Social Cognitive and Affective Neuroscience Advance Access published May 7, 2010

doi:10.1093/scan/nsq041 SCAN (2010) 1 of 10

Social stress reactivity alters reward and punishment learning

James F. Cavanagh,[1] Michael J. Frank,[2] and John J. B. Allen[1]
[1]University of Arizona, Department of Psychology, 1503 E University Blvd, Tucson AZ 85721 and [2]Brown University, Department of Psychology, 89 Waterman Street, Providence RI 02912, USA

To examine how stress affects cognitive functioning, individual differences in trait vulnerability (punishment sensitivity) and state reactivity (negative affect) to social evaluative threat were examined during concurrent reinforcement learning. Lower trait-level punishment sensitivity predicted better reward learning and poorer punishment learning; the opposite pattern was found in more punishment sensitive individuals. Increasing state-level negative affect was directly related to punishment learning accuracy in highly punishment sensitive individuals, but these measures were inversely related in less sensitive individuals. Combined electrophysiological measurement, performance accuracy and computational estimations of learning parameters suggest that trait and state vulnerability to stress alter cortico-striatal functioning during reinforcement learning, possibly mediated via medio-frontal cortical systems.

Keywords: social evaluative threat; computational psychiatry; reinforcement learning; anterior cingulate; EEG

How do individual differences in stress reactivity affect cognitive functioning? Substantial evidence exists to implicate stress reactivity as a curvilinear determinant of prefrontal executive and sub-cortical monoamine functioning, which are important for goal directed engagement with the environment (Arnsten, 1998). The varying effects of stress on performance are often described as an 'inverted-U', but few studies have investigated neural responses that differentiate facilitative from debilitating effects of stress. Increasing evidence has shown that both trait vulnerability and state emotional reactivity influence the psychobiological and neural response to social evaluative threat (Kemeny, 2003; Pruessner *et al.*, 2004; Lerner *et al.*, 2007; Cavanagh and Allen, 2008), suggesting that these individual differences may differentiate facilitative *vs* debilitating alterations in neural functioning. In this report, emotional reactivity to social evaluative stress is shown to facilitate punishment learning in highly stress vulnerable individuals, yet impede punishment learning in less vulnerable individuals. Stress-related alteration in a medio-frontal action monitoring system is identified as a possible mechanism by which this reinforcement learning bias is instantiated.

Received 16 December 2009; Accepted 8 April 2010

The authors thank Christina Figueroa, Amanda Halawani, Alhondra Felix, Devin Brooks, Rebecca Reed, Roxanne Raifepour, Katie Yeager and Olivia Sana for help running participants. Photographs courtesy of Purple X Photography. National Institute of Mental Health (F31MH082560 to J.F.C.); National Institute of Drug Addiction (R21DA022630 to M.J.F.); Infrastructure provided by National Institute of Mental Health (R01MH066902 to J.J.B.A.).

Correspondence should be addressed to James F. Cavanagh, Department of Psychology, University of Arizona, 1503 University Ave, Tucson, AZ, USA. E-mail: jim.f.cav@gmail.com

Stress and the anterior cingulate cortex

Located within the medial prefrontal cortex (mPFC), the anterior cingulate cortex (ACC) is consistently activated in neuroimaging investigations of stress reactivity (Dedovic *et al.*, 2009). It has been proposed that the ACC evaluates the uncontrollability of a stressor (Amat *et al.*, 2005, 2006), subsequently determining stress-related subcortical monoaminergic responses (Pascucci *et al.*, 2007) and hormonal reactivity (Diorio *et al.*, 1993; Radley *et al.*, 2009). Cavanagh and Allen (2008) previously investigated individual differences in stress reactivity and ACC functioning as reflected by the error-related negativity (ERN), a scalp-measured electrical voltage deflection occurring ~80 ms after an erroneous response (Gehring *et al.*, 1993). The ERN is generated by theta band phase resetting and enhancement (Trujillo and Allen, 2007) in the ACC and surrounding mPFC, and is thought to reflect the functions of an action monitoring system that uses signals of error, conflict or punishment to adapt future behavior (Debener *et al.*, 2005; Cavanagh *et al.*, 2009). Cavanagh and Allen (2008) observed that under social evaluative threat, ERN amplitudes were altered in an inverted-U type fashion as a joint function of trait vulnerability (punishment sensitivity) and state reactivity (negative affect), with the highly punishment sensitive group additionally characterized by poorer post-error accuracy. These findings suggest that ERN amplitudes may index stress-related alteration of ACC functioning, possibly predicting consequences in effective behavioral adaptation.

Learning and the anterior cingulate cortex

A novel opportunity to test the interaction between altered ACC functioning and cognitive/behavioral consequence

exists in the field of reinforcement learning, where larger ERN amplitudes reliably predict behavioral accuracy in punishment avoidance learning (Frank *et al.*, 2005, 2007a; Grundler *et al.*, 2009). During reinforcement learning, feedback- and response-related ACC activities become inversely related as learning progresses from reliance on external feedback (larger feedback-related activity to negative prediction errors) towards reliance on internal representations (larger response-related activity) (Holroyd *et al.*, 2004; Mars *et al.*, 2005). This inverse temporal relationship is reflected in feedback- and response-locked event-related voltage potential and theta band amplitudes, and is thought to depend on the slow cortico-striatal computation of 'action values' (Holroyd and Coles, 2002; Frank *et al.*, 2007c; Cavanagh *et al.*, 2010). These phasic theta band activities during reinforcement learning allow for a measurement of ACC and mPFC systems involved in punishment avoidance learning.

Stress, biased learning and the anterior cingulate cortex

We hypothesized that in addition to being an apparent determinant of stress reactivity, altered processing in the ACC could bias the cortico-striatal computation of action values. In line with this idea, recent computational efforts have detailed mechanisms by which PFC may bias striatal functioning during reinforcement learning, enhancing learning in the striatum when beliefs are consistent with outcomes and discounting learning when inconsistent (Doll *et al.*, 2009; Huys and Dayan, 2009). While learning can be ultimately assessed by performance accuracy, different brain regions may contribute to performance across different time scales. Computational estimations based on trial-to-trial performance patterns have proposed to parse PFC and striatal contributions during learning: separating the rapid, trial-to-trial adaptive (putatively PFC) learning rate expressed during training from the slow integrative (putatively striatal) learning rate revealed during subsequent testing (Frank *et al.*, 2007b). This latter, slowly integrative system is ultimately responsible for accurate action value computation, and is hypothesized here to be biased by stress-related ACC activities during training (akin to Doll *et al.*, 2009; Huys and Dayan, 2009).

