uncertain environments (Paulus and Yu, 2012; Mason et al., 2017). Relatedly, a recent study has reported that trait anxiety is negatively correlated with the ability to adjust learning rate in uncertain environment (Browning et al., 2015). However, the neural mechanisms by which learning rate is related to trait anxiety are still unknown. Furthermore, it is not clear whether emotionally aversive cues in the environment mediate such relation.

Here, we combine functional neuroimaging and computational modeling to investigate whether and how emotions modulate learning rate and whether those modulations depend on individual variation in the personality trait of social anxiety. A hybrid computational model was considered, in which error-driven learning depends on a learning rate containing both dynamic-, similar to Pearce-Hall, and constant-, similar to Rescorla-Wagner, components. Model-based analysis of task-related fMRI data was conducted to investigate the neural correlates of dynamic learning rate in the dACC, a region previously shown to encode dynamic learning rate in uncertain environments (Behrens et al., 2007,

2008). We hypothesized that the dynamic adjustment of learning rate and its neural correlates depend on emotional state and trait social anxiety.

# Methods

## Participants

Forty-five female volunteers gave written informed consent approved by the local ethical committee ("Comissie Mensgebonden Onderzoek" Arnhem-Nijmegen) and participated in the study. Only women have been recruited to have a relatively homogeneous sample in terms of emotional reactivity (Koch et al., 2007; Domes et al., 2010). Exclusion criteria were claustrophobia, neurological, cardiovascular or psychiatric disorders, regular use of medication or psychotropic drugs, heavy smoking and metal parts in the body. Participants were selected from an online pool of students based on their scores on the Liebowitz social anxiety scale (Liebowitz, 1987). Thus, participants were recruited to have either low (not greater than 13, n=23) or high scores (not smaller than 25, n=22) on this test. One participant did not finish the experiment due to headache (from the high score group). Data from all other 44 participants were analyzed (all right-handed, mean age of 20.7). We used data from a previously published study (Ly et al., 2014) focused on the association between emotional biasing of go/no-go responding and individual differences in social avoidance. Unlike the current study, Ly et al. (2014) did not consider any form of learning and only focused on behavioral inhibition.

## Probabilistic reversal learning task

Each participant completed 480 trials of a probabilistic learning task in the scanner. Each trial started with a face cue (happy or angry) presented on a color frame indicating the type of outcome valence (reward or punishment) at the end of the trial. Thus, there were

four trial-types in a 2x2 factorial design with factors emotion (happy or angry) and valence (reward or punishment). There were 120 trials per trial-type. Participants were instructed that the combination of emotional content of the face cue and color frame distinguished the four trial-types and that they had to learn the optimal response for each of the four cue-types separately. The response-outcome contingency was probabilistic and independent for each trial-type. The response-outcome contingency was reversed several times for each trial type, resulting in different degree of volatility in the course of experiment, while remaining counterbalanced across trial types. Specifically, each participant completed three sessions, with a 1-min break in between the sessions. Each session consisted of 160 trials, with 40 trials per trial-type. For each trial-type within a session, the probability of a positive outcome given a go-response could take one of the following combinations in two consecutive blocks: (i) 0.5, 0.2, 0.5, 0.2; (ii) 0.5, 0.2, 0.5, 0.8; (iii) 0.5, 0.8, 0.5, 0.8, where each session was associated with one of these combinations. The blocks with probability of 0.5 were short blocks with average length of 5 trials, and other blocks were long blocks with average length of 15 trials.

Emotional stimuli were adult Caucasian faces from 36 models (18 men) taken from several databases (Ekman and Friesen, 1976; Matsumoto and Ekman, 1988; Lundqvist et al., 1998; Martinez, and Benavente, 1998). Model faces were trimmed to exclude influence from hair and non-facial contours (van Peer et al., 2007; Roelofs et al., 2009). Model identity was counterbalanced, such that the model occurred equally often for each trial-type. The color frame (yellow or grey) indicating the possibility of reward or punishment was also counterbalanced across participants. On each trial, one of the face cues was presented centrally. Participants were then allowed to make a response 100 ms after cue onset, where they were required to make either a go- or a no-go-response within 1000 ms. If no response was made within 1000 ms, then a no-go-response was recorded. After a response-outcome delay of maximally 2000 ms (depending on the response time), the outcome was presented

for 1000 ms (+10 cents for reward, -10 cents for punishment, and 0 cents for omitted reward or avoided punishment). The inter-trial interval was jittered (2500 to 4500 ms).

The relatively long time window for responding (1000 ms) ensured that no-go responses are not due to failure in making a go response. To illustrate this point, we tested each participant response-time separately for go-responses in every trial-type. This test revealed that for all participants and all trial types, response-time are significantly lower than 1000 ms window (t-test, all *P*-values<$10^{-10}$).

## Computational models

In this section, we describe the computational learning models compared in this study. A common choice model was then used in combination with each of these learning models to predict the probability of choices, which will be presented later.

All learning models track expected value $x_t$ on trial $t$ of each stimulus and action pair. Thus, if $s_t$ is the stimulus presented on trial $t$, $c_t$ is the choice taken and $o_t$ is the received outcome, all models compute a prediction error signal and update the corresponding expected value:

$$\delta_t = o_t - x_t(s_t, c_t)$$

$$x_{t+1}(s_t, c_t) = x_t(s_t, c_t) + \alpha_t \delta_t$$

where $\delta_t$ is the prediction error on trial $t$ and $\alpha_t$ is the learning rate representing the degree to which the prediction error influences the current expected value. The learning models are different in how they conceptualize the learning rate.

**M1. Rescorla-Wagner model.** This model (Rescorla et al., 1972) is the simplest model among the tested models, containing only one free learning parameter as constant learning

rate, $\kappa$, bounded in the unit range, [0 1]. Therefore, for this model, $\alpha_t$ is equal to $\kappa$ on all trials.

**M2. Hybrid model.** This model and its variant (M4) are the main models of interest in this study. The hybrid model quantifies associability, $A_t$, and constructs the learning rate accordingly in two steps. First, it constructs $K_t$:

$$K_t = wA_t + (1 - w)$$

where w is the weight parameter constrained to lie in the unit range. Therefore, $K_t$ is a weighted combination of a constant- and a dynamic- component according to $w$. If $w = 0$, the dynamic component, $A_t$, has no influence on $K_t$ and therefore the learning rate is a constant. Conversely, if $w$=1, $K_t$ has no constant component and therefore it is fully dynamic. Note that, regardless of the value of $w$, the maximum possible value (i.e. the scale) of $K_t$ is 1. The learning rate is then defined as

$$\alpha_t = \kappa K_t$$

where $\kappa$ is another free parameter, which indicates the scale of learning rate. Thus, for any value of $\kappa$, the learning rate on every trial lies between 0 and $\kappa$.

In this model, the associability also gets updated. On every trial, two factors influence the associability update, similar to update rules in Bayesian dynamic models such as Kalman filter (e.g. see (Daw et al., 2006)). First, similar to the gain in the Bayesian models (e.g. Kalman gain), associability gradually reduces due to random diffusion:

$$A_t = \lambda A_t^{post}$$

Second, after observing the outcome of the trial, the associability gets updated according to the surprise (i.e. squared prediction error):

$$A_{t+1}^{post} = A_t \ + (1 - \lambda)\delta_t^2$$

Note that, on every trial, the learning rate, $\alpha_t$, depends on $A_t$, which itself depends on squared prediction errors from the past trials, but not the current one. Therefore, $\delta_t$ is not double counted in the value update.

Taken together, this learning model contains three free learning parameters, $\kappa$, $w$ and $\lambda$, which are all constrained to lie in the unit range. Moreover, since squared prediction errors in this task are between 0 and 1 (as outcomes are binary), associability will also always lie in the unit range. Consequently, learning rates will always be between 0 and 1 ensuring that expected values are well-defined for any set of parameters.

