NeuroImage 229 (2021) 117747

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Stress reduces both model-based and model-free neural computations during flexible learning

Anna Cremer[a], Felix Kalbe[a], Jan Gläscher[b,#], Lars Schwabe[a,#,*]

[a] *Department of Cognitive Psychology, Institute of Psychology, Universität Hamburg, Hamburg 20146, Germany*
[b] *Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany*

ARTICLE INFO

Keywords:
Stress
Cognitive flexibility
Decision-making
Reinforcement learning

ABSTRACT

Stressful events are thought to impair the flexible adaptation to changing environments, yet the underlying mechanisms are largely unknown. Here, we combined computational modeling and functional magnetic resonance imaging (fMRI) to elucidate the neurocomputational mechanisms underlying stress-induced deficits in flexible learning. Healthy participants underwent a stress or control manipulation before they completed, in the MRI scanner, a Markov decision task, frequently used to dissociate model-based and model-free contributions to choice, with repeated reversals of reward contingencies. Our results showed that stress attenuated the behavioral sensitivity to reversals in reward contingencies. Computational modeling further indicated that stress specifically affected the use of value computations for subsequent action selection. This reduced application of learned information on subsequent behavior was paralleled by a stress-induced reduction in inferolateral prefrontal cortex activity during model-free computations. For model-based learning, stress decreased specifically posterior, but not anterior, hippocampal activity, pointing to a functional segregation of model-based processing and its modulation by stress along the hippocampal longitudinal axis. Our findings shed light on the mechanisms underlying deficits in flexible learning under stress and indicate that, in highly dynamic environments, stress may hamper both model-based and model-free contributions to adaptive behavior.

Stressful events are a powerful modulator of learning and memory (Diamond et al., 2007; Luksys and Sandi, 2011; Lupien et al., 2009; Roozendaal et al., 2009; Schwabe et al., 2012). In particular, stress is thought to render learning and memory rather rigid, thus impairing the flexible adaptation to changing environments (Raio et al., 2017; Schwabe and Wolf, 2013; Wirz et al., 2018; Schwabe et al., 2013). Although such deficits in flexible learning under stress have far-reaching implications, not only for educational and clinical contexts (de Quervain et al., 2017; Goldfarb and Sinha, 2018; Goodman et al., 2012; Vogel and Schwabe, 2016), the exact mechanisms underlying the stress-induced impairments in flexible learning are still largely unclear.

Successful adaptation to dynamic environments depends on the complex interplay of at least two systems: (i) a reflective or goal-directed system that involves the consideration of prospective future courses of action and their consequences and (ii) a reflexive or habitual system that is guided by the retrospective experience of good and bad outcomes (Balleine and O'doherty, 2010; Sloman, 1996). Accumulating evidence from human and rodent studies shows that stress and stress hormones may bias the balance of these systems and favor habitual over goal-directed behavior (Braun and Hauber, 2013; Gourley et al., 2012; Schwabe et al., 2012; Schwabe and Wolf, 2009, 2011; Schwabe et al., 2012b). Computationally, goal-directed and habitual forms of behavioral control are assumed to overlap to some degree with model-based and model-free reinforcement learning systems (Dolan and Dayan, 2013). Within this framework, learning can be defined as the identification of a value function that selects the most rewarding options in the current environment. Therefore, the value function links the previous value of the options available with rewards that can be expected in the future. This results in a policy that maps different environments to action probabilities and therefore determines which actions are selected in each state (Gershman and Uchida, 2019). Specifically, a central aspect in both model-based and model-free learning is the computation of prediction error signals to update the value function. Therefore, previous experiences are used to form predictions, which are then updated by comparing the predicted outcome of an option to the actual outcome.

While a model-based policy acquires a cognitive map of the task structure (i.e., how different environments are linked to each other) and uses this to predict the most advantageous course of action, the model-free system encodes values by trial and error and uses the reward history to guide behavior (Daw et al., 2005, 2011; Gläscher et al., 2010).

[*] Corresponding author.
E-mail address: Lars.Schwabe@uni-hamburg.de (L. Schwabe).
[#] these authors contributed equally

https://doi.org/10.1016/j.neuroimage.2021.117747
Received 27 August 2020; Received in revised form 4 January 2021; Accepted 9 January 2021
Available online 14 January 2021

On the neural level, model-based processing is thought to rely on posterior inferior parietal as well as lateral prefrontal regions (Gläscher et al., 2010) and, as shown more recently, on hippocampal areas (Pfeiffer and Foster 2013; Garvert et al., 2017; Miller et al., 2017; Stachenfeld et al., 2017). Model-free learning, in turn, is assumed to be driven by prediction error signals of midbrain dopamine neurons mapping the difference between the actual and expected reward at a particular state and depends mainly on the ventral striatum (Bayer and Glimcher, 2005; Haruno and Kawato, 2006; McClure et al., 2003; O'Doherty et al., 2003). In terms of the flexible adaption to changes in the environment, we identified the medial prefrontal cortex (mPFC) as a potential key player, since it is linked to essential features of flexible learning (Nee et al., 2011). In particular, the mPFC is thought to be implicated in the anticipation of values of currently available actions (Aarts et al., 2008), the representation of possible outcomes (Brown, 2009), the association between actions and outcomes (Oliveira et al., 2007), error detection processes during contingency changes (Zarr and Brown, 2016), and the computation of likely action outcomes (Alexander and Brown, 2011; Croxson et al., 2009).

The computational conceptualization of reflexive and reflective systems of behavioral control in terms of model-based and model-free processing provided valuable insights into the mechanisms underlying each of these systems as well as their interplay. First behavioral studies suggested that acute stress may affect behavioral flexibility in general (Plessow et al., 2011; Schwabe and Wolf, 2011) and the contributions of model-based and model-free processes to aversive learning or learning from negative outcomes in particular (Park et al., 2017; Raio et al., 2017). However, how stress changes the contributions of model-based and model-free systems to flexible learning in a highly volatile environment and, in particular, the neural mechanisms underlying stress-induced alterations in model-based and model-free processing are largely unknown.

In the present experiment, we combined computational modeling and functional magnetic resonance imaging (fMRI) to elucidate the neurocomputational mechanisms underlying stress-induced deficits in flexible learning. Therefore, healthy participants first underwent a standardized stress or control procedure before they completed a two-step Markov decision task in the MRI scanner. This task allows a dissociation of model-based and model-free contributions to behavior (Daw et al., 2011) and requires two subsequent decisions which can ultimately lead to a reward. To explicitly probe the flexibility of learning, we used a modified version of this task that included repeated reversals of reward contingencies. Here, flexible learning was expressed as the ability to detect a reversal and adapt the choice behavior accordingly. We assumed that task performance would rely on both model-based and model-free computations and that stress would reduce their recruitment during learning. Because previous findings suggested that individuals with low working memory capacity were more susceptible to detrimental stress effects on model-based learning strategies than participants with high working memory capacity (Otto et al., 2013), we further included an n-back test to probe participants' baseline working memory performance.

1. Materials and methods

1.1. Participants and experimental design

Sixty-eight healthy volunteers participated in this experiment. Based on previous studies from our lab that reported effect sizes of Cohen's d from 0.66 to 0.98 for similar research questions (Schwabe and Wolf, 2009, 2012), we expected a medium to large effect of stress on flexible learning of Cohen's $d = 0.7$. A power analysis using G*power (Faul et al., 2007) indicated that using a two-tailed independent t-test with alpha = 0.05, a sample of 68 participants is required to detect such a medium-sized effect with a power of 0.80. All participants were right-handed, had normal or corrected-to-normal vision and were screened for possible MRI contraindications. Individuals with a current medical condition, current medication intake or lifetime history of any neurological or psychiatric disorders were excluded from participation. Moreover, we excluded smokers and women taking hormonal contraceptives as both can affect the stress response (Kirschbaum et al., 1999; Rohleder and Kirschbaum, 2006). Participants were asked not to drink coffee or other caffeinated beverages and not to do any exercise on the day of the experiment. In addition, they should not eat or drink anything except water 2 h before the appointment. All participants provided written informed consent before the beginning of testing and received a moderate monetary compensation. The study protocol was approved by the local ethics committee. Ten participants had to be excluded from the analysis because of excessive head movement (mean displacement > 5 mm) in the MRI ($n = 4$), because they missed more than 30% of the trials ($n = 3$) or because they chose the same action in more than 95% of the trials ($n = 3$), thus leaving a final sample of 58 participants (17 men and 12 women in each of the two groups, age 18–34, mean = 24.6, SD = 3.5, no age difference between groups, t(57) = 0.73, $p = 0.47$). Participants were pseudorandomly assigned to the stress and control groups, in order to achieve an identical number of men and women per group.

1.2. Stress induction

In order to control for the diurnal rhythm of the stress hormone cortisol, all testing took place in the afternoon and early evening, with the time of testing being counterbalanced across groups. Participants of the stress group underwent the Trier Social Stress Test (TSST; Kirschbaum et al., 1993), a standardized paradigm in experimental stress research that is known to activate both the autonomic nervous system and the hypothalamus-pituitary-adrenal axis. In brief, the TSST simulates a 15-min job interview, including a public speech about the participant's eligibility for a job tailored to his/her interests and a mental arithmetic task. During both tasks, participants were videotaped and evaluated by two rather cold, non-reinforcing committee members (1 man, 1 woman), dressed in white lab coats. In the control condition, participants spoke about a topic of their choice followed by a simple arithmetic task (counting forwards in steps of 15), without committee or video recordings.