In the current investigation, it was hypothesized that trait and state vulnerability to social evaluative threat may alter cortico-striatal functioning during reinforcement learning. Altered estimation of reward or punishment values, mediated by stress-altered ACC functioning, may be a mechanism by which stress reactivity affects cognitive functioning.

METHODS

Participants

Participants were 50 students (26 females) with a mean age of 18.9 years (s.d. = 1). All participants gave informed consent and the research ethics committee of the University of Arizona approved the study. Participants were invited to a screening session if they indicated low levels of depressive symptomatology on the Beck Depression Inventory (BDI score <7) during a pretest in introductory psychology classes. The screening session was used to identify participants who fit the recruitment criterion for the electroencephalogram (EEG) session: (i) aged 18–25, (ii) stable low BDI (<7) and no self-reported history of major depressive disorder, (iii) no current psychoactive medication use, (iv) no history of head trauma or seizures and (v) no self-reported symptoms indicating a possibility of an Axis I disorder, as indicated by self-reported computerized completion of the Electronic Mini International Neuropsychiatric Interview (eMINI; Medical Outcome Systems, Jacksonville, FL, USA). All experiments were run by the same male lead experimenter (J.F.C.) and a female assistant (varied). Additional methodological details can be found in the Supplementary Data.

Questionnaires

Pre-task questionnaires included questions about demographics and health, as well as the Carver and White (1994) Behavioral Inhibition Scale/Behavioral Activation Scale (BIS/BAS). The BIS scale was the primary measure of trait-level punishment sensitivity (Cavanagh and Allen, 2009). Post-task questionnaires included retrospective appraisals of emotional experience based on endorsement of an emotional adjective word list rated on a scale of 0 (not the slightest bit) to 8 (most in your life) (Gross and Levenson, 1995; Lerner *et al.*, 2007; Cavanagh and Allen, 2008). All negative emotionality measures were aggregated (fear, nervous, anxious, afraid, anger, irritation, frustration, ashamed, humiliated, embarrassed, self-conscious) the coefficient alpha for the measure of aggregate negative affect was $\alpha = 0.91$.

Probabilistic learning task

The probabilistic learning tasks consisted of a forced choice training phase followed by a subsequent testing phase (Frank *et al.*, 2004). During the training phase the participants were presented with three stimulus pairs, where each stimulus was associated with a different probabilistic chance of receiving 'Correct' or 'Incorrect' feedback. These stimulus pairs (and their probabilities of reward) were termed A/B (80%/20%), C/D (70%/30%) and E/F (60%/40%), see Figure 1. Over the course of the training phase, a participant usually learns to choose the optimal stimulus, solely due to adaptive responding based on the feedback. Under initial, benign conditions (T1), participants underwent training trials (consisting of 1–6 blocks of 60 stimuli each) until they reached a minimum criterion of choosing the probabilistically best stimulus in each pair ($AB \geq 65\%$, $CD \geq 60\%$ and $EF \geq 50\%$ correct choices). During the stress manipulation (T2), participants performed a fixed set of four training blocks before being moved to the test phase.

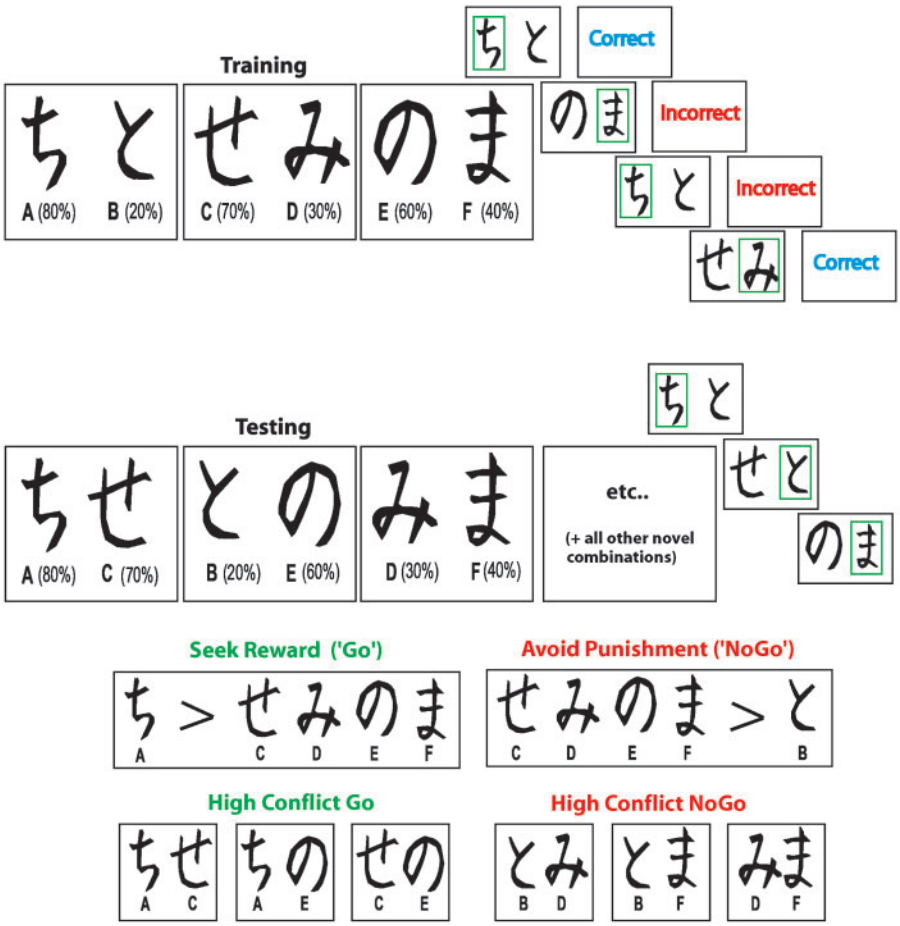

Fig. 1 Probabilistic learning task. During training, each pair is presented separately. Participants have to select one of the two stimuli, slowly integrating 'Correct' and 'Incorrect' feedback (each stimulus has a unique probabilistic chance of being 'Correct') in order to maximize their accuracy. During the testing phase, each stimulus is paired with all other stimuli and participants must choose the best one, without the aid of feedback. Note that the letter and percentage are not presented to the participant, nor are the green boxes surrounding the choice. During the training phase, participants must choose which stimulus they feel is correct, without the aid of feedback. Measures of reward and punishment learning are taken from the test phase, hypothesized to reflect the operations of a slow, probabilistic integrative system during training.