**M3. Reinforcement learning model of Li et al (2011).** This model also combines error-driven learning with an associability signal. The important difference between this model and M2 is that whereas in M2 the learning rate is a weighted combination of a dynamic and a constant component, M3 only contains a dynamic component. Also, the way that M3 quantifies surprise is slightly different compared with the M2 by updating associability according to the absolute value of previous prediction error (instead of squared value of prediction error).

$$A_t = (1 - \mu)A_{t-1} + \mu|\delta_{t-1}|$$

$$\alpha_t = \kappa A_t$$

where $\mu$ and $\kappa$ are free parameters (bounded in the unit range) determining the step-size for updating associability and the scale of learning rate, respectively.

**M4. Hybrid emotion-specific w model.** This model is identical to M2 except that it assumes two different weight parameters, $w_a$ and $w_h$, for angry and happy trials, respectively. Therefore, this model has one more free parameter compared with M2.

**M5. Hybrid emotion-specific *κ* model.** This model is also identical to M2 except that it assumes two different overall scale, *κ*, parameters for angry and happy trials.

**M6. Hybrid valence-specific w model.** This model is also identical to M2 except that it assumes two different weight, *w*, parameters for reward and punishment trials.

**Choice Model.** Each of the learning models was combined with a choice model to generate probabilistic predictions of choice data. Expected values were used to calculate the probability of actions, $a_1$ (go response) and $a_2$ (no-go response), according to a sigmoid (softmax) function:

$$p_t(a_1) = \frac{1}{1 + e^{-\beta(x_t(s_t,a_1) - x_t(s_t,a_2)) - b(s_t)}}$$

$$p_t(a_2) = 1 - p_t(a_1)$$

where $\beta$ is the decision noise parameter encoding the extent to which learned contingencies affect choice (constrained to be positive) and $b(s_t)$ is the bias towards $a_1$ due to the stimulus presented independent from learned values. The bias is defined based on three free parameters, representing bias due to the emotional content (happy or angry), $b_e$, bias due to the anticipated outcome valence (reward or punishment) cued by the stimulus, $b_v$, and bias due to the interaction of emotional content and outcome, $b_i$. No constraint was assumed for the three bias parameters. For example, a positive value of $b_e$ represents tendencies towards a go response for happy stimuli and for avoiding a go response for angry stimuli (regardless of the expected values). Similarly, a positive value of $b_v$ represents a tendency towards a go-response for rewarding stimuli regardless of the expected value of the go response. Critically, we also considered the possibility of an interaction effect in bias encoded by $b_i$. Therefore, the bias, $b(s_t)$, for the happy and rewarding stimulus is $b_e + b_v + b_i$, the bias for the angry and punishing stimulus is $-b_e - b_v + b_i$, the bias for the

221 happy and punishing stimulus is $b_e - b_v - b_i$ and the bias for the angry and rewarding
222 stimulus is $-b_e + b_v - b_i$.

223 ## Model fitting

224 We fitted parameters in the infinite real-space and transformed them to obtain *actual*
225 parameters fed to the models. Appropriate transform functions were used for this purpose:
226 the sigmoid function to transform parameters bounded in the unit range (the learning
227 parameters in all models) and the exponential function to transform the decision noise
228 parameter in the choice model. No transformation was needed for the bias parameters of
229 the choice model as they were not bounded.

230 Free parameters of each model were estimated in two stages. In the first stage, a set
231 of parameters, $\theta^n_{MAP}$, maximizing log-likelihood of data plus log-prior (maximum a posteriori,
232 MAP) was estimated for every participant separately ($n$ is the index of participant) similar to
233 our previous study (Piray et al., 2016). A wide Gaussian prior was assumed for all parameters
234 (with zero mean and a variance of 6.25). This initial variance is chosen to ensure that
235 the parameters could vary in a wide range with no substantial effect of prior. Specifically,
236 the log-effect of this prior is less than one chance-level choice (i.e log0.5) for any value of $w$
237 between 0.05 and 0.95. This is also the case for all other free parameters constrained in the
238 unit range. A non-linear derivative-based optimization algorithm (as implemented in the
239 fminunc routine in MATLAB, ©Mathwork) was used for fitting. To overcome bias of the
240 optimization algorithm to the initial point, the optimization was repeated multiple times and
241 the best set of parameters was selected.

242 In the second stage, a hierarchical fitting procedure was used to fit the models to
243 participants' choices. An expectation-maximization algorithm was used for optimizing
244 group– and individual– parameters in an iterative fashion, with Laplace approximation for
245 approximating the posterior distribution (Huys et al., 2011). This method estimates the

mean and the variance of parameters across all participants (group parameters) in the first step. In a subsequent step, that mean and variance is used to define a normal prior distribution of parameters and to estimate parameters of each individual participant using Laplace approximation. This procedure is then continued iteratively to reach convergence. Group parameters was initialized according to the mean and variance of the individual parameters, $\theta^n_{MAP}$, fitted in the first stage. This procedure regularizes individual fitted parameters according to group parameters, thereby decreases fitting noise and protects against outliers. The final estimated values for the group parameters, $\Theta$, were used to generate the regressors used in the fMRI analyses, as they are less biased by fitting noise. For details of the hierarchical fitting procedure, see Huys et al. (Huys et al., 2011).

All codes used for fitting are publically available online (https://github.com/payampiray/cbm_v0). The Gramm plotting tools (Morel, 2018) were used for visualization.

## Model selection

We employed a Bayesian model comparison approach to assess which model better captures participants' choices. This approach selects the most parsimonious model by quantifying model evidence, a metric which balances between model fits and complexity of the model (MacKay, 2003). Notably, this procedure penalizes complexity induced by both group and individual parameters using Laplace approximation and Bayesian information criterion (BIC), respectively. For each model fitted using the hierarchical fitting procedure, the log-model evidence (LME) is penalized for complexities at both individual and group levels, which can be quantified using Laplace approximation and Bayesian information criterion, respectively (Piray et al., 2014):

$$LME = \sum_n \log P(D^n|\theta^n) + \sum_n \log N(\theta^n|\Theta, \Sigma) + \frac{1}{2} dN \log 2\pi - \frac{1}{2}\sum_n \log|H_n| - d \log N$$

where $D^n$ is the set of choice data for the $n$th participant$\theta^n$, is the fitted individual parameters for $n$th participant, $\Theta$ and $\Sigma$ is the mean and variance for the group distribution, respectively, $d$ is number of free parameters of the model, $N$ is the number of participants and $|H_n|$ is the determinant of the Hessian matrix of the log-posterior function at $\theta^n$. The log-likelihood function is the predicted probability of choice data given the model and parameters defined as $\log p(D^n|\theta^n) = \sum_t \log p_t(c_t)$, where the sum is over all trials. Therefore, the first term on the right-hand side of the equation is how well the model predicts data. The sum of the next three terms together is the penalty due to individual parameters. The last term represents the penalty approximated for $2d$ (mean and variance together) group parameters as quantified using Bayesian information criterion.

## fMRI data acquisition and preprocessing

Whole-brain imaging was performed on a 3T MR scanner (Magnetom Trio Tim; Siemens Medical Systems) equipped with a 32-channel head coil using a multi-echo GRAPPA sequence (Poser et al., 2006) [repetition time (TR): 2.32 ms, echo times (TEs, 4): 9.0/19.3/30/40 ms, 38 axial oblique slices, ascending acquisition, distance factor: 17%, voxel size 3.3_3.3_2.5 mm, field of view (FoV): 211 mm; flip angle, 908]. At the end of the experimental session, high-resolution anatomical images were acquired using a magnetization prepared rapid gradient echo sequence (TR: 2300 ms, TE: 3.03 ms, 192 sagittal slices, voxel size 1.0_1.0_1.0 mm, FoV: 256 mm).