To evaluate the successful stress induction through the TSST, subjective and physiological measurements were taken at several time points across the experiment (see Fig. 1). Baseline was assessed 10 min after the start of the appointment, so that the subjects were able to acclimatize to the situation. Directly after the TSST/control manipulation, participants rated the difficulty, stressfulness, and unpleasantness of the experimental treatment on a scale from 0 ("not at all difficult/stressful/unpleasant") to 100 ("very difficult/stressful/unpleasant"). Blood pressure and pulse were measured at baseline, during the TSST, directly after the TSST and after the fMRI scanning session using a digital blood pressure device (OMRON model M500 (HEM-7321-D); Healthcare Europe BV, Hoofddorp, The Netherlands) with a cuff applied around the right upper arm, when subjects were standing. Finally, we collected saliva samples using Salivette® collection devices (Sarstedt) at baseline, 18 min after stressor onset (shortly before the learning task started), and after each block of the Markov decision task (i.e., 40, 60 and 90 min after the treatment, Fig. 1). Saliva samples were stored at −18 °C until the end of data collection, when we analyzed saliva cortisol concentrations using a luminescence assay (IBL, Germany).

1.3. Markov decision task

Twenty minutes after the beginning of the stress/control manipulation, when stress-induced cortisol concentrations were expected to peak, participants performed a modified version of a two-step Markov decision task in the MRI scanner. This task was designed to dissociate between model-based and model-free learning mechanisms (Daw et al., 2011). Each trial consisted of two successive stages, in each of which

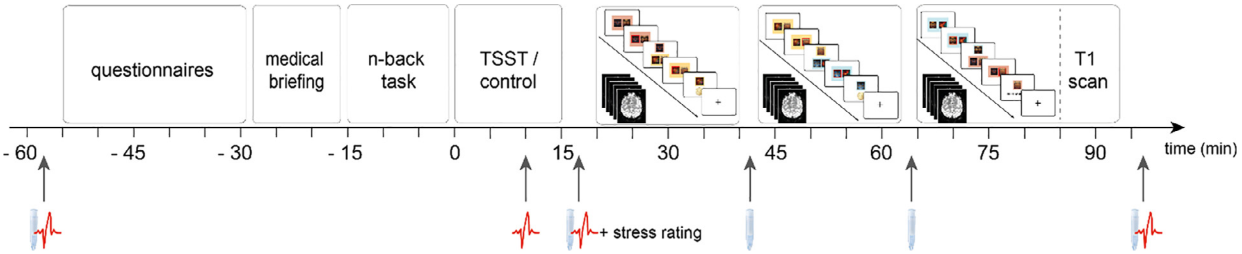
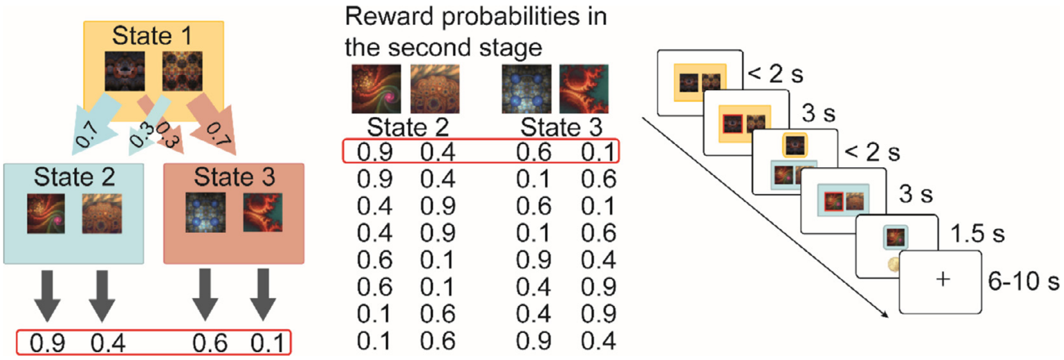


Fig. 1. Experimental procedure. Stress was induced by the Trier Social Stress Test (TSST). Before the stress/control treatment, participants completed several questionnaires and performed an n-back task. After the stress/control procedure, participants completed three blocks of a modified Markov decision task (MDT) in the MRI scanner. Stress reactivity was assessed by subjective and physiological measures (salivary cortisol, blood pressure, pulse), which were taken at several time points across the experiment.



Reward probabilities in the second stage

State 2		State 3	
0.9	0.4	0.6	0.1
0.9	0.4	0.1	0.6
0.4	0.9	0.6	0.1
0.4	0.9	0.1	0.6
0.6	0.1	0.9	0.4
0.6	0.1	0.4	0.9
0.1	0.6	0.4	0.9
0.1	0.6	0.9	0.4

Fig. 2. Experimental task. Left: State transition structure. Each first stage (state 1) action is predominantly associated with one or the other second stage states (state 2 and state 3), and leads there in 70% of the time. The different states were marked by differently colored boxes. Reward probabilities in the second stage undergo frequent reversals. Middle: Reversal patterns. Reward probabilities of 0.4 and 0.9, and 0.1 and 0.6 stay together in one state. Out of these possible combinations, six patterns (two per block) occurred over the course of the experiment per participant. Right: Timeline of events per trial. A first stage choice between two options leads to a second stage choice which is probabilistically reinforced with monetary rewards.

the participant had to choose between two options (“left” or right”), represented by fractal images (Fig. 2). The first stage decision (state 1) led to one of two possible states in the second stage (state 2 and state 3), requiring another choice between two fractals, which were associated with different probabilities of receiving a monetary reward. Each first stage option was predominantly (70%) associated with one or the other second stage state. The choice options in the second stage led to a monetary reward with a probability of 0.9, 0.6, 0.4 and 0.1, with the probabilities of 0.9 and 0.4 being paired in one state (“good state”), and 0.6 and 0.1 in the other (“bad state”). A win was depicted by a 10 cents coin, otherwise “no reward” appeared on the screen. Upon each state, participants had 2 s to submit their choice on an MRI compatible button box. If they failed to enter a choice within this time window, the trial was aborted and the next trial started. Trials were separated by an inter trial interval of randomized length, between 6 and 10 s (Fig. 2, right).

Whether the transition structure is included in the decision or not provides insights into the engaged learning strategy. While the model-free learner evaluates actions retrospectively by repeating previously rewarded choices, the model-based learner also takes the task structure into account. Consider a first stage choice that led to a second state via a rare transition, followed by a second state choice that led to a reward. A purely model-free agent would repeat the action because it was rewarded. A purely model-based learner, however, would switch to the other first stage option because it takes into account that the previous first stage action only leads to the rewarded second stage state via a rare transition. Thus, first stage decisions provide the opportunity to determine the extent to which model-based vs. model-free computations contribute to decisions. In order to explicitly test the flexibility of learning, we modified the original task by introducing repeated reversals of reward contingencies (Fig. 2), requiring the flexible adaptation of behavior.

Participants performed 202 trials, distributed over three blocks (70 / 66 / 66 trials) and separated by breaks. In order to explicitly test the flexibility of learning, defined as adaptation to a changing environment, we modified the original task by introducing repeated reversals of reward contingencies. Specifically, the reward probabilities associated with second stage actions were reversed twice per block, fixed at trial numbers 27, 49, 93, 115, 159 and 181. To ensure that reversals are detectable despite the probabilistic reward structure, the reversals only take place within one of the two second stage states. Fig. 2 (middle) shows all possible reversal combinations. Note that the three blocks were separated by a short break in which the subjects were briefly moved out of the scanner to collect saliva samples. The experimenter then placed the Salivette in the participant’s mouth using sterile plastic tweezers, paralleled by the instruction to move as little as possible. We applied the same criteria for the movement parameters between the blocks as during the task, i.e. excluding participants with a mean displacement > 5 mm. To make sure that participants did not continue to apply their previously learned contingencies, each block began with a new stimulus allocation. That is, the same six fractals were randomly assigned to the three states. Likewise, the background colors of the states were reassigned. The second stage reward probabilities were randomly attached to the new second stage stimuli. The assignment of colors and stimuli to the states was counterbalanced across participants. The stimulus pairs within the states stayed the same within one block, and so did the background colors of the states and the transitions between first and second stage. The location of the two options in one state was randomized from trial to trial to ensure that the participants learned stimulus – state and stimulus – reward contingencies rather than the stimulus position.

Participants were instructed that they had to make two decisions in a row in each trial, with the second decision possibly leading to a reward and that the aim of the task was to gain as many rewards as possible. They were told that the first decision was not directly associated with the reward, but that it leads to one of two possible states, in which they again have the choice between two pictures. They were also instructed that each first stage option was primarily associated with one or the other state in the second stage, but not with which one. Further, the instructions stated that within the two second stage states, each picture leads with a certain probability to a reward and in both states there is a slightly better and a slightly less favorable option and that they should find the best option. Most importantly, participants were informed that this option would change several times throughout the experiment and that they should detect the changes and adapt their behavior accordingly. Lastly, they learned that none of the images would lead to a reward in all trials, so that it is possible that they do not get a reward for an answer that has been correct many times before - but that this does not necessarily mean that a reversal took place. The exact task instructions that participants received are provided in the supplemental material.

Before the TSST or control manipulation, participants performed a brief training session for the learning task (out of the MRI scanner). The training consisted of three parts with 10 trials each, introducing the task iteratively. In each phase, the trial structure was the same as in the experimental task. In the first part, the participants should find out which image would lead to a reward with one of the four images being rewarded while the other three were not. The second part was identical, except that the rewarded picture changed at some point. The participants were informed about the reversal and were instructed to adapt their behavior. The third phase was identical to the experimental task. We used the same stimuli and transition structure as in the experiment, the image position and colors of the states were randomized. All three phases had a fixed number of trials, no learning criterion was applied.

1.4. Working memory assessment

Because previous research suggested that the influence of stress on the control of learning may be moderated by the individual working memory capacity (Otto et al., 2013), we measured working memory using an n-back task (Kirchner, 1958) before participants underwent the stress or control manipulation. Participants were presented a random sequence of one-digit numbers from “0” to “9” and asked to indicate via button press (“yes” or “no”) whether the currently presented number was the same as the one presented n-trials before. Participants received 10 stimulus blocks in total (2 practice blocks with feedback and 8 experimental blocks without feedback), in which working memory load varied by alternately using a 2-back and a 3-back condition. Each block consisted of 24 stimulus trials. Stimuli were displayed for 500 ms and responses were recorded within 1500 ms, followed by 2000 ms fixation cross.