To test whether the participants learned more from seeking reward (Go learning) or avoiding punishment (NoGo learning), a testing phase followed the training phase. During the testing phase all possible stimulus pairs (i.e. AD, CF, etc.) were presented eight times (120 trials total). Go learning was defined as the accuracy of choosing A over C, D, E and F (i.e. seeking A), whereas NoGo learning was defined as the accuracy of choosing C, D, E and F over B (i.e. avoiding B). These test phase accuracies are suggested to reflect the slow probabilistic cortico-striatal integration that occurred during the training phase, where effective estimation of stimulus-action value is particularly needed to resolve the difference between stimuli with similar reinforcement values. Accordingly, high conflict trials were defined based on the reinforcement value difference between stimuli (a difficult choice relates to a small difference in reinforcement value) for high conflict Go (AC, AE and CE) and high conflict NoGo (BD, BF and DF), similar to previous research (Frank *et al.*, 2005, 2007b).

Stress manipulation

The stress manipulation was initiated between the T1 (benign) and T2 (stress) tasks. The social stress manipulation was designed to follow the criterion of Dickerson and Kemeny (2004) to create a socially evaluative environment with overt displays of exposed failure while maximizing motivated performance. Similar to the Trier Social Stress Task, this situation was uncontrollable in that the participant could not alleviate the evaluative tone. Probabilistic tasks are often opaque and non-memorizable, facilitating the environment of uncontrollability. Following the T1 (benign) task, participants were informed that they were going to 'need to do this task again' and that 'it is important that they try harder'. The assistant set up a video camera on top of the monitor, and the participants were told that the experimenters would be directly monitoring their performance this time. Both experimenters directly watched the participant and the video feed during the T2 (stress) task (Figure 2), the lead experimenter additionally made a check mark every

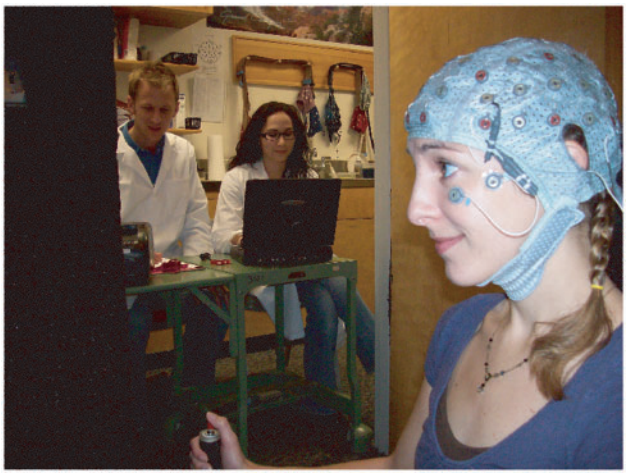

Fig. 2 Depiction of the social evaluative threat stress manipulation during the second performance of the task (T2).

time the participant received incorrect feedback. Additional standardized prompts were used to pressure the participant: 'Please keep your attention on the monitor', 'Just try to keep up with the pace', etc.

Abstract computational modeling of performance data

The trial-by-trial sequence of choices for each subject, for each task (T1 or T2) was fit by a Q-learning reinforcement learning model (Watkins, 1992; Sutton and Barto, 1998). As in Frank *et al.* (2007c), this model includes separate learning rate parameters for gain and loss (correct and incorrect feedback trials) in the training phase of the task. These separate gain/loss learning rates (αG/αL) scaled the updating of the stimulus-action values separately for rewards and punishments. The expected value (Q) of any stimulus (i) at time (t) was computed after each re-inforcer ($R = 1$ for Correct, $R = 0$ for Incorrect): where αG and αL were learning rates from gains and losses, respectively, and which were multiplied by prediction errors to update Q-values. These Q-values were entered into a softmax logistic function to produce probabilities of responses for each trial. These probabilities were then used to compute the log likelihood estimate of the subject having chosen that set of responses. The best-fitting parameters were found with a standard hill-climbing search algorithm (simplex method) by maximizing this log likelihood estimate.

It has been hypothesized that PFC and striatal systems learn separate estimates of state-action values in parallel, with different effective learning rates (Frank *et al.*, 2007c). To address the issue of the hypothesized cortico-striatal system which slowly integrates reinforcement information during training, but expresses these learned probabilities during the test phase, a parallel Q learning model was used to estimate end-of-training Q-values that correspond to test phase choices [see Supplementary Data and Frank *et al.* (2007c) for details]. Hereafter, these different types of learning rates are referred to by the phase they modeled to differentiate the putative functions of the rapidly adaptive system of the PFC reflected in 'fit-to-train' learning rates and the slow implicit cortico-striatal integrating system reflected in 'fit-to-test' learning rates.