Given the multiecho GRAPPA MR sequence (Poser et al., 2006), the head motion parameters were estimated on the MR images with the shortest TE (9.0 ms), because these images are the least affected by BOLD signals. These motion-correction parameters, estimated using a least-squares approach with six rigid body transformation parameters (translations, rotations), were then applied to the four echo images collected for each excitation. After spatial realignment, the four echo images were combined into a single MR

volume using an optimized echo weighting method (Poser et al., 2006). Noise effects in data were removed using FMRIB's ICA-based Xnoiseifier tool (FIX), which uses independent component analysis (ICA) and classification techniques to identify noise components in data (Salimi-Khorshidi et al., 2014). Other preprocessing steps were carried out in SPM12. The T1-weighted image was spatially coregistered to the mean of the functional images. The fMRI time series were transformed and resampled at an isotropic voxel size of 2mm into the standard Montreal Neurological Institute (MNI) space using both linear and nonlinear transformation parameters as determined in a probabilistic generative model that combines image registration, tissue classification, and bias correction (i.e. unified segmentation and normalization) of the coregistered T1-weighted image (Ashburner and Friston, 2005). The normalized functional images were spatially smoothed using an isotropic 6mm full-width at half-maximum Gaussian kernel.

## Statistical analysis of imaging data

General linear model (GLM) was used to model effects at the single-subject level (first-level analysis). Four sets of four regressors, each containing one regressor per trial-type, were considered: one set was time-locked to the visual presentation of cues; one set was time-locked to the visual presentation of outcomes; one set was parametrically modulated by prediction error and time-locked to the presentation of the trial outcome; one set was parametrically modulated by dynamic learning rate and time-locked to the presentation of the trial outcome. Group parameters obtained through the hierarchical fitting procedure, $\Theta$, were used to generate these signals. Twelve motion regressors representing six motion parameters obtained from the brain-realignment procedure and their first derivative were also included.

Contrasts of interests were estimated at the subject-level. These contrast images were then used in a second-level GLM to make inference at the group level (t-test). The region-of-

interest analysis in the dorsal anterior cingulate was performed in anatomically defined mask of the rostral cingulate motor area, which has been shown to correlate with learning rate and has distinct connectional fingerprints. The rostral cingulate motor area mask was created based on a diffusion-parcellation atlas of human medial and ventral frontal cortex (thresholded at $p<0.25$) (Neubert et al., 2015).

## Results

Forty-four participants carried out a probabilistic learning task. Participants were selected from an online pool of students based on their scores on the Liebowitz social anxiety scale (Liebowitz, 1987). Thus, participants were recruited to have either low (not greater than 13) or high scores (not smaller than 25) on this test. Participants were accordingly divided into two groups with low (n=23, mean=8.26, SE=0.76) or high (n=21, mean=31.00, SE=1.37) social anxiety.

In the experiment (Figure 1), participants were presented with validated images of faces (happy or angry) and were asked to make either a go- or a no-go- response (i.e. press a button, or withhold a button press, respectively) for each of these facial cues in order to obtain monetary reward or avoid monetary punishment. There were 4 trial types: happy face – reward outcome trials, happy punishment, angry reward and angry punishment trials. Participants were also informed about outcome valence at the start of each trial by presenting the face image in a background color (yellow or white) indicating whether, at the end of a trial, a win outcome consisted of obtaining a reward or avoiding a punishment. Crucially, the response-outcome contingencies for the cues were probabilistic and manipulated independently, and reversed after a number of trials, varying between 5 and 15 trials, so that the experiment consisted of a number of blocks with varying trial length (Figure 1C). Within each block, the probability of a win was fixed. There were matched

numbers of action-outcome contingency reversals across trial types, with 120 trials in each of the four trial types (see Methods for details).

[Figure 1 about here]

Participants learned the task effectively: performance quantified as the number of correct decisions given the true underlying probability was significantly higher than chance across the group ($t(43)=14.68$, $p<0.001$). Importantly, participants responded to reversals. As Figure 2 shows, their performance was approximately at chance level immediately after reversals and improved slowly for all trial types and both type of responses. Note that, as Figure 2 shows, the effects of reversal learning on performance is not different between go and no-go responses as the slope of the two curves is not substantially different.

The emotional cues did not influence overall task performance ($t(43)=-0.37$, $p=0.71$), nor participants' bias towards go-responses ($t(43)=-1.39$, $p=0.17$). However, longer latencies of go responses following the presentation of angry face cues relative to happy face cues indicated that participants did process the emotional content of those cues ($t(43)= 3.72$, $p<0.001$). Latencies of go responses, however, did not vary as a function of social anxiety ($t(43)=0.68$, $p=0.5$).

[Figure 2 about here]

## Emotional cues modulate adaptive learning rate

We tested whether participants adjusted their learning rate dynamically according to the history of surprises. First, we considered a Rescorla-Wagner model in which expected value is updated by the product of prediction errors and a constant learning rate (model M1). We then focused on assessing the additional explanatory power of a class of an augmented hybrid Pearce-Hall Rescorla-Wagner models in which the learning rate depends on another variable, $K_t$, that combines the learning rate of Rescorla-Wagner with that of Pearce-Hall

model. The dynamic component of $K_t$ was adjusted according to the history of surprises (or sample variance equal to squared prediction error), similar to the Pearce-Hall associability rule.

Therefore, we built a model (model M2) in which $K_t$ is a weighted combination of a constant- and a dynamic- component according to a weight parameter, w. The weight parameter, w, indicates the degree to which this dynamic associability component influences on $K_t$ and thereby contributes to the learning rate. If w=0, the dynamic component has no influence on $K_t$ and therefore the learning rate is a constant. Conversely, if w=1, $K_t$ has no constant component and therefore the learning rate is fully dynamic.

On every trial, the product of $K_t$ with another free parameter, $\kappa$, indicates the learning rate on that trial, in which $\kappa$ indicates the overall scale of learning rate (also constrained to lie in the unit range). Thus, while w indicates the degree to which learning rate is changing over time, $\kappa$ determines the maximum of learning rate. In other words, on every trial, learning rate lies between zero and $\kappa$. In sum, this augmented hybrid model contains both a model with a constant learning rate (if w=0) for which the learning rate is always κ, and a model with a fully dynamic learning rate (if w=1) as special cases.

We used a choice model to generate probability of choice data according to action values derived for each model. Note that the choice model controlled value-independent biases in making or avoiding a go response due to the emotional or reinforcing content of the cues (see Methods for formal definition). We then used a hierarchical Bayesian estimation algorithm (Huys et al., 2011, 2012; Piray et al., 2014) to obtain parameters of the model given the data. This is an algorithm with the advantage that fits to individual subjects are constrained according to the group-level distribution. For each model, this procedure also calculates its evidence (Piray et al., 2014), a measure of goodness of fit of the model penalized by the complexity of the model (MacKay, 2003), which can be used for Bayesian

model comparison. This analysis revealed that the hybrid model explains data better than the simpler model with a constant learning rate (Table 1). As a control analysis, we compared M2 with two other models. First, we considered the reinforcement learning model implemented by Li et al. (2011) (model M3), which was inferior to our original model. Unlike M2, this reinforcement learning model contains only a dynamic component in its learning rate. Note that whereas the weight parameter of M2 enables us to quantify individual differences in the degree to which participants followed the Pearce-Hall associability rule, M3 does not have such parameter. In other words, under M3, all individuals equally follow the Pearce-Hall associability rule.

We then asked whether emotional cues modulate learning rate. Specifically, we considered a variant of the hybrid model M2 with emotion-specific weight parameters (model M4). This dual weight model contains separate weight parameters for happy and angry trials. We used the same Bayesian model comparison procedure to compare this model with model M2. We found that this model outperformed M2 despite the penalty for one extra parameter. We also used classical likelihood ratio tests for comparing this model (M4) with the original hybrid model (M2), as M2 is nested within M4. The results confirmed the Bayesian model comparison results indicating that the hybrid model with emotion-specific w parameters (M4) is better given the data ($\chi^2(2)=21.84$, $p<0.0001$).