1.5. Behavioral data analyses

To test whether the TSST successfully induced stress, data on subjective ratings, vital signs, and salivary cortisol were analyzed using mixed-design ANOVAs with the between-subjects factor treatment and the within-subjects factor time after stress/control manipulation onset. T-tests were used to investigate post-hoc group differences in these measures. Learning performance was quantified by the proportion of first stage choices for the option that led predominantly, with a probability of 0.7, to the second stage state with the overall higher probability to obtain a reward. Likewise, the proportion of choices for the option with the higher reward probability (either 0.9 or 0.6) in the second stage reflected successful learning. We further computed the sensitivity to detect changing contingencies as a difference index between the mean number of advantageous choices in the four trials before a reversal relative to the four trials after a reversal. We chose this number of trial before and after a reversal to ensure that the participants had enough trials to learn the contingencies and to specifically capture the reversal related behavior. The results remained the same if we used, for instance, 5 trials before/after a reversal instead. In order to identify the model-based and model-free contributions to behavior and whether these contributions differed between the stress and control groups, we used a mixed design ANOVA with the between-subjects factor treatment (stress vs. control manipulation) and the within-subject factors reward (rewarded vs. not rewarded) and transition (common vs. rare). Further, we performed a mixed-effects logistic regression to explain the first stage choice on each trial. First stage choice was coded as stay vs. switch and was explained as a function of previous trial’s outcome (rewarded or not rewarded) and previous trial’s transition type (common or rare). Within-subject factors (the intercept, main effects of reward and transition, and their interaction) were taken as random effects across subjects, and estimates and statistics reported at the population level. The experimental treatment (stress vs. control) was taken as a fixed effect.

We also performed exploratory analyses to test whether anxiety, depression, chronic stress or working memory capacity influence the susceptibility to stress effects on flexible learning. We tested whether these measures correlated with the sensitivity index or the model parameters. Additionally, we subdivided the stress group and the control group based on a median split on these measures, and analyzed whether individuals with particular high or low scores differed in their behavior around the reversals by using a mixed design ANOVA with the between-subjects factors treatment (stress vs. control manipulation) and level (high vs. low) and the within-subject factor time (pre reversal vs. post reversal). All analyses were performed in R (R Core Team, 2019). Greenhouse-Geisser correction was applied when sphericity was violated. Logistic regressions were conducted as mixed-effects models and were performed using the lme4 package (Pinheiro and Bates, 2000).

1.6. Computational modeling

We used reinforcement learning models to dissociate model-free and model-based contributions to subject’s trial by trial choices. We fit choice behavior to a dual-system reinforcement learning model which includes both model-free and model based learning strategies, assuming that choices derive from a weighted combination of both model-free and model-based value computations (Daw et al., 2011; Gläscher et al., 2010). Therefore, the algorithms learn a value function $Q(s,a)$ for each of the stimulus-action pairs in the two stages (three states, first stage: s_A, second stage: s_B and s_C; each with two actions). On trial t, the first stage state (always s_A) is followed by the first stage action which leads to the second stage state (s_B or s_C). The second stage action a_2 is probabilistically connected to a reward $r_{2,t}$. At each stage i of each trial t, the value for the visited state-action pair $Q(s,a)$ was updated according to both a model-free and a model-based algorithm.

Model-free values were computed with a SARSA (λ) temporal difference algorithm. As stated before, model-free choices derive from repeating previously rewarded actions. In the first trial, each state-action pair (s, a) at stage i and trial t has a Q-value of zero. In each following trial $t+1$ the value for the visited state-action pair $Q_{MF}(s_{i,\,t+1},\ a_{i,\,t+1})$ is updated based on whether the particular pairing was rewarded in the previous trial t. Therefore, the general form of the model-free value update for chosen stimulus-action pair is:

$$Q_{MF}\left(s_{i,\,t+1},\ a_{i,\,t+1}\right) = Q_{MF}\left(s_{i,\,t},\ a_{i,\,t}\right) + \alpha_i \delta_{i,\,t} \tag{1}$$

where

$$\delta_{i,\,t} = r_{i,\,t} + Q_{MF}\left(s_{i+1,\,t},\ a_{i+1,\,t}\right) - Q_{MF}\left(s_{i,\,t},\ a_{i,\,t}\right) \tag{2}$$

δ refers to the reward prediction error and α indicates the learning rates. However, this general form of value update and prediction error is narrowed down in the different stages of the task, which is explained next.

The prediction error is different for the two stages of the task. Since $r_{1,t}$ is always zero, the prediction error at the first stage is driven by the value of the selected second stage action:

$$\delta_{1,\ t} = Q_{MF}\left(s_{2,\ t},\ a_{2,\ t}\right) - \ Q_{MF}\left(s_{1,\ t},\ a_{1,\ t}\right) \tag{3}$$

This prediction error $\delta_{1,\ t}$ is used to update $Q_{MF}(s_{1,\ t},\ a_{1,\ t})$ immediately after the first choice has been made:

$$Q_{MF}\left(s_{1,\ t+1},\ a_{1,\ t+1}\right) = Q_{MF}\left(s_{1,\ t},\ a_{1,\ t}\right) + \ \alpha_1\delta_{1,\ t} \tag{4}$$

Since there is no third stage, the second stage prediction error is driven by the reward $r_{2,t}$:

$$\delta_{2,\ t} = r_{2,t} - \ Q_{MF}\left(s_{2,\ t},\ a_{2,\ t}\right) \tag{5}$$

This prediction error $\delta_{2,\ t}$ at the second stage is used to update the first and second stage model-free values, once the reward information of the outcome has become available.

$$Q_{MF}\left(s_{2,\ t+1},\ a_{2,\ t+1}\right) = Q_{MF}\left(s_{2,\ t},\ a_{2,\ t}\right) + \ \alpha_2\delta_{2,\ t} \tag{6}$$

$$Q_{MF}\left(s_{1,\ t+1},\ a_{1,\ t+1}\right) = Q_{MF}\left(s_{1,\ t},\ a_{1,\ t}\right) + \ \alpha_1\lambda\delta_{2,\ t} \tag{7}$$

Note that this update uses the already updated $Q_{MF}(s_{1,\ t},\ a_{1,\ t})$ from above, thus constituting a second update of first stage values.

The learning rates a_1 and a_2, estimated for both stages, control how much the Q-value is updated by the prediction error and therefore indicate to what extent newly acquired information overwrites old information. The learning rates are constrained between 0 and 1 with an α parameter = 0 indicating no learning and $\alpha = 1$ indicating the agent considers only the most recent information. Further, at the end of each trial the eligibility parameter λ (range 0 to 1) modulates a_1 in the second update in light of the reward information that has become available at the end of the trial . Higher values of lambda indicate more reliance to further back states and actions. In other words, λ performs a down-weighting of the first stage action based on the temporal distance from the current trial. Both the first- and second stage Q_{MF} values are updated at the second stage, with the first stage values receiving the prediction error values that were decayed by λ (see supplement in Daw et al. (2011) for details).

The model-based agent learns cumulative state-action values with a FORWARD algorithm. As described before, the model-based learner's decisions are not only determined by the reward, but also include the path that lead to the second stage's state, i.e. whether the transition was common or rare. Specifically, the algorithm computes a transition function for the first stage state-action pairs and then combines it with the second stage's reward predictions. Referring to our experimental task, this means that a model-based learner would first consider which first stage action leads to which second stage state, and then learn the reward values for the second stage actions. At the first stage, the transition function T contains the information of which first stage action maps to which second stage state. Note that in our model, the transition structure with common and rare transitions leading to 70 and 30 percent in one of the two states in the second stage was predetermined and not learned by the model (see below for a test of this supposition). At the second stage, Q_{MB} values are calculated similar to the Q_{MF} values: comparing the actual outcome of the visited state with the predicted outcome, weighted by the learning rate α (to which extent will the old information be overwritten by the new information) and the eligibility parameter λ (how far is the distance from the current trial). Model-based value expectation depends on the specification of first stage Q-values in terms of Bellman's equation (Sutton and Barto, 1998) using the transition structure P:

$$\begin{aligned} Q_{MB}\left(\mathrm{s}_{A,t+1}, a_{1,t+1}\right) = \ & \mathrm{P}(s_B \ |\mathrm{s}_{A,\ t})\ max\ Q_{MF}\left(s_{B,t},\ a_{2,t}\right) \\ & + P\left(s_C\ \mathrm{s}_A, a_{1,t}\right)\ max\ Q_{MF}\left(s_{C,t},\ a_{2,t}\right) \end{aligned} \tag{8}$$

and is recomputed at each trial, based on the current estimates of the transition probabilities and second stage reward values. Because model-based and model-free algorithms coincide at the second stage, we set $Q_{MB} = \ Q_{MF}$ at this level.

Finally, we assume that behavior derives from a weighted combination of both model-based and model-free value computations. Therefore, we define net action values at the first stage as the weighted sum of model-based and model-free values

$$Q_{net}\left(\mathrm{s}_{A,t+1}, a_{1,t+1}\right) = wQ_{MB}\left(\mathrm{s}_{A,t}, a_{1,t}\right) + (1-w)Q_{MF}\left(\mathrm{s}_{A,t}, a_{1,t}\right), \tag{9}$$

where w is a weighting parameter. This parameter is assumed to be constant across trials, with $w = 0$ reflecting purely model-free value computing and $w = 1$ purely model-based reinforcement learning. The probability of a choice is composed by a softmax for Q_{net} at the first stage:

$$\mathrm{P}(a_{1,\ t} = a\ |\mathrm{s}_{A,\ t}) = \frac{\exp\left(\beta_1\left[Q_{net}\left(s_{1,\ t},\ a\right) + p\ *rep(a)\right]\right)}{\sum_{a'}\exp\left(\beta_1\left[Q_{net}\left(s_{1,\ t},\ a'\right) + p\ *rep(a')\right]\right)} \tag{10}$$

where the inverse temperature parameters β_1 and β_2 indicate the randomness of the choice by specifying the extent to which the values are updated based on the learned information. Temperature parameters are set from 0 to ∞ with lower values indicating more randomness in choice behavior. The stay parameter p, ranging from 0 to 1, captures first-order perseveration in the first stage, together with the indicator function rep that is 1 when the current first stage action is the same as in the previous trial. The stay parameter was omitted at the second stage and hence the softmax is defined as:

$$\mathrm{P}(a_{2,\ t} = a\ |\mathrm{s}_{2,\ t}) = \frac{\exp\left(\beta_2\left[Q_{net}\left(s_{2,\ t},\ a\right)\right]\right)}{\sum_{a'}\exp\left(\beta_2\left[Q_{net}\left(s_{2,\ t},\ a'\right)\right]\right)} \tag{11}$$

In total, the algorithm contains 7 free parameters (a_1, a_2, β_1, β_2, λ, p, w) which were fit separately for each participant using the probabilistic programming language Stan through its MATLAB interface (Carpenter et al., 2017).