Electrophysiological recording and processing

Scalp voltage was measured using 64 Ag/AgCl electrodes referenced to a site immediately posterior to Cz using a Synamps^2 system (bandpass filter 0.5–100 Hz, 500 Hz sampling rate, impedances <10 kΩ). Eyeblinks were removed with independent components analysis (Delorme and Makeig, 2004). Epochs were then transformed to Current Source Density (CSD) using the methods of Kayser and Tenke (2006). CSD computes the second spatial derivative of voltage between nearby electrode sites, acting as a reference-free spatial filter. Time-frequency calculations were computed using custom-written Matlab routines (Cohen *et al.*, 2008; Cavanagh *et al.*, 2009). The CSD-EEG time series in each epoch was convolved with a set of complex Morlet wavelets, defined as a Gaussian-windowed complex sine wave: $e^{-i2\pi tf}e^{-t^2/(2\times\sigma^2)}$. Power was normalized by decibel (dB) conversion. Epochs were baseline corrected for each frequency by the average power from −300 to −200 ms prior to the onset of the time-locking event. Values for statistical analysis were averaged over time and frequency in windows defined by the grand average wavelet plots in Figure 4 (over the broadly ranged theta band of 3–9 Hz: 0–100 ms for response locked, 250–450 ms for feedback locked). The electrocardiogram (ECG) was recorded from a bipolar clavicle lead with the Synamps^2 system and processed using QRSTool and CMET software (Allen *et al.*, 2007). Reactivity scores were computed as the difference between the average heart rate during the stress and the baseline tasks, computed separately for training and test phases of the tasks (T2–T1 for training, T2–T1 for testing), to highlight stress-specific heart rate change during task performance.

Reinforcement terminology

Hereafter, the terms reward and punishment are used as general terms for phenomena which either increase the probability of behavior (reward) or decrease the probability of behavior (punishment). While more precise and historically accurate terms would be positive reinforcement or positive punishment, the field of 'reinforcement learning' is construed to reflect both reinforcers (rewards) and punishers. Thus, reward learning was measured by the phenomena involved in processing feedback indicating a response was 'correct', test phase 'Go' accuracy, and learning rates for 'gain'. Punishment learning was measured by phenomena involved in processing feedback indicating a response was 'incorrect', test phase 'NoGo' accuracy and learning rates for 'loss'.

Statistical analysis

Data from each task were only analyzed if participants selected the most rewarding stimulus (A) over the most punishing stimulus (B) >50% of the time during the testing phase, since data from participants who fail this basic criterion are not interpretable (Frank *et al.*, 2005, 2007a; Gründler *et al.*, 2009; Cavanagh *et al.*, 2010). This criterion removed 8 participants from T1 analyses, 5 participants from T2 analysis and 12 participants from joint T1 and T2 analyses. For simplicity, Low and High BIS groups were created from a median split (median and mean rounded BIS score was 19). General Linear Models (GLMs) were used to examine change over time (T2–T1) of high conflict accuracies and learning rates. To investigate the role of state-dependent affect, negative affect was used as a continuous moderator in ANOVAs for the T2 period specifically. For simplicity, bivariate correlations were displayed for significant relationships between variables and Fisher's *z*-tests were used to test for significant differences between correlations. Multiple regression was used to predict test-phase punishment learning accuracy based on frontal theta during training (values were centered and multiplied to test for interactive effects), additionally, BIS group differences in this brain–behavior relationship were tested.

RESULTS

Demographics and performance

Table 1 presents performance means and s.ds on each task. There were no between-BIS group differences in T1 or T2 measures of performance (all *t*-values <2), with the exception of T1 fit-to-test learning rate for gain, where the low BIS group had a higher learning rate [low BIS $M = 0.55$; high BIS $M = 0.29$; $t_{(42)} = 2.18$, $P < 0.05$]. Instead, stress-related effects relied on the moderating influence of within-subject variables, as detailed below.

Table 1 Task performance and learning rates (means and s.d.) for T1 and T2 conditions

	Task	
Measure	T1 (Benign)	T2 (Stress)
Training blocks	3.36 (1.82)	4
Training RT (ms)	1083 (310)	1219 (380)
Test RT (ms)	1167 (420)	1517 (510)
Train accuracy (%)	68 (10)	72 (12)
Test accuracy (%)	68 (10)	74 (10)
Test Go accuracy (%)	68 (23)	75 (22)
Test NoGo accuracy (%)	69 (23)	73 (21)
Test Hi Conflict Go accuracy (%)	56 (23)	66 (20)
Test Hi Conflict NoGo accuracy (%)	60 (18)	62 (23)
Training loss learning rate	0.16 (0.22)	0.14 (0.24)
Training gain learning rate	0.29 (0.31)	0.31 (0.30)
Test loss learning rate	0.30 (0.38)	0.27 (0.32)
Test gain learning rate	0.43 (0.40)	0.44 (0.36)

Combined psychobiological stress response

There was no difference in the mean heart rate reactivity in the stress *vs* benign condition for training or test phases. There were no between-BIS group differences in mean heart rate reactivity or negative affect due to the stress manipulation, yet significant variance in test phase heart rate reactivity was accounted for in a regression by the interaction of BIS group and negative affect ($\beta = 0.26$, $t = 3.19$, $P < 0.01$). Heart rate reactivity significantly correlated with negative affect in the High BIS group only [$r_{(20)} = 0.61$, $P < 0.01$], which was significantly different than the null relationship in the Low BIS group [$r_{(27)} = -0.13$, $P > 0.50$; Fisher's $z = 2.63$, $P < 0.01$]. This relationship between state-dependent emotionality and physiological reactivity demonstrates a combined psychobiological stress response in the high BIS group.

Stress effects on reinforcement learning

A 2 (valence: reward, punishment) * 2 (group: Low BIS, High BIS) GLM for the difference measure (T2–T1) of high-conflict accuracy revealed a significant interaction [$F_{(1,36)} = 4.06$, $P < 0.05$] with a significant simple contrast for the Low BIS group ($P < 0.05$), absent any main effects. Figure 3A shows how the Low BIS group was characterized by both change towards higher high-conflict Go accuracy and lower high-conflict NoGo accuracy under stress; a non-significant trend in the opposite direction was found in the High BIS group. A similar 2 (valence: reward, punishment) * 2 (group: Low BIS, High BIS) GLM for the difference measure (T2–T1) of fit-to-test learning rates revealed a significant interaction [$F_{(1,36)} = 4.72$, $P < 0.05$] and no main effects. Decomposing the interaction found no significant simple contrasts, but the pattern depicted in the interaction is shown in Figure 3A, which shows how the Low BIS group was characterized by both change towards lower gain and higher loss learning rates under stress, the opposite pattern was found in the High BIS group. This pattern is consistent with the idea that lower learning rates are required to integrate probabilities over trials and to successfully discriminate between subtle differences in probability, as required in high conflict trials. Similar GLMs for high conflict reaction times, low conflict accuracies and fit-to-train learning rates were non-significant (F's < 1), demonstrating the specificity of these findings.