[Table 1 about here]

We also considered control analyses to test modulation of M2 parameters across different factors. First, we fitted a model in which $\kappa$ rather than w was assumed to be emotion-specific (M5). This model tested the idea that emotions reduce or increase scale of learning rate regardless of the dynamics of the environment. The evidence for this model, however, was lower than that for the original one (M2) ruling out that emotions affect the overall scale of learning rate rather than its sensitivity to environmental dynamics (Table 1).

Second, we tested a control model in which the weight parameters varied as a function of the valence of the outcome (model M6). In this model, w was different for reward and punishment trials. This model also did not outperform the original model, M2. Altogether, these results suggest that emotional state modulates the degree to which people adapt their learning rate dynamically as a function of the history of surprises.

[Table 2 about here]

## Trait social anxiety predicts dynamic learning rate in states evoked by angry face cues

Trait social anxiety is a predictor of vulnerability to depression and anxiety disorders (Mineka and Oehlberg, 2008), pathologies hypothesized to be related to disrupted learning in uncertain environments (Paulus and Yu, 2012; Huys et al., 2015). Furthermore, a recent study has shown that variability in learning rate in a probabilistic learning task is associated with individual differences in trait anxiety (Browning et al., 2015). Here, we build on these prior findings by assessing whether individual differences in the effect of emotional cues on the dynamic learning rate, w, are related to individual variability in social anxiety. To this end, we tested how individual differences in parameters of the winning model, M4, are related to social anxiety. We analyzed estimated weights, w, using individually fitted parameters. Unlike parameters estimated by the hierarchical Bayesian procedure that are regularized according to all subjects' data, the individually fitted parameters are independently estimated and therefore can be used in regular statistical tests. Nonparametric Wilcoxon rank (two-tailed) tests were employed, because of the non-Gaussian distribution of the weight parameters (as they were constrained to lie in the unit range).

The weight, w, differed significantly between the low and high social anxiety groups on angry trials ($p=0.001$, $z=3.20$; Figure 3A), but not on happy trials ($p=0.56$, $z=-0.59$; Figure 3B) and the difference in weights (angry vs. happy) was also significantly different between

the two groups (p=0.033, z=2.14). Thus, participants with high versus low social anxiety exhibited reduced dynamic adjustment of learning rate on trials starting with an angry, but not a happy, face. No significant difference between the two groups was found for the other parameters of the model (all p>0.05).

[Figure 3 about here]

An obvious next question is how the low weight parameter in the high socially anxious group affected their choice. Since the weight parameter, w, indicates sensitivity of the learning rate to changes in the environment, its effects on learning is manifested in the relative performance in the stable versus volatile epochs. For example, a model with a low weight, w, would change its decisions on the basis of a few bad outcomes that could be due to noise. This model feature can cause poor performance especially in relatively stable conditions in which the action-outcome contingency does not change and optimal learning relies on a reduced learning rate.

To demonstrate this quantitatively and in a relatively theory-neutral fashion, we analyzed performance of participants on the angry trials in two different conditions. We dissociated stable and volatile epochs, depending on whether there has been at least a change in action-outcome contingencies in the last 10 preceding trials. Thus, a trial was defined as stable if no change occurred in the action-outcome contingency in the last 10 trials. Otherwise, it was defined as a volatile trial. Performance in the stable and volatile epochs was quantified in terms of the average optimal choice (i.e. the probability of choosing the action with the highest probability of winning). Since our task is stochastic (action-outcome probability is never more than 80% and there are frequent reversals) and the average length of stable blocks (with probability of 80%) was 15 trials, the window of 10 trials provide a reasonable criterion for defining stability. Note that the modeling results presented above are not sensitive to such criteria in defining stability versus volatility and

467 rather define volatility based on the sequences of choices and surprises. Nevertheless, to
468 ensure that the results presented here are robust against the 10-trial criterion, we
469 considered other definition of stability in which the window length was more than 10 trials.
470 The pattern of results found for those alternatives were consistent with the one presented
471 here.

472 First, we analyzed optimal choice probability on angry trials as a function of condition
473 (stable vs. volatile) using non-parametric Wilcoxon tests (due to its non-Gaussian
474 distribution, all tests are two-tailed). Across all participants, optimal choice probability was
475 higher for stable than volatile trials ($p<0.0001$, $z=4.04$). This is expected because making an
476 optimal choice after a change in action-outcome contingency (i.e. in volatile trials) is more
477 difficult than the stable condition in which there is no change in contingency. The important
478 question, however, is whether this analysis confirms the model-based results, which suggest
479 that social anxiety affects optimal choice probability differentially for the stable and volatile
480 conditions. As predicted, we found a significant interaction between social anxiety and
481 epoch, with the high social anxiety group showing less difference between optimal choice
482 probability in stable and volatile epochs than the low social anxiety group ($p=0.02$, $z=2.33$;
483 Figure 3C). Post-hoc tests revealed that the low social anxiety group benefited from stability
484 of the environment as their performance was significantly better in the stable than the
485 volatile epoch ($p<0.0001$, $z=3.83$). This effect was not present in the high social anxiety
486 group ($p=0.12$, $z=1.55$). Note that the difference in relative performance is not due to better
487 performance of the high social anxiety group in volatile conditions. Specifically, no significant
488 difference in optimal choice probability on the volatile epoch was found between the two
489 groups ($p=0.88$, $z=-0.15$) indicating that the high social anxiety group did not perform better
490 in volatile conditions. Significant effects were found when we considered different window
491 length for defining stability (windows with 11 to 14 trials, all p-values<0.05).

We also performed the same analysis for the happy trials, which, as predicted by the model-based analyses, did not reveal any group by epoch interaction effect ($p=0.91$, $z=-0.11$; Figure 3D).

### Trait social anxiety predicts dorsal anterior cingulate cortex activity related to learning rate in states evoked by angry face cues

The dACC has been proposed to contribute to learning from experience by computing learning rate (Behrens et al., 2007, 2008; Rushworth et al., 2011). In nonhuman primates, lesions to dACC results in an inability to use more than the most recent outcome to guide decisions (Kennerley et al., 2006). In humans, blood oxygenation level dependent (BOLD) responses in the dACC have been shown to correlate with learning rate in a probabilistic learning task. Another study using the same task has reported that the dynamic learning rate depends on trait anxiety scores (Browning et al., 2015). The next question we ask here is whether learning rate-related signals in the dACC depend on emotion-related traits, such as social anxiety, and emotional states, as manipulated using emotional facial cues.

To answer this question, we performed model-based fMRI analysis (Cohen et al., 2017) to isolate BOLD signals that correlate with learning rate in different emotional contexts. Our linear regression model included not just dynamic learning rate, but also prediction error to control for prediction error-related effects. These model-derived time series were considered as parametric regressors at the time of outcome, separately for each of the four trial-types, leading to 8 regressors. Eight regressors of no-interest were added to account for trial-type specific effects at the time of cue presentation (4 regressors) and of outcome presentation (4 regressors). To generate regressors for fMRI analysis on a common scale, we used the average parameters estimated by the hierarchical Bayesian procedure across all subjects as the common values for all parameters. This is a common approach in model-based neuroimaging, which enables us to draw inferences about individual differences in the

neural correlates of model-derived regressors (Daw et al., 2006; Daw, 2011). In other words, any effect regarding individual differences in neural correlates should be attributed to neural signal rather than the parameters used to generate regressors correlating with those signals. Importantly, we used parameters of the hybrid model M2 (rather than M4) to ensure that any difference in correlation between BOLD and learning rate in angry versus happy trials is not confounded with different weight parameters. An anatomically defined mask of the dACC (the rostral cingulate motor area in the connectivity-based parcellation atlas of medial frontal cortex (Neubert et al., 2015)) was employed for region-of-interest analysis.