With the help of the model-parameters determined for each subject we were able to draw conclusions about the learning strategies used. We conducted group comparisons for each parameter to identify general differences in behavioral tendencies between the stress group and the control group.

To determine different learning strategies in the neuroimaging data we calculated three different prediction errors. Therefore, we extracted each participant's best fitting parameters and reran the task, resulting in model predictions on a trial basis. In addition to the actual prediction with the individual w-parameter, we also created model-based or model-free predictions by again inserting the parameters in the task, but this time not the fitted w-parameter, but with $w = 0$ and $w = 1$, reflecting pure model-free and pure model-based behavior, respectively. Thus, we obtain three predicted datasets for each subject. This allows us to distinguish between model-based and model-free prediction errors for the value update at stage 1.

For Q_{MB}, we set $w = 1$:

$$Q_{MB}\left(\mathrm{s}_{i,t+1}, a_{i,t+1}\right) = 1{*}Q_{MB}\left(\mathrm{s}_{i,t}, a_{i,t}\right) + (1-1)Q_{MF}\left(\mathrm{s}_{i,\ t}, a_{i,t}\right) \tag{12}$$

Likewise, for Q_{MF}, we determine $w = 0$:

$$Q_{MF}\left(\mathrm{s}_{i,t+1}, a_{i,t+1}\right) = 0{*}Q_{MB}\left(\mathrm{s}_{i,t}, a_{i,t}\right) + (1-0)Q_{MF}\left(\mathrm{s}_{i,t}, a_{i,t}\right) \tag{13}$$

These predicted Q-values are used to derive prediction errors (see Eq. (3)):

$$PE_{MB}\left(s_{1,t},\ a_{1,\ t}\right) = Q_{MB}\left(\mathrm{s}_{2,t}, a_{2,t}\right) - Q_{MB}\left(\mathrm{s}_{1,t}, a_{1,t}\right) \tag{14}$$

$$PE_{MF}\left(s_{1,t},\ a_{1,\ t}\right) = Q_{MF}\left(\mathrm{s}_{2,t}, a_{2,t}\right) - Q_{MF}\left(\mathrm{s}_{1,t}, a_{1,t}\right) \tag{15}$$

Finally, we identify the reward prediction error RPE which is calculated when the outcome is presented:

$$RPE\left(s_{2,\ t},\ a_{2,\ t}\right) = r_{2,t} - \ Q_{net}\left(s_{2,\ t},\ a_{2,\ t}\right) \tag{16}$$

Table 1
Model Comparison using WAIC.

Model Name	α	β	p	w	λ	ε	nParams	Control	Stress
Full	**2**	**2**	**1**	**1**	**1**	**0**	**7**	**11,191.15**	**12,056.88**
full + state space	2	2	1	1	1	1	8	11,202.57	12,062.26
no p	2	2	0	1	1	0	6	11,436.80	12,359.40
one α	1	2	1	1	1	0	6	11,214.76	12,072.60
one β	2	1	1	1	1	0	6	11,218.68	12,077.35
no p_one α	1	2	0	1	1	0	5	11,495.35	12,450.99
no p_one β	2	1	0	1	1	0	5	11,542.43	12,433.98
one α_one β	1	1	1	1	1	0	5	11,541.47	12,436.59
no p_one α_one β	1	1	0	1	1	0	4	11,628.99	12,509.57

1.7. Model validation

To validate the model fit, we compared our fully parameterized hybrid model (Daw et al., 2011; Gläscher et al., 2010) to various reduced nested versions. The model should be as complex as necessary to adequately represent behavior, but only as complex as justified by the data. We compared our model to several other models that are simplified by removing different parameters, e.g. a version without the stay bias (p), a version with only one learning rate (α), a version with only one temperature parameter (β), and combinations of these reductions (Table 1). We also explicitly tested whether state space learning played a significant role in task performance because the participants did not know the transition probabilities for common and rare transitions at the beginning of the experiment. We therefore included a mechanism by which a participant can learn the transition probabilities during task execution, which we have used in a prior publication (Gläscher et al., 2010). Transition probabilities are stored in a transition matrix T, which can be learned using a state prediction error (see Gläscher et al., 2010). Specifically, each element of T specifies the probability of the reached second stage state s_2 from the first stage state s_1 via an action a_1 ($T(s_{1,t}, a_{1,t}, s_{2,t})$). In the state space learning model, all transition probabilities are initially set to 0.5 reflecting no prior knowledge about the transitions by the participants. Upon every trial all possible transitions following action $a_{1,t}$ are updated according to the following learning rules:

$$T\left(s_{1,t+1}, a_{1,t+1}, s_{2,t+1}\right) = T\left(s_{1,t}, a_{1,t}, s_{2,t}\right) + \varepsilon(1 - T\left(s_{1,t}, a_{1,t}, s_{2,t}\right) \tag{17}$$

$$T\left(s_{1,t+1}, a_{1,t+1}, \neg s_{2,t+1}\right) = T\left(s_{1,t}, a_{1,t}, \neg s_{2,t}\right) - \varepsilon\ T\left(s_{1,t}, a_{1,t}, \neg s_{2,t}\right) \tag{18}$$

where $T(s_{1,t}, a_1, s_{2,t})$ is the probability of transitions from the first stage state $s_{1,t}$ to the second stage state $s_{2,t}$ using action $a_{1,t}$ on trial t, $T(s_{1,t}, a_1, \neg s_{2,t})$ is the unrealized transition to the other possible second stage state, and ε is the learning rate for state space learning, modeled with an initial uniform Beta(1,1) prior. We think that updating both the realized and the unrealized state transition following the same action $a_{1,t}$ is a reasonable approach given that participants are probably aware (at least in the latter parts of the experiment) that action $a_{1,t}$ could have also resulted in a different transition. All other components of the state space learning model are identical to the full learning model, including the linear weighting of model-free and model-based learning (see equations above). Model comparisons were performed by calculating the widely applicable information criterion (WAIC; Watanabe, 2010) which indicates prediction performance and assesses the quality of a model, relative to the quality of other candidate models by estimating the posterior likelihood, followed by a correction for the effective number of parameters to adjust for overfitting. This approach is often used for comparing models estimated using Markov Chain Monte Carlo sampling as in our case and confirmed that the full model outperformed all competing versions (Table 1).

A fully parameterized hybrid model without a state space learning component fitted subjects' choices best in a model comparison that considers differences in model complexity. Model performance is indicated by the widely applicable information criterion (WAIC), presented separately for the stress group and the control group. Lower values represent a better fit. The full model contains two learning rates (α), two temperature parameters (β), the stay bias (p), the weighting parameter (w) and the eligibility parameter (λ) and was compared to several other models that are simplified by removing different parameters, respectively, or included the state space learning rate ε.

1.8. MRI data acquisition and analysis

Functional imaging was conducted using a 3 T Siemens (Erlangen, Germany) MAGNETOM Prisma scanner, equipped with a 64-channel head coil, to acquire gradient echo T2*-weighted echo-planar-images (EPI) with BOLD contrast. For each of the three functional runs, we collected about 600 vol with the following parameters: 60 slices, slice thickness = 2 mm, flip angle 60%, FOV 224 × 224, repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, voxel size 2.0 mm isotropic. Slice orientation was tilted −30° from the line connecting the anterior and posterior commissure to alleviate signal drop out in the orbitofrontal cortex (Deichmann et al., 2003). We additionally acquired a high-resolution T1-weighted anatomical image (TR = 2.5 s, TE = 2.12 ms, 256 slices, voxel size = 0.8 × 0.8 × 0.9 mm). Preprocessing of functional images was performed with MATLAB and SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first five functional images were discarded from the analysis to allow for T1 saturation effects. The remaining functional images were first spatially realigned, then coregistered to the structural image, followed by a normalization to the MNI space. The images were additionally spatially smoothed using a 4 mm full-width half-maximum Gaussian kernel.

Subject-specific design matrices were defined using general linear modeling (GLM) as implemented in SPM12. We entered three regressors coding the average BOLD response at each of the three states (two choice states, one outcome state). The model-derived prediction error signals (model-free prediction error PE_{MF} and model-based prediction error PE_{MB}) were entered as parametric modulators, modeled at the onset of the second stage. We chose this time point because we assume that the relevant learning computations are integrated to a value update when the second decision is required. A parametric regressor coding the received outcomes (reward = +1, no reward = 0) was modeled at the time of the outcome. Model-derived reward prediction errors (RPE) were modeled as parametric modulators on the outcome onsets, because here we are specifically interested in the process of reward-related value updating. Moreover, we subdivided behavioral data into "advantageous-choice-trials" and "disadvantageous-choice-trials", separately for the first and the second stage and entered their onsets into our GLM. For the second-level models, the contrasts of interest "model-based prediction error", "model-free prediction error", "reward prediction error", "reward", "advantageous choice stage 1" and "advantageous choice stage 2" were defined. These difference contrasts were taken to a second-level group two-sample t-test, allowing a direct comparison between the stress and control group.