In summary, under stress, the Low BIS group was characterized by higher accuracy on reward trials and lower accuracy on punishment trials, the opposite pattern of results was found in the High BIS group (although only significant in learning rate). These patterns were specific to difficult valenced choices (high conflict, requiring a slow learning rate), fitting with the hypothesized long-term integrative process of the cortico-striatal system.

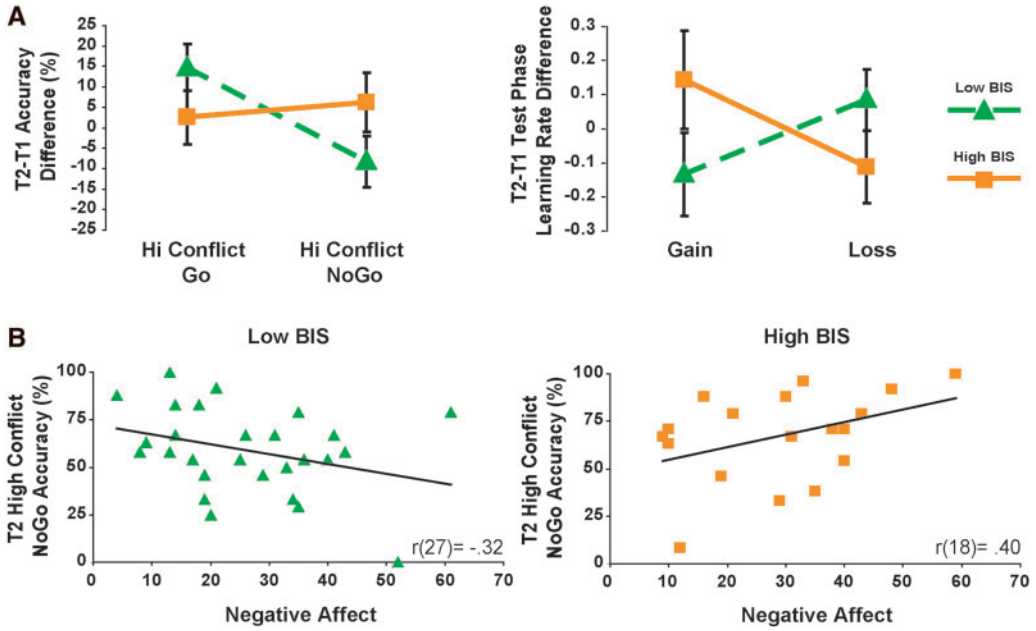

Fig. 3 Performance and learning rate changes due to stress reactivity. (**A**) Under stress, the Low BIS group was characterized by the tendency for more accurate reward learning and less accurate punishment learning; this pattern trended towards the opposite direction High BIS group. These same patterns are reflected more strongly in test phase learning rate changes, where lower learning rates are hypothesized to reflect more effective slow probabilistic integration. (**B**) The degree of negative affect during stress also differentially altered punishment learning: relating to poorer punishment learning in the Low BIS group yet better punishment learning in the High BIS group.

Negative affect moderates stress effects on reinforcement learning

Under stress conditions (T2) specifically, similar 2 (valence: reward, punishment) * 2 (group: Low BIS, High BIS) GLMs were run with negative affect as a continuous moderator to investigate the interactive role of trait and state stress vulnerability. A three-way interaction was found for high-conflict accuracy [$F_{(1,41)} = 5.01$, $P < 0.05$], yet the three-way interaction for fit-to-test learning rate was only a trend [$F_{(1,41)} = 2.86$, $P = 0.10$]. A follow-up ANOVA for accuracy was split by valence (see Supplementary Data for learning rate analyses). No effects were found for high conflict Go accuracy, whereas there was a significant trait (BIS group) * state (negative affect) interaction for high-conflict NoGo accuracy [$F_{(1,41)} = 6.02$, $P < 0.05$], absent any significant main effects. Figure 3B shows how increasing negative affect was related to poorer high-conflict NoGo accuracy in the Low BIS group [$r_{(27)} = -0.32$, $P = 0.10$] but better high-conflict NoGo accuracy in the High BIS group [$r_{(18)} = 0.40$, $P = 0.10$; Fisher's $z = -2.29$, $P < 0.05$]. In summary, increasing negative affect moderated the diminishment in punishment accuracy in the Low BIS group yet the enhancement of punishment accuracy in the High BIS group.

Negative affect facilitates internalization of punishment

To identify neural systems hypothesized to be altered by stress reactivity, theta-band power metrics locked to the response and negative feedback during the training phase (Figure 4) were investigated within the context of trait and state stress reactivity. Negative affect was included as a continuous moderator in a 2 (locus: response, negative feedback) * 2 (group: Low BIS, High BIS) GLM for theta power. There was a significant negative affect * locus interaction, $F_{(1,41)} = 4.64$, $P < 0.05$, absent any other interactive or main effects. Figure 5 demonstrates how increasing negative affect was related to increasing response-locked theta power [$r_{(45)} = 0.32$, $P < 0.05$], yet decreasing negative feedback-locked theta power [$r_{(45)} = -0.21$, $P = 0.17$; Fisher's $z = 2.5$, $P < 0.01$], demonstrating greater state-dependent internalization of value across all participants (no group effects). Importantly, negative affect did not correlate with T2 accuracy during the training phase ($P > 0.47$).