In line with previous findings, we found that BOLD signal in the dACC, across all trials and participants, correlated with learning rate (bilaterally, peak at x=8, y=26 z=42, voxel-level familywise small-volume corrected at $p<0.05$; Figure 4A). Post-hoc test at the peak revealed that the effects are significantly stronger for the angry than happy trials ($t(43)=2.11$, $p=0.041$; Figure 4B). Similar effects were found when considering activity of all voxels showing a significant (at $p<0.001$ uncorrected) learning rate activity ($t(43)=2.11$, $p=0.041$). Further tests also revealed that dACC correlation with learning rate was driven by the angry trials. Specifically, BOLD signal in the dACC exhibited a significant correlation with learning rate during angry trials (bilaterally, peak at x=-8, y=24 z=40, voxel-level familywise small-volume corrected at $p<0.05$), but not during happy trials (no voxel survived uncorrected threshold of 0.001). Therefore, we focused on angry trials and asked whether high social anxiety individuals show weaker learning rate related activity than the low social anxiety group, as suggested by the modeling findings.

We found that individual differences in social anxiety covaried strongly with learning rate-related signals in the dACC on angry trials (Figure 4C). Specifically, the learning rate signal in the dACC during angry trials (at the peak voxel x=-8, y=24, z=40) was stronger for the low than the high social anxiety group ($t(42)=3.05$, $p=0.004$). Similar effects were found

when considering activity of all voxels showing a significant (at $p<0.001$ uncorrected) learning rate activity on angry trials ($t(42)=2.37$, $p=0.023$). Post-hoc tests at the peak voxel revealed that the high social anxiety group did not show a significant correlation ($t(20)=0.93$, $p=0.36$). These results demonstrate that, compared with the low social anxiety group, the high social anxiety dynamically adapted their learning rate to a lesser degree on trials involving presentation of an angry face. Moreover, unlike the low social anxiety group, their dACC BOLD signal did not covary with the learning rate on these trials.

[Figure 4 about here]

We looked at two control contrasts in the above neuroimaging analysis. First, we found strong prediction error related signal in the ventral striatum (bilaterally, peak at 14, 12, -8, voxel-level familywise small-volume corrected at $p<0.05$), consistent with previous studies (McClure et al., 2003; O'Doherty et al., 2003; Daw et al., 2006). Second, we performed a region-of-interest analysis in the amygdala. We focused on the amygdala given its important role in emotional processing (Weiskrantz, 1956; Ledoux, 1996; Phelps and LeDoux, 2005), and previous reports on amygdala sensitivity to learning rate (Li et al., 2011). Despite the presence of clear emotion-related main effects of cue in the amygdala (bilaterally, peak at -14, -8, -16, voxel-level familywise small-volume corrected at $p<0.05$), with stronger signal during the presentation of the angry faces, there were no significant effects of learning rate in the amygdala ($p<0.001$ uncorrected).

[Table 3 about here]

## Discussion

In daily life, it is important to adaptively learn from the outcomes of our decisions, even in environments with threat cues. The adaptation should depend on the history of

565 outcomes and the degree to which those previous outcomes were surprising. When the
566 environment is full of surprises, recent experiences are more predictive of future events
567 than remote experiences. In those circumstances, a higher learning rate is optimal. We
568 found evidence that social anxiety is associated with reduced adaptation of learning rate,
569 particularly in aversive states, such as those evoked here by exposure to images of angry
570 faces.

571 Our findings are in line with theories looking at psychiatric disorders linked to social
572 anxiety from the perspective of decision neuroscience (Hartley and Phelps, 2012; Paulus and
573 Yu, 2012; Huys et al., 2015). These disorders are hypothesized to be accompanied by deficits
574 in learning and decision making, particularly in uncertain environments requiring dynamic
575 learning (Paulus and Yu, 2012; Browning et al., 2015). Here, we focused on trait social
576 anxiety in healthy participants, as trait social anxiety is a factor predicting vulnerability to
577 anxiety and depression (Barlow, 2004; Mineka and Zinbarg, 2006; Mineka and Oehlberg,
578 2008). Our data indicate the presence of maladaptive biases in learning, at both
579 computational and neural levels, even in healthy individuals. These findings suggest a
580 particular computational mechanism by which social anxiety might impact decisions in
581 threatening situations. In those situations, the weight of dynamic learning rate is too low for
582 anxious individuals, making them oversensitive to noisy outcomes of their decisions.
583 Suboptimal decisions and oversensitivity to outcomes exacerbate each other, generating a
584 dysfunctional loop.

585 Inspired by these modeling results, we found signatures of disrupted adaptation of
586 learning rate in the behavioral data (Figure 3C). In threatening situations evoked by angry
587 face images, the high social anxiety group did not benefit from stability in the environment
588 and showed similar levels of performance in both stable and volatile situations. In contrast,
589 the low social anxiety group showed a much better performance in the stable situation

compared with the volatile situation. These results are consistent with a recent report by Browning and colleagues (Browning et al., 2015). They showed that anxiety is associated with inability to adjust learning in stable and volatile situations. Our data adds to those findings by showing that inability in optimal learning is also a function of emotional cues. Furthermore, our findings elucidate corresponding neural mechanisms in socially anxious individuals by showing that disruption in optimal learning is accompanied by disruption in dACC activity related to learning rate. The dACC has been argued to specifically contribute to reinforcement learning by computing learning rate in uncertain environments (Behrens et al., 2007, 2008; Rushworth et al., 2011). However, so far, it has remained unclear whether dACC computations of learning rate are modulated by emotional cues or by traits such as social anxiety. Showing those modulations is particularly important, because the dACC is a central node of the brain system processing negative affect (Shackman et al., 2011), suggesting that its computations might be sensitive to negative emotions. Here, we replicated previous findings, namely covariation between dACC activity and learning rate (Behrens et al., 2007, 2008). Furthermore, we added to those reports by demonstrating that learning rate-related computations are stronger when responding to emotional cues. More importantly, our results suggest that high socially anxious individuals show disrupted dACC activity in relation to learning rate.

Influences of emotional conditioned stimuli on optimal learning, as found in this study, might be due to effects of those stimuli on emotions, and subsequent effects of negative emotions on optimal learning and decision making. Another possibility is that social threat cues disrupt optimal learning directly, even when they are not accompanied with negative emotions. Future studies should address this question, in particular by analyzing choice data and simultaneously-recorded physiological signals related to experienced emotions, such as skin conductance response. Importantly, although current research on defensive behavior is mainly focused on elicited reactions, new theories emphasize active responses to threat

cues (LeDoux and Daw, 2018). The neural processes underlying those active responses are not yet clear, although amygdala is hypothesized to influence active decisions by signaling threats to the striatum (LeDoux and Daw, 2018), which plays a key role in learning and decision making. The role of the dACC in these neural processes are not yet known, although dACC has dense connectivity with both the amygdala and the striatum (Draganski et al., 2008; Shackman et al., 2011).

In this study, in addition to emotional content of conditioned stimuli, we manipulated valence of outcomes independently. However, no significant effect of outcome valence on optimal tuning of learning rate was found. Nevertheless, further studies are needed to investigate effects of outcome valence on optimal learning. First, optimal learning might be more sensitive to primary punishments such as shocks. In this study, however, we used monetary `outcomes as instrumental reinforcers both as reward and punishment. Second, the outcome manipulation of the present study might not be sufficiently powerful to be detected in our sample size. Third, in our paradigm, the punishment is avoidable (outcome contingency is instrumental), while the facial expression is not. This difference might lead to potentiated effects for the negative facial expression versus the negative outcome.