Based on our a-priori hypotheses, analyses were restricted to brain areas that have previously been implicated in model-based and model-free reinforcement learning (Daw et al., 2011; Gläscher et al., 2010; Lee et al., 2014). We used the following anatomical masks from the Harvard-Oxford atlas: putamen, caudate, and hippocampus. In order to test for potential differential involvement of anterior and posterior regions of the hippocampus in model-based and model-free learning under stress, we divided a hippocampal mask along the y-axis into three parts with approximately equal lengths, using the WFU pick-atlas (Lancaster et al., 2000; Maldjian et al., 2003): posterior hippocampus from $Y = -40$ to -30, medial hippocampus from $Y = -29$ to -19, and anterior hippocampus from $Y = -18$ to -4. For a more detailed description see Collin et al. (2015) and Dandolo and Schwabe (2018). Moreover, we used anatomical masks for lateral orbitofrontal cortex from the Montreal atlas and combined the AAL atlas-masks for frontal superior medial, frontal middle and frontal superior to a medial prefrontal cortex mask, as implemented in the WFU PickAtlas Tool (Maldjian et al., 2003). 10 mm spheres centered on the peak voxel of bilateral ventral striatum (left peak: −9 2 8, right peak: 9 5 −8), bilateral insulae (left peak: −30 20 −2, right peak: 33 29 7) and ilPFC (left peak: −54 38 3, right peak: 48 35 −2) were created, because they were previously associated with model-free and model-based learning strategies (Lee et al., 2014). We applied a small volume correction (svc) for the areas of interest with an initial uncorrected threshold of 0.05 on whole-brain-level. The svc was applied on voxel level. Voxels were regarded as significant, when falling below a corrected voxel threshold of 0.05 (family wise error (FWE) corrected) adjusted for the small volume.

2. Control variables

To control for personality traits and behavioral tendencies that may affect flexible learning and decision-making in general, participants filled out several questionnaires at the beginning of the experiment. In particular, participants completed German versions of the State-Trait Anxiety Inventory (STAI; Spielberger et al. 1970), the Trier Inventory of Chronic Stress (TICS; Schulz & Schlotz 1999) and the Beck Depression Inventory (BDI; Beck et al. 1961).

3. Results

3.1. Successful stress induction

Participants first underwent the TSST, a standardized stress protocol consisting of a mock job interview, or a non-stressful control procedure. Subjective and physiological measurements confirmed the successful stress induction through the TSST (Fig. 3A-E). The TSST was experienced as significantly more difficult ($t(56) = 5.73$, $p = 4.12e^{-07}$, $d = 1.51$), unpleasant ($t(56) = 6.70$, $p = 1.09\, e^{-08}$, $d = 1.76$), and stressful ($t(56) = 5.55$, $p = 8.14e^{-07}$, $d = 1.46$) than the control manipulation. Moreover, the TSST, but not the control procedure, led to increased systolic blood pressure (treatment × time: $F(3168) = 16.67$, $p = 1.59e^{-09}$; $\eta^2_{ges} = 0.059$), diastolic blood pressure ($F(2.64, 148.01) = 15.67$, $p = 3.29e^{-08}$ (Greenhouse-Geisser corrected), $\eta^2_{ges} = 0.080$), and pulse ($F(2.41, 134.77) = 14.39$, $p = 3.83e^{-07}$(Greenhouse-Geisser corrected), $\eta^2_{ges} = 0.048$), indicating significant autonomic activation in response to the TSST. Finally, the TSST, but not the control manipulation, induced a pronounced increase in salivary cortisol (treatment × time: $F(2.43, 136.31) = 10.70$, $p = 3.83e^{-07}$ (Greenhouse-Geisser corrected), $\eta^2_{ges} = 0.0475$). While groups did not differ in cortisol concentrations before the TSST ($t(56) = -0.35$, $p = 0.73$, $d = -0.09$), cortisol concentrations were significantly higher in the stress group than in the control group at all time points of measurement after the manipulation (all $p \leq 0.05$). Peak cortisol levels were reached ∼18 min after stressor onset, shortly before the Markov decision task in the MRI began, and cortisol levels remained significantly elevated throughout the task.

3.2. Stress reduces the behavioral sensitivity to reversals of reward contingencies

In order to examine how stress changes the flexibility of learning, participants completed a modified Markov decision task in the MRI scanner about 20 min after the onset of the stress or control manipulation. This task was designed to dissociate model-free and model-based learning (Daw et al., 2011; Gläscher et al., 2010) and involved two subsequent choices, each between two fractal stimuli (Fig. 2). The first stage decision led to a second stage, requiring another choice between two options which were associated with different probabilities of monetary reward. Each of the first stage options was predominantly associated with one or the other state in the second stage. Whether or not the transition between the first and the second stage is considered in the decision allows conclusions to be drawn about the underlying learning strategy. While a purely model-free learning strategy only accounts for whether the previous action led to a reward in the second stage, a model-based learner would also include the path that led to the result in the subsequent decision. Learning performance was quantified by the proportion of first stage choices for the stimulus that led predominantly to the second stage state with the overall higher probability to obtain a reward (0.9 | 0.4 vs. 0.6 | 0.1). Likewise, successful learning in the second stage was associated with the proportion of choices for the option with the higher reward probability (either 0.9 or 0.6).

The stress and control groups did not differ in the overall proportion of advantageous choices, neither in the first stage ($t(56) = 1.123$, $p = 0.266$, $d = 0.295$), nor in the second stage ($t(56) = -0.239$, $p = 0.81$, $d = -0.062$). This pattern of results is generally in line with previous findings suggesting that the stress-induced alteration in the nature of learning becomes apparent only when the environment changes and the flexibility of behavior is probed (Kim et al., 2001; Schwabe and Wolf, 2009; Schwabe et al., 2010). The proportion of advantageous first stage choices did not differ between blocks (main effect block: $F(1, 56) = 0.04$), $p = 0.84$, $\eta^2_{ges} = 0.0003$; treatment × block: $F(1, 56) = 0.03$, $p = 0.87$, $\eta^2_{ges} = 0.0002$), neither did the proportion of advantageous second stage choices (main effect block: $F(1, 56) = 1.72$), $p = 0.20$, $\eta^2_{ges} = 0.10$; treatment × block: $F(1, 56) = 1.23$, $p = 0.27$, $\eta^2_{ges} = 0.007$).

In a next step, we analyzed participants' behavioral response to the reversal by comparing the proportion of advantageous choices in the four trials before a reversal relative to the four trials after a reversal between the stress and control groups (Fig. 4). For the first stage choices, the proportion of advantageous choices was – as expected – overall significantly lower after a reversal than before (main effect of time; $F(3168) = 25.018$, $p = 5.95e^{-06}$, $\eta^2_{ges} = 0.23$), post-hoc *t*-test pre vs. post: $t(57) = 4.87$, $p = 9.21\ e^{-06}$, $d = 0.64$). Interestingly, the change in first stage choices from pre- to post-reversal differed significantly between groups (treatment × time: $F(1, 56) = 4.104$, $p = 0.048$, $\eta^2_{ges} = 0.047$). Post-hoc t-tests revealed that the proportion of advantageous choices in the pre-reversal trials was significantly lower in the stress group than in the control group ($t(56) = -2.25$, $p = 0.03$, $d = -0.59$), while groups did not significantly differ in the proportion of advantageous choices after a reversal ($t(56) = 1.19$, $p = 0.24$, $d = 0.31$). To verify that the predictions of our model matched the actual data around the reversals, we performed posterior predictive checks. Therefore, we generated 50 simulations for each participant, in which we entered each individual set of optimized parameters into our version of the Markov decision task. Averaging over these simulations we obtained the posterior predictive peri-reversal time course of advantageous choices. This showed a pattern very similar to the actual data (Fig. 4, right panel).

To test whether the differential influence of reversals on choice behavior in the stress group relative to the control group cannot be explained by a general learning impairment in the stress group, we tested whether the proportion of advantageous choice in the stress group differed from chance level. The proportion of advantageous choices in the first stage was significantly different from chance (i.e. 50 percent; $t(28) = 3.03$, $p = 0.005$), indicating that the stress group had learned the contingencies before a reversal took place. Moreover, the proportion of advantageous choices in the stress group differed in the four trials before a reversal versus four trials after a reversal ($t(28) = 2.14$, $p = 0.04$, $d = 0.40$), indicating that the reversal did affect the behavior of stressed participants, but to a lesser extent than in controls.