Internalization of punishment differentially predicts punishment learning

Multiple regression was used to investigate if T2 high conflict NoGo accuracy was predicted by response-locked and feedback-locked theta power (here, the difference score between feedbacks was used: incorrect-correct), and if this relationship differed between groups. The three-way interaction was significant ($\beta = 0.20$, $t = 2.03$, $P < 0.05$). Separate between-group regressions for each theta measure were non-significant, yet within-group regressions were revealing. In the Low BIS group, feedback-locked theta power alone predicted punishment learning (main effect: $\beta = 0.06$, $t = 2.11$, $P < 0.05$), the other main and interaction effects

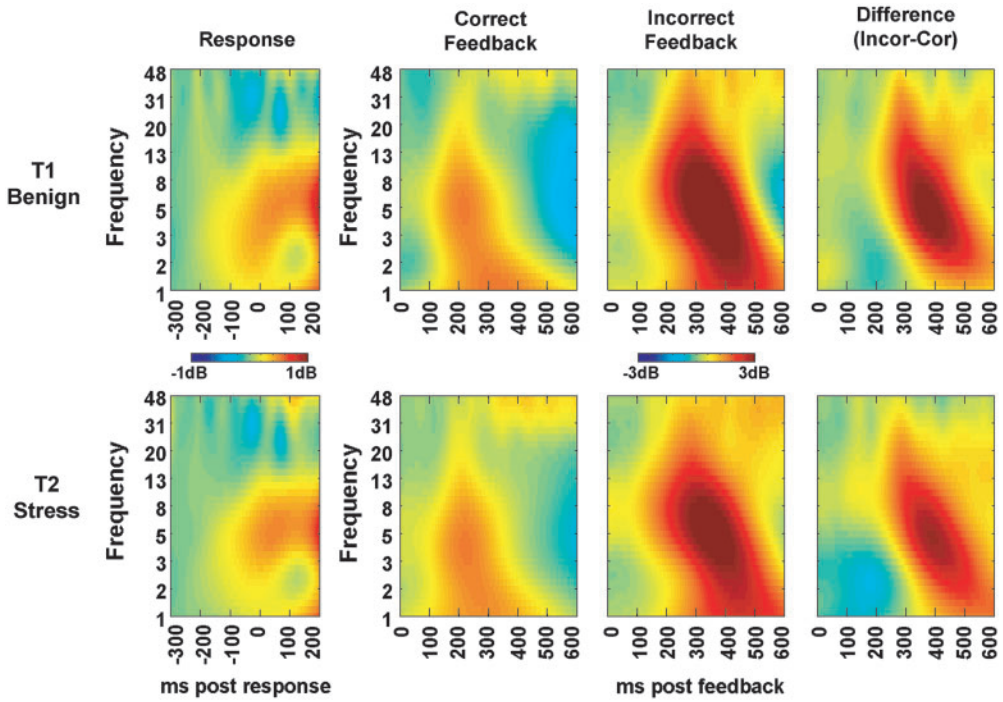

Fig. 4 Time-frequency representations of EEG power at the FCz electrode. EEG plots are shown for response- and feedback-locked events during training (T1 and T2).

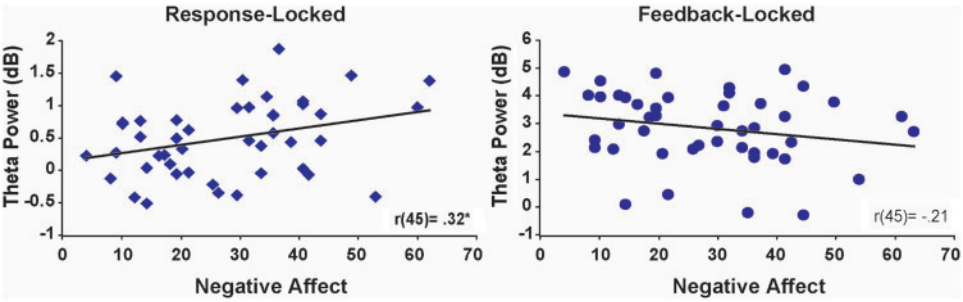

Fig. 5 Increasing negative affect was related to a general pattern of greater internalization of punishment during training, as reflected by an increase in response-locked theta power and a concurrent decrease in negative feedback-locked theta power.

were non-significant (t's < 1). In the High BIS group, there was a significant interaction ($\beta = 0.21$, $t = 2.50$, $P < 0.05$) where greater interactive feedback- and response-locked theta power predicted punishment learning, absent any main effects (t's < 1). Figure 6 summarizes this interaction with median split values (± s.e.). In the Low BIS group, only theta power to negative feedback predicted punishment learning. In the High BIS group, an inverted-U effect of interactive response and feedback processing predicted punishment learning, where high total theta activities or low total theta activities predicted better learning.

DISCUSSION

This investigation revealed that trait level punishment sensitivity and state-related negative affect moderate the ability to learn to seek reward and avoid punishment during social stress. Low trait-level punishment sensitivity was related to a tendency towards better reward learning and poorer punishment learning; the opposite pattern was found in highly punishment sensitive individuals (at least in learning rate). These inverted-U effects were further bolstered by the finding that negative affect was inversely related to effective punishment learning in low punishment sensitive individuals, but these measures were directly related in more sensitive individuals. These reinforcement-related learning alterations were specific to high conflict choices and fit-to-test learning rates, suggesting an alteration in the slow integrative process of the cortico-striatal system.

Cortical bias of reinforcement

One candidate mechanism by which individual differences in stress reactivity affect reinforcement learning may be due to