In this study, unlike the recent study by Li et al. (2011), we did not find associability related activity in the amygdala, even when we focused only on angry trials. However, there are important differences between the paradigm used in this study and that of Li et al. First, Li et al. used shocks as negative outcomes, whereas we used financial losses as negative outcomes. Second, Li et al., fitted their model to skin conductance response data, whereas we fitted models to choice data. Finally, Li and colleagues examined amygdala activation in the context of a Pavlovian task that did not require making decisions, whereas the current study required decision making. Consistent with our findings, a recent study in monkeys did not find significant effects of amygdala lesions on associability in a stochastic two-arm

641 bandit task (Costa et al., 2016). It should be noted, however, that the role of amygdala
642 regarding associability computations in threat situations might be to signal presence of
643 threat to other regions (Fox et al., 2015), such as dACC.

644 The biases induced by threatening social cues, such as angry faces, reflect Pavlovian
645 biases in learning. These Pavlovian biases are not always the most rational responses, but
646 they are generally useful heuristics as they reflect predominant statistics of the environment
647 around us, for example threatening angry cues are more likely to be followed by negative
648 outcomes. Importantly, unlike Pavlovian response biases, such Pavlovian learning biases
649 affect causal inference. Therefore, our findings suggest that threatening angry cues affect
650 how high trait social anxiety individuals make causal inference. In the context of social threat
651 cues, those individuals are unable to dissociate a bad outcome that happened by chance
652 from an actual mistake caused by their own actions. This might be related to symptoms of
653 "self-blame" in anxiety and depression disorders(Beck, 1967), although further studies are
654 needed to investigate this somewhat speculative hypothesis.

655 Previous works have linked Pavlovian biases to neuromodulatory systems (den Ouden
656 et al., 2013; Swart et al., 2017), particularly dopaminergic (although see the recent study by
657 Rutledge et al. (Rutledge et al., 2017)) and serotonergic systems. Whether and how these, or
658 other neuromodulatory (Iglesias et al., 2013; Payzan-LeNestour et al., 2013), systems
659 modulate such Pavlovian biases in learning rate in socially anxious individuals are open
660 questions for future studies.

661 Psychological, temporal difference and Bayesian accounts of learning suggest that
662 learning rate is a crucial element of learning, which should be adaptively adjusted according
663 to the history of surprises to support optimal learning (Pearce and Hall, 1980; Yu and Dayan,
664 2005; Behrens et al., 2007; Li et al., 2011; Mathys et al., 2011; Iglesias et al., 2013). Here, we
665 used an augmented hybrid Rescorla-Wagner model in which learning rate was a weighted

combination of a dynamic and a constant component. The dynamic component was gradually updated according to the sample variance (squared error) on every trial. The hybrid model can be treated as a proxy model of fully Bayesian accounts, which has the benefit to be close to classical psychological models. An important open question for future studies is whether the inability to adjust learning rate in socially anxious individuals is caused by disruptions in computationally higher levels of reasoning that are responsible for detecting changes in the environment. Hierarchical Bayesian models are particularly useful to address this question (Behrens et al., 2007). Another important question remained to be addressed is whether these hierarchically-computed learning rates vary as a function of the valence of prediction errors, which is shown to influence baseline learning rates in humans (Frank et al., 2004, 2007; Piray et al., 2014) as well as monkeys (Piray, 2011) and supported by neural models of prefrontal cortex–basal ganglia (Frank et al., 2004; O'Reilly and Frank, 2006) and mesostriatal circuits (Haber et al., 2000; Piray et al., 2017).

In this study, we characterized the computational and neural mechanisms by which emotional context modulated optimal learning in an uncertain environment and how those mechanisms are disrupted in high trait social anxious individuals. These findings open the way to test and modify the neurobiological underpinnings of maladaptive learning in pathologies related to social anxiety.

685 ## References


686 Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26:839–851.

687 Barlow DH (2004) Anxiety and Its Disorders: The Nature and Treatment of Anxiety and
688 Panic, 2 edition. New York, NY: The Guilford Press.

689 Beck AT (1967) Depression: Clinical, Experimental, and Theoretical Aspects. University of
690 Pennsylvania Press.

691 Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008) Associative learning of social
692 value. Nature 456:245–249.

693 Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of
694 information in an uncertain world. Nat Neurosci 10:1214–1221.

695 Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ (2015) Anxious individuals
696 have difficulty learning the causal statistics of aversive environments. Nat Neurosci
697 18:590–596.

698 Cohen JD (2005) The Vulcanization of the Human Brain: A Neural Perspective on
699 Interactions Between Cognition and Emotion. J Econ Perspect 19:3–24.

700 Cohen JD, Daw N, Engelhardt B, Hasson U, Li K, Niv Y, Norman KA, Pillow J, Ramadge PJ,
701 Turk-Browne NB, Willke TL (2017) Computational approaches to fMRI analysis.
702 Nat Neurosci 20:304–313.

703 Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB (2016) Amygdala and Ventral
704 Striatum Make Distinct Contributions to Reinforcement Learning. Neuron 92:505–
705 517.

706 Daw ND (2011) Trial-by-trial data analysis using computational models. In: Decision Making,
707 Affect, and Learning: Attention and Performance XXIII (Delgado MR, Phelps EA,
708 Robbins TW, eds), pp 3–38. New York: Oxford University Press.

709 Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for
710 exploratory decisions in humans. Nature 441:876–879.

711 de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, Bestmann S (2016)
712 Computations of uncertainty mediate acute stress responses in humans. Nat Commun
713 7:10996.

714 den Ouden HEM, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Franke B,
715 Cools R (2013) Dissociable effects of dopamine and serotonin on reversal learning.
716 Neuron 80:1090–1100.

717 Domes G, Schulze L, Böttger M, Grossmann A, Hauenstein K, Wirtz PH, Heinrichs M,
718 Herpertz SC (2010) The neural correlates of sex differences in emotional reactivity
719 and emotion regulation. Hum Brain Mapp 31:758–769.

720 Draganski B, Kherif F, Klöppel S, Cook PA, Alexander DC, Parker GJM, Deichmann R,
721 Ashburner J, Frackowiak RSJ (2008) Evidence for Segregated and Integrative
722 Connectivity Patterns in the Human Basal Ganglia. J Neurosci 28:7143–7152.

723 Dreisbach G, Goschke T (2004) How positive affect modulates cognitive control: reduced
724 perseveration at the cost of increased distractibility. J Exp Psychol Learn Mem Cogn
725 30:343–353.

726 Ekman P, Friesen, WV (1976) Pictures of Facial Affect. Palo Alto, CA: Consulting
727 Psychologist Press. Available at: http://www.paulekman.com/product/pictures-of-
728 facial-affect-pofa/ [Accessed December 11, 2015].

729 Fox AS, Oler JA, Tromp DPM, Fudge JL, Kalin NH (2015) Extending the amygdala in
730 theories of threat processing. Trends Neurosci 38:319–329.

731 Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple
732 dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl
733 Acad Sci U S A 104:16311–16316.

734 Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement
735 learning in parkinsonism. Science 306:1940–1943.

736 Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an
737 ascending spiral from the shell to the dorsolateral striatum. J Neurosci 20:2369–2382.

738 Hartley CA, Phelps EA (2012) Anxiety and decision-making. Biol Psychiatry 72:113–118.

739 Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling
740 the roles of approach, activation and valence in instrumental and pavlovian
741 responding. PLoS Comput Biol 7:e1002028.

742 Huys QJM, Daw ND, Dayan P (2015) Depression: a decision-theoretic analysis. Annu Rev
743 Neurosci 38:1–23.

744 Huys QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai trees in your
745 head: how the pavlovian system sculpts goal-directed choices by pruning decision
746 trees. PLoS Comput Biol 8:e1002410.

747 Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, Stephan KE
748 (2013) Hierarchical prediction errors in midbrain and basal forebrain during sensory
749 learning. Neuron 80:519–530.