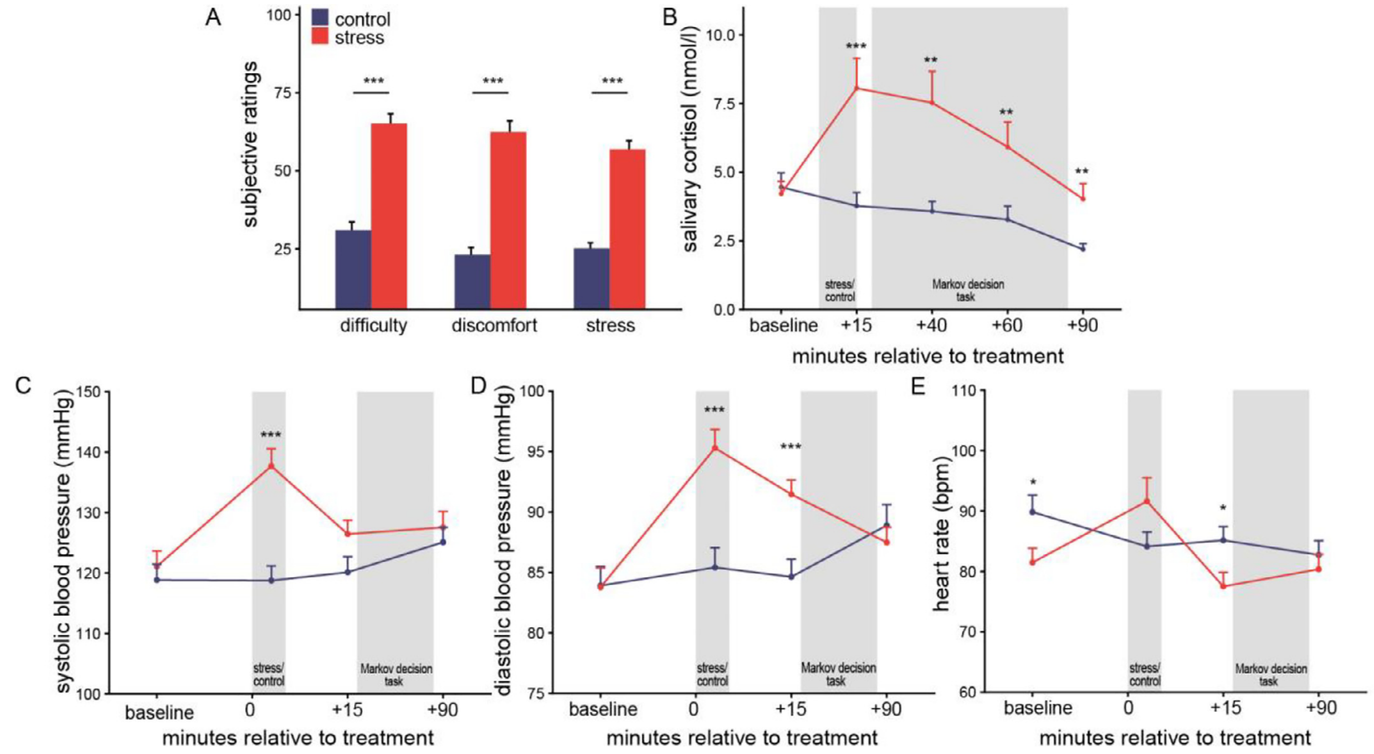


Fig. 3. Successful stress induction. (A) Participants in the stress condition rated the treatment as significantly more difficult, unpleasant, and stressful than participants in the control condition. The exposure to the stressor led further to significant increases in (B) salivary cortisol levels, (C) systolic blood pressure, and (D) diastolic blood pressure. (E) Heart rate measures were significantly lower in the stress group than in the control group at baseline (t(56) = −2.28, $p = 0.03$). Measures increased significantly after stress, relative to baseline (t(28) = 3.34, $p = 0.002$), but decreased at control treatment (t(28) = −4.37, $p = 0.0002$); error bars represent standard errors of the mean, $^{**}p < 0.01$, $^{***}p < 0.001$ for the comparison between the stress group and the control group.

The observed group differences in the first stage are particularly intriguing as the first stage choice indicates the integration of the task structure into the decision. A large proportion of decisions that lead to the better second state suggest an understanding of the state space and the associated transitions (model-based learning) - regardless of the reward obtained in the end. We further tested whether participants' behavior around the reversal, expressed as mean number of advantageous first stage choices in the four trials before a reversal minus the four trials after a reversal, differed between blocks. This analysis showed that neither the stress group (F(1, 28) = 0.04, $p = 0.84$, $\eta^2_{ges} = 0.002$), nor the control group (F(1, 28) = 1.89, $p = 0.18$, $\eta^2_{ges} = 0.06$) changed in their sensitivity to reversals across the three blocks of the experiment.

The proportion of choices for the option with the higher reward probability in the second stage was also significantly lower after a reversal than before (main effect of time; F(1, 56) = 89.948, $p = 3.007e^{-13}$, $\eta^2_{ges} = 0.424$), but did not differ between groups (treatment × time: F(1, 56) = 0.099, $p = 0.755$, $\eta^2_{ges} = 0.0008$). Accordingly, the change in the proportion of advantageous second stage choices from before to after the reversal did not differ between the stress and control groups (t(56) = −0.289, $p = 0.773$, $d = -0.076$; Fig. 4B).

3.3. Model-based and model-free contributions to behavior

In order to capture model-free and model-based contributions to choice behavior, we conducted a logistic regression analysis. The previous trial's transition type and outcome were used to explain whether participants chose the same action again or whether they switched to the other option. This analysis allows a dissociation of model-free and model-based contributions because both learning strategies make qualitatively distinct predictions about how the previous trial's characteristics influence the first stage choice in the following trial. Fig. 5 (left) shows the theory-based choice behavior of purely model-free and model-based learners. A pure model-free strategy predicts that a rewarded action will be repeated, regardless of the transition type (main effect of reward). A model-based agent, on the other hand, uses its knowledge of the task structure and therefore predicts an interaction between transition and reward. The data predicted by our model suggest a mixture of model-free and model-based learning strategies, without differences between the stress group and the control group (Fig. 5, middle).

The logistic regression analysis confirmed the basic signature of model-free reinforcement learning to behavior, indicated by an increased probability to stay when the previous trial was rewarded ($z = 5.715$, $p = 1.10e^{-08,}$ $\beta = 1.295$), as well as the contribution of model-based strategies as indicated by a reward × transition interaction with an additional increase in stay probabilities when a reward was obtained after a common transition ($z = 2.586$, $p = 0.0097$, $\beta = 0.380$). Thus, participants demonstrated both model-based and model-free elements of learning. However, as shown in Fig. 5 (right), the balance of model-based and model-free contributions appeared to be overall biased towards more model-free learning, without significant differences between groups (stress × reward, $z = -1.048$, $p = 0.295$, $\beta = -0.330$; stress × reward × transition, $z = -1.181$, $p = 0.238$, $\beta = -0.235$).

3.4. Stress effects on model-based and model-free parameters

In a next step, we used reinforcement learning models to dissociate model-free and model-based contributions to participants' trial-by-trial choices. We fitted choice behavior to a dual-system reinforcement learning model which includes both model-free and model based learning strategies (Daw et al., 2011; Gläscher et al., 2010). The algorithm contained 7 parameters, fitted individually for each participant.

We assumed that choices were driven by the weighted average of these two computations. The weighting parameter w shows a predominance of model-free proportions in choice behavior (mean control

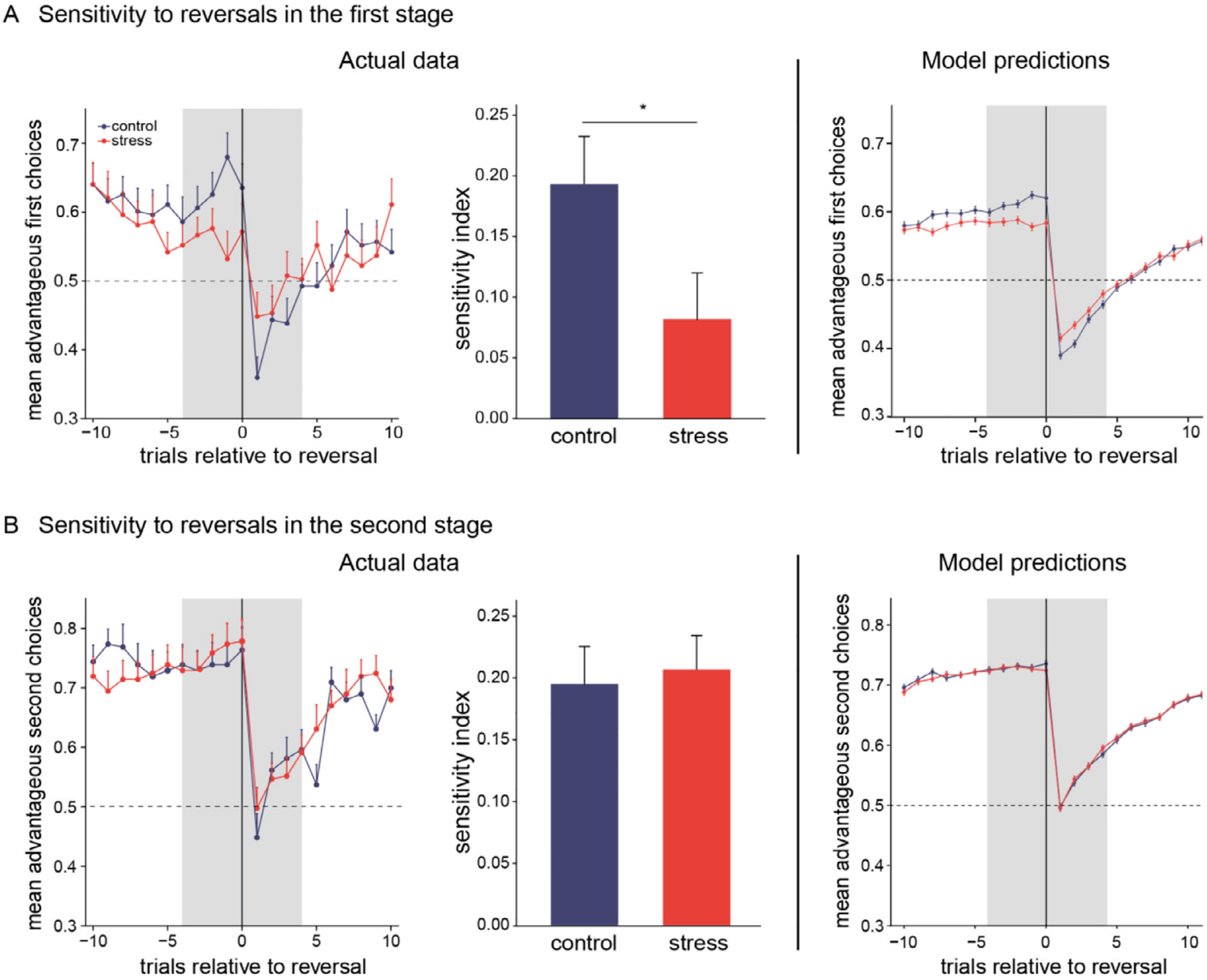


Fig. 4. Stress reduces the behavioral sensitivity to reversals in the first stage. The proportion of advantageous first stage choices is higher in the four trials before a reversal than in the four trials after a reversal, indicating that the reversals have an effect on behavior (A, B). The sensitivity index, computed as the mean of advantageous choices before vs. after a reversal, is significantly higher in the control group than in the stress group in the first stage (A), while the sensitivity index for the second choice does not differ between the stress group and the control group (B). Right panels: Model simulations with best fitting parameters for the trials around the reversals show a pattern similar to the actual behavioral data.