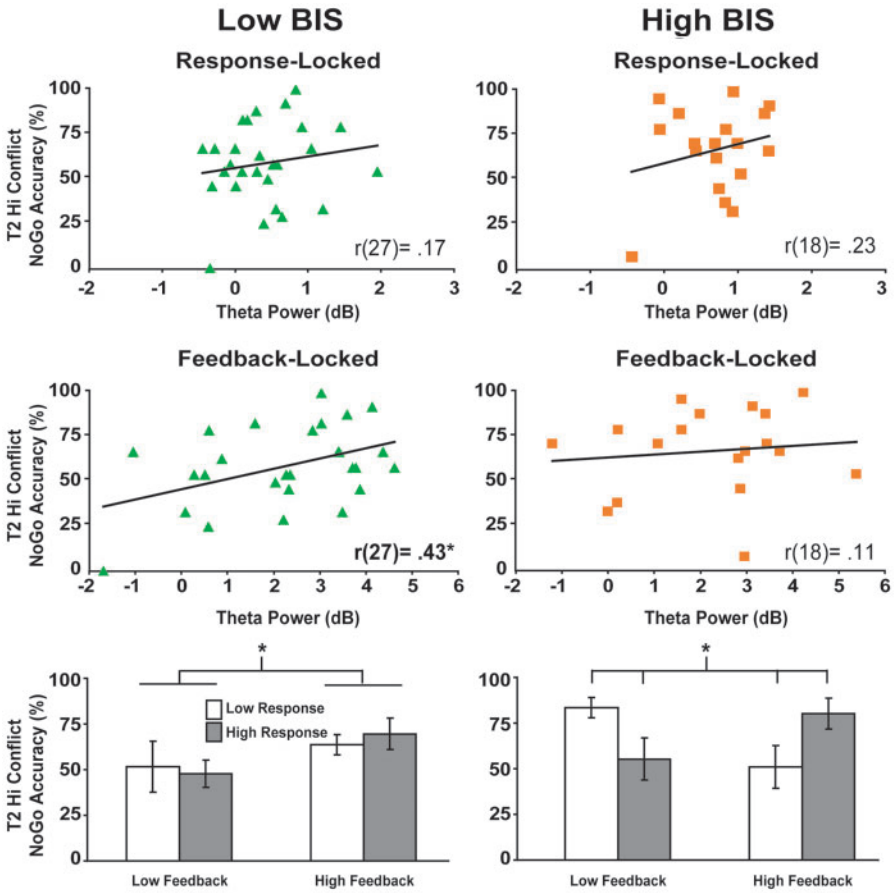

Fig. 6 Response and negative-feedback locked theta in relation to performance (values for interaction plots are median splits ±SE). In the Low BIS group (left column), theta power to feedback alone predicted subsequent high conflict NoGo learning (asterisk indicates main effect $p < .05$). In the High BIS group (right column), an inverted-U type interaction between response- and feedback-locked theta predicted subsequent accuracy in high conflict NoGo learning, where high total theta activities or low total theta activities predicted better learning (asterisk indicates interaction effect $p < .05$).

altered internalization of error, conflict and punishment in cortico-striatal action monitoring systems. Mediofrontal EEG theta power demonstrated a negative affect-dependent shift towards diminished processing of external punishment cues and heightened internal processing of error and conflict during training. While this shift was present in all participants regardless of trait-level punishment sensitivity, only feedback-locked theta power predicted punishment learning in the Low BIS group; the interaction of these theta metrics predicted punishment learning in the High BIS group.

In the Low BIS group, increasing negative affect may have led to poorer punishment learning due to diminished utilization of external punishment feedback. In the High BIS group, increased sensitivity to internal indicators of performance in conjunction with continued attention to external punishment cues led to better punishment learning. Curiously, the exact opposite pattern of theta response (suggesting both lower internal and external salience) was also shown to predict better punishment learning in High BIS participants, possibly indicating that stress-related PFC activities at both ends of the inverted-U facilitate integration of action values.

Reinforcement and sub-cortical activity

An acute stress-related increase in mesolimbic dopamine (DA) tone has been previously described in human neuroimaging experiments (Pruessner *et al.*, 2004; Soliman *et al.*, 2008). Here, an enhanced reward learning bias in the Low BIS group and decreased reward learning bias in the High BIS group could also be related to tonic dopaminergic mechanisms. A computational model of cortico-striatal functioning during this same learning task suggests that phasic DA bursts to reward are heightened by increased DA tone, increasing 'Go' learning via D1 receptor activities and facilitating motor execution to seek rewards (Frank *et al.*, 2004). Conversely, phasic DA dips to punishment are exacerbated by decreased DA tone, supporting 'NoGo' learning via D2 autoreceptor activity and inhibiting motor execution for punishment avoidance.

Increased DA tone is suggested to reflect coping with a controllable stressor, whereas decreased tonic DA is suggested to reflect reaction to an uncontrollable stressor (Cabib and Puglisi-Allegra, 1994). Critically, both the appraisal of controllability and subsequent regulation of DA tone may be determined by mPFC (Amat *et al.*, 2005;

Amat *et al.*, 2006; Pascucci *et al.*, 2007). The ACC has been shown to appraise environmental uncertainty and subsequently alter learning rates (Behrens *et al.*, 2007), and the PFC may even directly bias subcortical functioning to enhance belief-specific learning (Doll *et al.*, 2009; Huys and Dayan, 2009). Here, an appraisal of stressor uncontrollability or a general expectation of failure in highly punishment sensitive participants may have lead to the bias towards punishment learning (and away from reward learning), partially mediated by a diminishment of mesolimbic DA tone and/or effective learning rate.

Limitations and future directions

Manipulations of stress controllability are difficult to manage with human participants, especially when the task requires veritable interaction. Here, we regretfully did not measure an appraisal of perceived controllability. Future investigations could gather appraisals of stress controllability as well as manipulate the veracity or intensity of task feedback. While we did demonstrate an important role of emotionality in mediofrontal systems during reinforcement learning, some aspects of hypothesized striatal and dopaminergic functions remain difficult or impossible to assess with human neuroimaging. Social, cognitive and behavioral neuroscientists should continue to strive towards hypotheses that are able to be cross-validated and translated between methodologies and species.

SUMMARY

An integrative explanation of the findings and possible mechanisms here revolves around the fact that the mPFC is intimately involved in appraising stressor controllability and environmental uncertainty, as well as adapting behavior to reinforcement. The combined activities of this particular cortico-striatal system identify it as a focal node by which stress may be internalized to affect cognitive, emotional and behavioral functioning. Dysfunctional stress reactivity may be a risk factor for prolonged affective distress, yet mechanisms underlying this process remain under-explained. Stress-related alteration of reward and punishment learning systems—particularly the ACC—is a viable candidate for how dysfunctional stress reactive responses are translated into ongoing cognitive and affective distress in mental illness and addiction.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Allen, J.J.B., Chambers, A.S., Towers, D.N. (2007). The many metrics of cardiac chronotropy: a pragmatic primer and a brief comparison of metrics. *Biological Psychology*, *74*, 243–62.