750 Kahneman D (2011) Thinking, Fast and Slow, 1st ed. New York: Farrar, Straus and Giroux.

751 Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth MFS (2006) Optimal
752 decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

753 Koch K, Pauly K, Kellermann T, Seiferth NY, Reske M, Backes V, Stöcker T, Shah NJ,
754 Amunts K, Kircher T, Schneider F, Habel U (2007) Gender differences in the
755 cognitive control of emotion: An fMRI study. Neuropsychologia 45:2744–2754.

756 Ledoux J (1996) The EMOTIONAL BRAIN: The Mysterious Underpinnings of Emotional
757 Life, 1st ed. New York: Simon & Schuster.

758 LeDoux J, Daw ND (2018) Surviving threats: neural circuit and computational implications
759 of a new taxonomy of defensive behaviour. Nat Rev Neurosci 19:269–282.

760 Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. Annu Rev
761 Psychol 66:799–823.

762 Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human
763 striatum and amygdala in associative learning. Nat Neurosci 14:1250–1252.

764 Liebowitz MR (1987) Social phobia. Mod Probl Pharmacopsychiatry 22:141–173.

765 Lundqvist D, Flykt A, Öhman A (1998) Karolinska Directed Emotional Faces (KDEF). CD
766 ROM from Department of Clinical Neuroscience, Psychology section, Karolinska
767 Institutet. Available at: http://www.emotionlab.se/resources/kdef [Accessed
768 December 11, 2015].

769 Ly V, Cools R, Roelofs K (2014) Aversive disinhibition of behavior and striatal signaling in
770 social avoidance. Soc Cogn Affect Neurosci 9:1530–1536.

771 MacKay DJC (2003) Information Theory, Inference, and Learning Algorithms. Cambridge
772 University Press.

773 Martinez, AM, Benavente, R (1998) The AR Face Database. CVC Tech Rep No 24.

774 Mason L, Eldar E, Rutledge RB (2017) Mood Instability and Reward Dysregulation-A
775 Neurocomputational Model of Bipolar Disorder. JAMA Psychiatry 74:1275–1276.

776 Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A bayesian foundation for individual
777 learning under uncertainty. Front Hum Neurosci 5:39.

778 Matsumoto D, Ekman P (1988) Japanese and Caucasian facial expressions of emotion
779 (JACFEE). San Francisco (CA): University of California: Human Interaction
780 Laboratory.

781 McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive
782 learning task activate human striatum. Neuron 38:339–346.

783 Mineka S, Oehlberg K (2008) The relevance of recent developments in classical conditioning
784 to understanding the etiology and maintenance of anxiety disorders. Acta Psychol
785 (Amst) 127:567–580.

786 Mineka S, Zinbarg R (2006) A contemporary learning theory perspective on the etiology of
787 anxiety disorders: it's not what you thought it was. Am Psychol 61:10–26.

788 Morel P (2018) Gramm: grammar of graphics plotting in Matlab. J Open Source Softw 3:568.

789 Neubert F-X, Mars RB, Sallet J, Rushworth MFS (2015) Connectivity reveals relationship of
790 brain areas for reward-guided learning and decision making in human and monkey
791 frontal cortex. Proc Natl Acad Sci U S A 112:E2695-2704.

792 O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models
793 and reward-related learning in the human brain. Neuron 38:329–337.

794 O'Reilly RC, Frank MJ (2006) Making working memory work: a computational model of
795 learning in the prefrontal cortex and basal ganglia. Neural Comput 18:283–328.

796 Paulus MP, Yu AJ (2012) Emotion and decision-making: affect-driven belief systems in
797 anxiety and depression. Trends Cogn Sci 16:476–483.

798 Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP (2013) The neural representation
799 of unexpected uncertainty during value-based decision making. Neuron 79:191–201.

800 Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of
801 conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.

802 Phelps EA, LeDoux JE (2005) Contributions of the amygdala to emotion processing: from
803 animal models to human behavior. Neuron 48:175–187.

804 Phelps EA, Lempert KM, Sokol-Hessner P (2014) Emotion and decision making: multiple
805 modulatory neural circuits. Annu Rev Neurosci 37:263–287.

806 Piray P (2011) The role of dorsal striatal D2-like receptors in reversal learning: a
807 reinforcement learning viewpoint. J Neurosci 31:14049–14050.

808 Piray P, den Ouden HEM, van der Schaaf ME, Toni I, Cools R (2017) Dopaminergic
809 Modulation of the Functional Ventrodorsal Architecture of the Human Striatum.
810 Cereb Cortex 27:485–495.

811 Piray P, Toni I, Cools R (2016) Human Choice Strategy Varies with Anatomical Projections
812 from Ventromedial Prefrontal Cortex to Medial Striatum. J Neurosci 36:2857–2867.

813 Piray P, Zeighami Y, Bahrami F, Eissa AM, Hewedi DH, Moustafa AA (2014) Impulse
814 control disorders in Parkinson's disease are associated with dysfunction in stimulus
815 valuation but not action valuation. J Neurosci 34:7814–7824.

816 Poser BA, Versluis MJ, Hoogduin JM, Norris DG (2006) BOLD contrast sensitivity
817 enhancement and artifact reduction with multiecho EPI: parallel-acquired
818 inhomogeneity-desensitized fMRI. Magn Reson Med 55:1227–1235.

819 Rescorla RA, Wagner AR, Black AH, Prokasy WF (1972) A theory of Pavlovian
820 conditioning: Variations in the effectiveness of reinforcement. In: Classical
821 Conditioning II: Current Research and Theory, pp 64–69. New York: Appleton
822 Century-Crofts.

823 Roelofs K, Minelli A, Mars RB, van Peer J, Toni I (2009) On the neural control of social
824 emotional behavior. Soc Cogn Affect Neurosci 4:50–58.

825 Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex
826 and reward-guided learning and decision-making. Neuron 70:1054–1069.

827 Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, Lam J,
828 Skandali N, Siegel JZ, Ousdal OT, Prabhu G, Dayan P, Fonagy P, Dolan RJ (2017)
829 Association of Neural and Emotional Impacts of Reward Prediction Errors With
830 Major Depression. JAMA Psychiatry 74:790–797.

831 Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014)
832 Automatic denoising of functional MRI data: combining independent component
833 analysis and hierarchical fusion of classifiers. NeuroImage 90:449–468.

834 Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ (2011) The
835 integration of negative affect, pain and cognitive control in the cingulate cortex. Nat
836 Rev Neurosci 12:154–167.

837 Swart JC, Froböse MI, Cook JL, Geurts DE, Frank MJ, Cools R, den Ouden HE (2017)
838 Catecholaminergic challenge uncovers distinct Pavlovian and instrumental
839 mechanisms of motivated (in)action. eLife 6.

840 van Peer JM, Roelofs K, Rotteveel M, van Dijk JG, Spinhoven P, Ridderinkhof KR (2007)
841 The effects of cortisol administration on approach-avoidance behavior: an event-
842 related potential study. Biol Psychol 76:135–146.

843 van Steenbergen H, Band GPH, Hommel B (2010) In the mood for adaptation: how affect
844 regulates conflict-driven control. Psychol Sci 21:1629–1634.

845 Weiskrantz L (1956) Behavioral changes associated with ablation of the amygdaloid complex
846 in monkeys. J Comp Physiol Psychol 49:381–391.

847 Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.

848
849

## Legends

Table 1. Bayesian model comparison. For each model, differential log-model evidence is shown. Higher values indicate more evidence in favor of the model. The hybrid model with emotion-specific w (M4) has the highest Bayesian model evidence among all models. Note that models are only different in the number of learning parameters. Additionally, all models contain 4 parameters for generating choice including 3 value-independent biases in making a go or no-go response and 1 inverse-temperature parameter. See Methods for formal definition of all models. See Table 2 for further statistics on fitted parameters of the winning model.