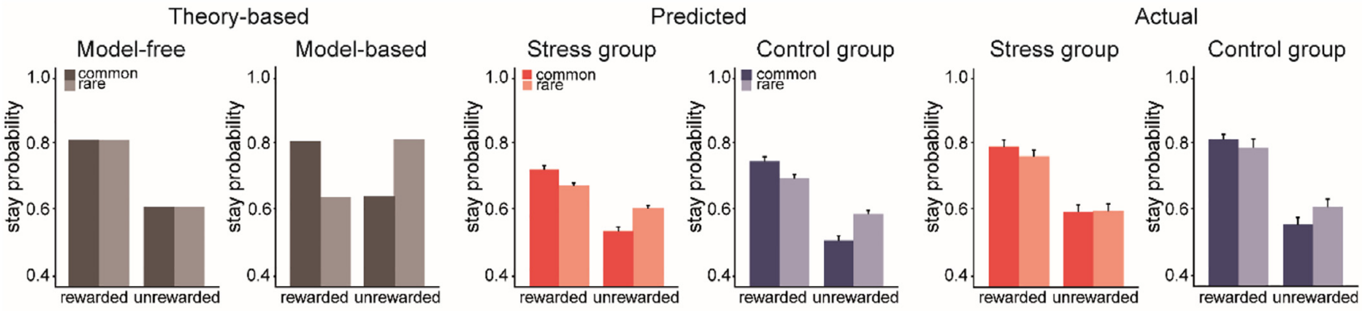


Fig. 5. Factorial analysis of choice behavior. Left: Pure model-free reinforcement learning predicts that a previously rewarded action is more likely to be repeated on the subsequent trial, regardless of whether that reward occurred after a common or a rare transition. Pure model-based behavior comprises a knowledge of the task structure: a reward obtained via a rare transition predicts a switch to the other option. Middle: Data obtained from a posterior predictive check using the set of model parameters estimated for each participant suggests a mixture of both model-free and model-based learning strategies. Right: Actual Data. Participants show both model-based and model-free learning with an overall bias toward model-free learning, independent of group assignment.

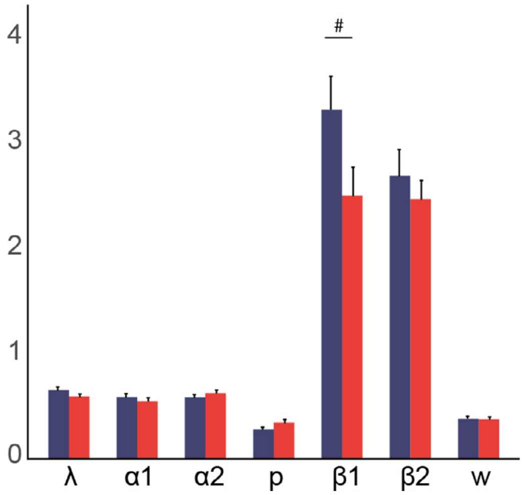


Fig. 6. Stress effects on the model parameters. Best-fitting parameter estimates, shown across subjects. The stress group tended to show a reduced temperature in the first stage, compared to the control group ($t(56) = 1.96$, $p = 0.056$, $d = 0.51$), indicating more random or exploring choice behavior; no group differences in the eligibility parameter λ, the two learning rates α_1 and α_2, the stay bias p, or the weighting parameter w; error bars represent standard errors of the mean, # $p < 0.06$ for the comparison between the stress group and the control group.

group: 0.38, mean stress group: 0.37, test against 0.50: both $p < 0.001$), which was comparable in the stress and control groups ($t(56) = 0.10$, $p = 0.918$, $d = 0.03$), indicating that acute stress did not alter the weight of model-free and model-based contributions to learning per se (Fig. 6). However, the temperature parameter for the first stage choice tended to be lower in the stress group ($t(56) = 1.96$, $p = 0.056$, $d = 0.51$; all other parameters remained unaffected by stress, all $p > 0.20$, Fig. 6). This temperature parameter was significantly positively associated with both the proportion of advantageous first stage choices ($r(56) = 0.45$, $p = 0.0004$) and the sensitivity index ($r(56) = 0.44$, $p = 0.0006$). In the second stage, the corresponding temperature parameter was also positively correlated with the proportion of advantageous choices ($r(56) = 0.48$, $p = 0.0002$), as well as the sensitivity index ($r = 0.48$, $p = 0.0002$). Furthermore, the inverse temperature parameter reflects the extent to which the underlying value computations are used to guide choices, in the sense of an exploration – exploitation trade off or a measure of choice stochasticity. Our results thus point to a rather explorative choice behavior in the stress group, or more random first stage decisions, suggesting that the stressed participants did not use the first decision as a planning step for the second stage, but may have randomly made the first decision in order to proceed to the reward-guided second choice. In other words, stress appeared to affect the utilization of value computations for the first stage choice.

These results are in line with our findings that stress reduced the sensitivity to reversals in the first stage. In addition, we tested whether the sensitivity index correlates with the weighing parameter w. Our results showed no such correlation ($r = -0.17$, $p = 0.19$). The absence of a correlation between the weighing parameter and participants' sensitivity to a reversal was not surprising given that we assume that both model-based and model-free processes may contribute to flexible learning and the sensitivity to changes in the environment. Further modeling parameters did not correlate with the sensitivity index (all r ⟨ |0.22|, all p ⟩ 0.1).

3.5. *Stress affects the neural underpinnings of both model-based and model-free learning*

Our behavioral results suggested that the stress group tended to show more explorative or random choice behavior at the first stage than the control group. Directly building on this pattern of results, we compared the brain activity at advantageous first stage choices with disadvantageous at that time point between the stress and the control group. This analysis showed that stressed participants had significantly reduced activity in the medial prefrontal cortex (mPFC; peak −16 10 62, $p_{svc} = 0.03$, FWE, Fig. 7B), compared to the control group. Comparing advantageous choices to disadvantageous choices in the second stage, the control group tended to show a higher activity in the ventral striatum (peak 2 10 −8, $p_{svc} = 0.07$, FWE).

Next, we regressed the model-derived prediction errors against the fMRI data collected during the Markov task. Corroborating earlier reports (Daw et al., 2011; Gläscher et al., 2010; Lee et al., 2014), our data pooled over both groups showed that reward prediction-errors were computed in the lateral OFC, ilPFC, mPFC, ventral striatum, putamen, insula and in the hippocampus (all $p_{svc} < 0.016$, FWE). Reward onsets were associated with activity in the ilPFC, mPFC and insula (all $p_{svc} < 0.03$, FWE). Interestingly, reward onsets were associated with increased activity in the posterior hippocampus (peak −22 −34 −2, $p_{svc} = 0.018$, FWE, Fig. 7C) in the stress group. The computation of model-free prediction errors was associated with the lateral OFC, the ventral striatum and the anterior hippocampus (all $p_{svc} < 0.017$, FWE) and model-based prediction errors with activity in the hippocampus, lateral OFC, mPFC and putamen (all $p_{svc} < 0.009$, FWE).

The table shows MNI (Montreal Neurological Institute) coordinates for local maxima in mm. All areas with $k > 5$ significant voxels are reported. For our regions of interest (ROIs), we implemented small volume correction (SVC) using an initial threshold of $p < 0.05$, uncorrected. The significance threshold was set to $p < 0.05$, family wise error (FWE) corrected.

Most interestingly, these neural underpinnings of both model-free and model-based learning were affected by stress (Table 2). Compared to controls, stressed participants showed reduced correlations between BOLD activity and model-free prediction errors in the right ilPFC (peak 48 32 −8, $p_{svc} = 0.005$, FWE; Fig. 7A) and a tendency to reduced activation in the left amygdala (peak −24 −8 −18, $p_{svc} = 0.059$, FWE). For model-based prediction errors, stressed participants showed, relative to controls, reduced activity in the right putamen (peak 30 −10 12, $p_{svc} = 0.032$, FWE, Fig. 7D) and a higher activation in the right ilPFC (peak 48 32 −8, $p_{svc} = 0.005$, FWE, Fig. 7D). At trend level, stress increased activity in the right insula (peak 32 30 −2, $p_{svc} = 0.059$, FWE) and led to a decrease in the activity of the right amygdala (peak 30 −4 −20, $p_{svc} = 0.054$, FWE). Moreover, stress tended to reduce activity in the hippocampus (peak −24 −34 −4, $p_{svc} = 0.078$, FWE), a region only rather recently implicated in model-based behavior (Vikbladh et al., 2019). Because it is assumed that there is a functional separation along the hippocampal anterior-posterior axis (Fanselow and Dong, 2010; Poppenk et al., 2013; Strange et al., 2014), we further subdivided the hippocampus into anterior and posterior parts, in accordance with previous studies (Collins et al., 2015; Dandolo and Schwabe, 2018), and tested whether the obtained stress effect was specific to the anterior or posterior hippocampus. This analysis revealed that stress affected indeed solely the posterior hippocampal contribution to model-based behavior (peak −24 −34 −2, $p_{svc} = 0.019$, FWE), while there was no stress effect on the anterior hippocampus (left: $p_{svc} = 0.78$, right: $p_{svc} = 0.17$, FWE).