Amat, J., Paul, E., Zarza, C., Watkins, L.R., Maier, S.F. (2006). Previous experience with behavioral control over stress blocks the behavioral and dorsal raphe nucleus activating effects of later uncontrollable stress: role of the ventral medial prefrontal cortex. *Journal of Neuroscience*, *26*, 13264–72.

Amat, J., Baratta, M.V., Paul, E., Bland, S.T., Watkins, L.R., Maier, S.F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience*, *8*, 365–71.

Arnsten, A. (1998). Catecholamine modulation of prefrontal cortical cognitive function. *Trends in Cognitive Sciences*, *2*, 11.

Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–21.

Cabib, S., Puglisi-Allegra, S. (1994). Opposite responses of mesolimbic dopamine system to controllable and uncontrollable aversive experiences. *Journal of Neuroscience*, *14*, 3333–40.

Carver, C.S., White, T.L. (1994). Behavioral-inhibition, behavioral activation, and affective responses to impending reward and punishment-the Bis Bas scales. *Journal of Personality and Social Psychology*, *67*, 319–33.

Cavanagh, J.F., Allen, J.J.B. (2008). Multiple aspects of the stress response under social evaluative threat: an electrophysiological investigation. *Psychoneuroendocrinology*, *33*, 41–53.

Cavanagh, J.F., Allen, J.J.B. (2009). The Behavioural Inhibition System. In: Sander, D., Scherer, K.R., editors. *The Oxford Companion to Emotion and the Affective Sciences*. New York: Oxford University Press, pp. 73–4.

Cavanagh, J.F., Cohen, M.X., Allen, J.J.B. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *Journal of Neuroscience*, *29*, 98–105.

Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J.B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, *49*, 3198–209.

Cohen, M.X., Ridderinkhof, K.R., Haupt, S., Elger, C.E., Fell, J. (2008). Medial frontal cortex and response conflict: evidence from human intracranial EEG and medial frontal cortex lesion. *Brain Research*, *1238*, 127–42.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D.Y., Engel, A.K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*, 11730–7.

Dedovic, K., D'Aguiar, C., Pruessner, J.C. (2009). What stress does to your brain: a review of neuroimaging studies. *Canadian Journal of Psychiatry*, *54*, 6–15.

Delorme, A, Makeig, S (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Dickerson, S.S., Kemeny, M.E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355–91.

Diorio, D., Viau, V., Meaney, M.J. (1993). The role of the medial prefrontal cortex (cingulate gyrus) in the regulation of hypothalamic-pituitary-adrenal responses to stress. *Journal of Neuroscience*, *13*, 3839–47.

Doll, B.B., Jacobs, W.J., Sanfey, A.G., Frank, M.J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Research*, *1299*, 74–94.

Frank, M.J., Seeberger, L.C., O'Reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*, 1940–3.

Frank, M.J., Woroch, B.S., Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron*, *47*, 495–501.

Frank, M.J., D'Lauro, C., Curran, T. (2007a). Cross-task individual differences in error processing: neural, electrophysiological, and genetic components. *Cognitive Affect Behavior Neuroscience*, *7*, 297–308.

Frank, M.J., Santamaria, A., O'Reilly, R.C., Willcutt, E. (2007b). Testing computational models of dopamine and noradrenaline dysfunction in

attention deficit/hyperactivity disorder. *Neuropsychopharmacology*, *32*, 1583–99.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., Hutchison, K.E. (2007c). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of Natational Academy of Sciences of the United States of America*, *104*, 16311–6.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E. (1993). A neural system for error-detection and compensation. *Psychological Science*, *4*, 385–90.

Gross, J.J., Levenson, R.W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*, 87–108.

Gründler, T.O.J., Cavanagh, J.F., Figueroa, C.M., Frank, M.J., Allen, J.J.B. (2009). Task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia*, *47*, 1978–87.

Holroyd, C.B., Coles, M.G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychology Review*, *109*, 679–709.

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*, 497–8.

Huys, Q.J., Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*, 314–28.

Kayser, J., Tenke, C.E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, *117*, 348–68.

Kemeny, M. (2003). The psychobiology of stress. *Current Directions in Psychological Science*, *12*, 4.

Krugel, L.K., Biele, G., Mohr, P.N.C., Li, S.C., Heekeren, H.R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 17951–6.

Lerner, J.S., Dahl, R.E., Hariri, A.R., Taylor, S.E. (2007). Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. *Biological Psychiatry*, *61*, 253–60.

Mars, R.B., Coles, M.G., Grol, M.J., et al. (2005). Neural dynamics of error processing in medial frontal cortex. *Neuroimage*, *28*, 1007–13.

Pascucci, T., Ventura, R., Latagliata, E.C., Cabib, S., Puglisi-Allegra, S. (2007). The medial prefrontal cortex determines the accumbens dopamine response to stress through the opposing influences of norepinephrine and dopamine. *Cerebral Cortex*, *17*, 2796–804.

Pruessner, J.C., Champagne, F., Meaney, M.J., Dagher, A. (2004). Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: a positron emission tomography study using [11C]raclopride. *Journal of Neuroscience*, *24*, 2825–31.

Radley, J.J., Gosselink, K.L., Sawchenko, P.E. (2009). A discrete GABAergic relay mediates medial prefrontal cortical inhibition of the neuroendocrine stress response. *Journal of Neuroscience*, *29*, 7330–40.

Soliman, A., O'Driscoll, G.A., Pruessner, J., et al. (2008). Stress-induced dopamine release in humans at risk of psychosis: a [11C]raclopride PET study. *Neuropsychopharmacology*, *33*, 2033–41.

Sutton, R.S., Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, Mass: MIT Press.

Trujillo, L.T., Allen, J.J. (2007). Theta EEG dynamics of the error-related negativity. *Clinical Neurophysiology*, *118*, 645–68.

Watkins, C.J.C.H.D, P. (1992). Technical Note: Q-Learning. *Machine Learning*, *8*, 279.