Table 2. Fitted parameters of the winning model (Hybrid emotion-specific w model) individually using maximum a posteriori (MAP) and the hierarchical fitting procedure (HFP).

Table 3. Statistics of the neuroimaging analysis of the main effects of learning rate in the dACC mask (small-volume voxel-level familywise error corrected).

Figure 1. Probabilistic reversal learning task. A) Timeline of the task. Participants had to respond (either go or no-go) after a face cue was presented. A probabilistic outcome was presented following a delay. Importantly, the presented cues were used as conditioned stimuli and the optimal response is a function of the probability of a win given a go, which varied across trials, independent of the emotional content of the cue (see panel C). B) There were four different trial-types in the task. The cue of each trial could vary in emotional content (an angry or happy face) and in color (grey, yellow). The color indicates whether the outcome of a trial is a reward or punishment. C) An example of probability sequence of win given a go response for one of the four trial-types. The dots show the actual feedback seen by the participant, which are drawn from this probability distribution. Probability of a win given the no-go response is the reverse of the probability of a win given the go response (e.g.

874 when probability of a win given go is 0.8, probability of win given no-go is 0.2). Note that the
875 probability sequence (and thus the optimal response) is reversed multiple times for each cue.
876 The underlying probability sequence is manipulated independently for each cue.

877 Figure 2. Performance after reversals. Data has been shown separately for each trial type
878 and response type (go vs. nog-go). Performance was about the chance level immediately
879 after reversal and improved over the course of learning. For each participant, a learning
880 curve is defined by averaging performance after multiple reversals occurred for the
881 corresponding trial type and response type. Mean learning curve across all participants and
882 corresponding standard error of the mean are plotted. Note that the x-axis displays trials
883 after reversals (reversal occurred at trial 0). Abbreviations: HR, happy and reward trial type;
884 HP, happy and punishment trial type; AR, angry and reward trial type; AP, angry and
885 punishment trial type.

886 Figure 3. Effects of social anxiety on learning rate: A and B) Effects of social anxiety (SA) on
887 the weight given to the dynamic component of the learning rate in angry (A) and happy (B)
888 trials. High socially anxious individuals showed less dynamic adjustment of learning rate
889 (indexed by parameter w) on angry trials. There was no effect of social anxiety on dynamic
890 learning rate on happy trials. C and D: Effects of social anxiety on performance in stable and
891 volatile epochs in angry (C) and happy (D) trials. In line with the weight parameter results,
892 high socially anxious individuals showed less benefit of stability in their performance than
893 the low social anxiety group on angry trials. There was no significant interaction on the
894 happy trials. Standard boxplots are plotted, in which the box is drawn between the 25 and
895 75 percentiles with a line indicating the median. The distribution of data is also shown.

896 Figure 4. Effects of social anxiety (SA) on learning rate-related activity in the dACC. A) Across
897 all trials, dACC correlated with the learning rate (LR), although B) the effects were stronger
898 during angry trials and was mainly driven by the angry trials. C) On angry trials, learning rate-

related activity in the dACC was present in the low SA group, which showed significantly greater activity than the high SA group. For a, *t*-map (degrees of freedom=43) is shown with all voxels within the anatomical mask of dACC with $p<0.005$ (uncorrected for display). For B, the effect size is defined as the beta coefficient at the peak voxel of learning rate-related activity in the dACC. For C, the effect size is defined as the beta coefficient at the peak voxel of learning rate-related activity in the dACC on angry trials. In C and C, the corresponding mean and standard errors of the mean are plotted. The distribution of data is also shown. See Table 3 for further details of statistical analysis.

908 ## Tables

| | Model | No free parameters | Relative log model evidence |
|---|---|---|---|
| M1 | Rescorla-Wagner | 5 | -15.02 |
| M2 | Hybrid | 7 | -7.13 |
| M3 | Li et al. (2011) model | 6 | -14.78 |
| **M4** | **Hybrid emotion-specific w** | 8 | 0 |
| M5 | Hybrid emotion-specific $\kappa$ | 8 | -7.77 |
| M6 | Hybrid valence-specific w | 8 | -8.85 |

909

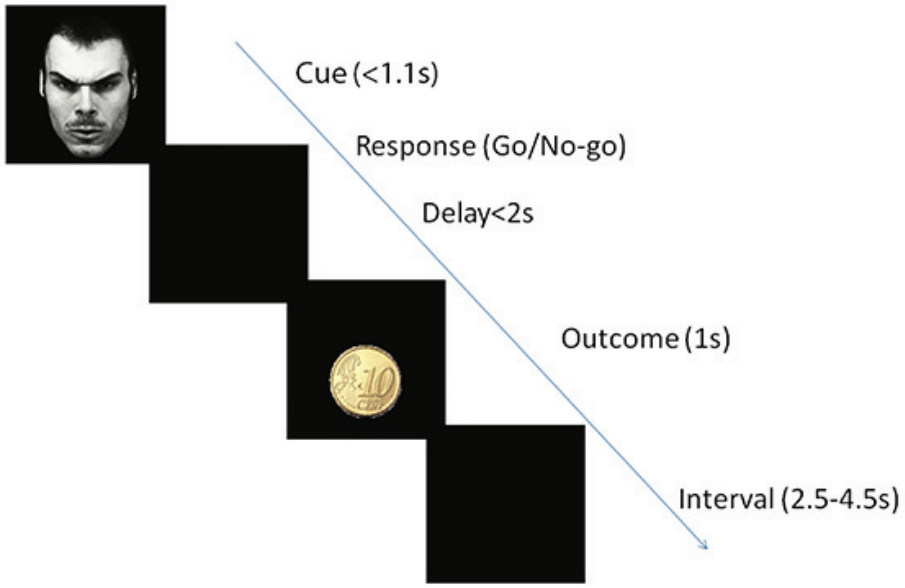
| | MAP 25$^{th}$ percentile | MAP median | MAP 75$^{th}$ percentile | HFP group mean ( $\Theta$ ) |
|---|---|---|---|---|
| $w_h$ | 0.2 | 0.372 | 0.546 | 0.397 |
| $w_a$ | 0.21 | 0.426 | 0.649 | 0.403 |
| $\kappa$ | 0.789 | 0.868 | 0.909 | 0.922 |
| $\lambda$ | 0.404 | 0.539 | 0.673 | 0.458 |
| $\beta$ | 1.321 | 1.822 | 2.619 | 1.646 |
| $b_v$ | 0.021 | 0.181 | 0.339 | 0.152 |
| $b_e$ | -0.12 | 0.029 | 0.181 | 0.032 |
| $b_i$ | -0.052 | 0.056 | 0.14 | 0.045 |

910

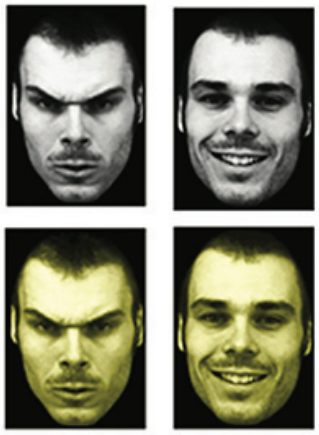
| | Cluster-level statistics | | Voxel-level statistics | | | | |
|---|---|---|---|---|---|---|---|
| | $P_{FWE}$ | k | $P_{FWE}$ | T(43) | Peak (mm) | | |
| Learning rate across all trials | 0.034 | 38 | 0.032 | +3.75 | 8 | 26 | 42 |
| | 0.038 | 32 | 0.035 | +3.72 | -10 | 18 | 44 |
| Learning rate in angry trials | 0.017 | 77 | 0.013 | +4.14 | -8 | 24 | 40 |
| | 0.033 | 38 | 0.017 | +4.03 | 10 | 28 | 42 |

911

A




B



C