3.6. *Exploratory analysis of control variables and working memory influences*

To control for personality traits and behavioral tendencies that may affect flexible learning or modulate stress effects on flexible learning, we measured state anxiety, trait anxiety, depressive symptoms and chronic stress via the STAI-S, STAI-T, BDI and TICS, respectively. Because one subject code was mistakenly assigned twice, we could not use the questionnaire data of two participants, resulting in $n = 56$ for the follow-

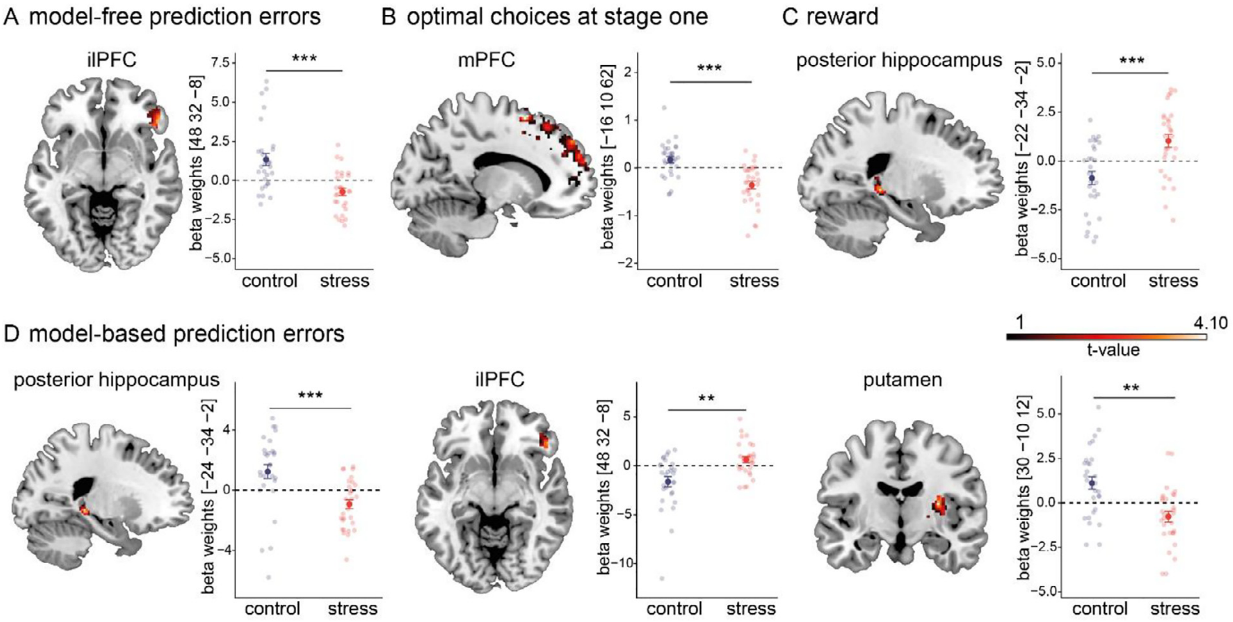


Fig. 7. Stress reduces posterior hippocampal activity during model-based learning and inferiorlateral prefrontal activity during model-free learning. (A) The stress group showed reduced activity during model-free error computation contrary to the control group in the right ilPFC. (B) The stress group showed reduced activity in the mPFC during advantageous choices in the first stage. (C) The stress group showed a higher activity in the posterior hippocampus during reward computations, compared to the control group. (D) Model-based prediction errors were associated with a stress-induced reduction of the posterior hippocampus and the putamen, while the stress group showed an increased activity in the ilPFC, compared to the control group. Data are thresholded at $p < 0.05$, uncorrected, for display purposes only. Parameter estimates were extracted for the peak voxel; error bars represent standard errors of the mean, $^{***}p < 0.001$ for the comparison between the stress group and the control group.

Table 2
Stress effects on neural representations of learning computations.

contrast name	ROI name	Cluster	P_{FWE}	t_{max}	MNI coordinates		
					X	Y	Z
Model-based prediction errors							
control > stress	posterior hippocampus (L)	64	0.019	3.98	−22	−34	−2
control > stress	putamen (R)	74	0.032	4.02	30	−10	12
stress > control	ilPFC (R)	72	0.047	3.61	48	32	−8
Model-free prediction errors							
control > stress	ilPFC (R)	128	0.005	4.58	48	32	−8
Rewards							
stress > control	posterior hippocampus (L)	35	0.018	4.00	−22	−34	−2
Optimal first stage choices							
control > stress	mPFC (L)	26	0.035	4.88	−16	10	62

ing analyses. Importantly, stress and control groups did not differ in depressive symptoms (t(54) = 1.62, p = 0.11, d = 0.43), state anxiety (t(54) = 0.33, p = 0.74, d = 0.089), trait anxiety (t(54) = 1.16, p = 0.25, d = 0.31), or subjective chronic stress (t(54) = 0.89, p = 0.38, d = 0.24, Table 1). Furthermore, in light of previous evidence suggesting that anxiety, depressive symptoms or chronic stress may be associated with the vulnerability to stress and changes in model-based behavior (Nasca et al., 2015; Radenbach et al., 2015; Weger and Sandi, 2018), we further tested whether the questionnaire data correlated with the sensitivity index or model-derived parameters. These analyses yielded no significant correlations between the sensitivity index and state / trait anxiety, chronic stress, or depressive symptoms (stress: all |r| < 0.16, all p > 0.4, control: all |r| 〈 0.37, all p 〉 0.06, all participants: all |r| 〈 0.25, all p 〉 0.06), except for a significant negative correlation between STAI-S scores and the sensitivity index in the control group (r = −0.418, p = 0.03), which would however not survive a correction for multiple comparisons. When we subdivided participants into subgroups based on a median-split on the respective questionnaire score, we obtained evidence suggesting that acute stress might influence participants' behavioral response to the reversal in particular in individuals with high trait or state anxiety. Further, stress and control groups appeared to differ in particular when participants reported low chronic stress and low levels of depressive mood (see supplemental Figure S1 and supplemental Table S2). These analyses, however, were exploratory and need to be interpreted with great caution.

Because there is evidence that high baseline working memory might protect model-based learning from deleterious stress effects (Otto et al., 2013), participants completed an n-back test, as common measure of working memory (Owen et al., 2005), before they underwent the stress or control manipulation. The working memory data of four participants are missing due to technical failure. Importantly, groups did not differ in baseline working memory performance (t(52) = −1.38, p = 0.17, d = −0.38). When we analyzed correlations between baseline working memory performance on the one hand and the task performance (i.e. the sensitivity index) on the other hand, we obtained no significant correlations, neither within the stress or control groups (stress: r(23) = 0.293, p = 0.155; control: r(27) = −0.004, p = 0.984), nor across all participants (r = 0.178, p = 0.197). These correlational data suggest that base-

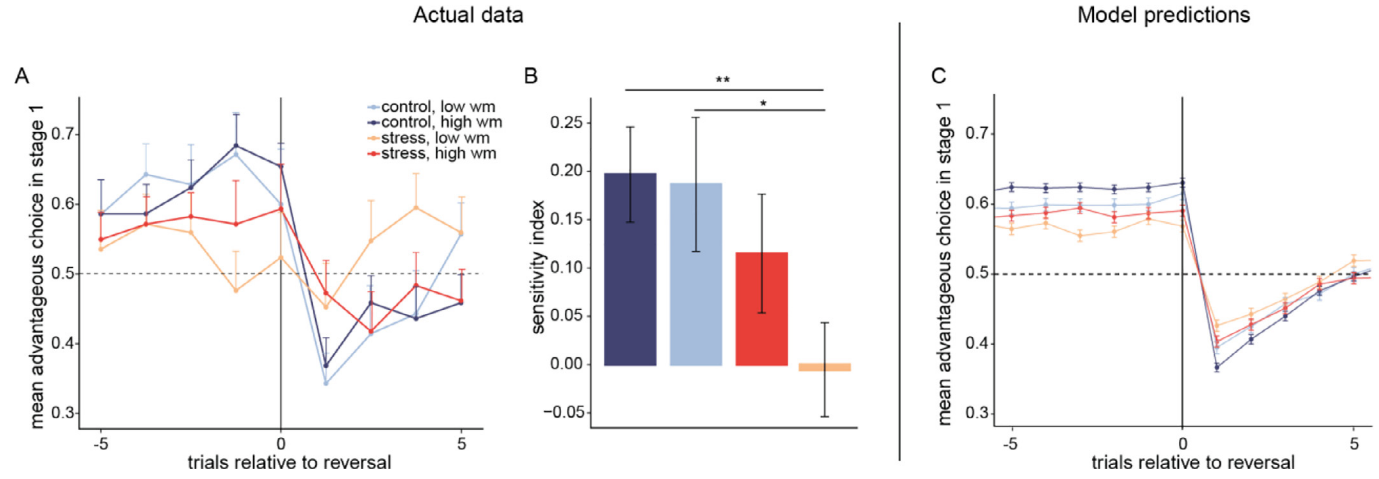


Fig. 8. Stress effects, separately for high and low baseline working memory capacity, as measured with an n-back task. (A) Our data suggest that subjects with low working memory are particularly susceptible to stress effects on flexible learning, yet the interaction between stress and working memory is not statistically significant. (B) The sensitivity index, computed by the mean of advantageous choices before vs. after a reversal, is significantly higher in the control group (high and low working memory) than in the low working memory stress group (t(39) = 2.88, p = 0.006. d = 0.99). (C) Posterior predictive behavior in the trials around the reversal, separately for individuals with high and low working memory capacity in the stress group and in the control group, confirms that the model predictions match the actual behavior, except for a deviation in the stress/low working memory group; $^{***}p < 0.001$, $^{**}p < 0.01$ and $^{*}p < 0.05$ for the comparison between the groups; error bars represent standard errors of the mean.

line working memory does not modulate the impact of stress on learning performance. However, it may also be assumed that a differential susceptibility to stress effects is less modulated by gradual differences in the working memory, but is rather apparent at particularly high or particularly low scores of working memory. Therefore, we tested in a next step whether stress affected the proportion of advantageous first stage choices 4 trials before vs. 4 trials after a reversal differently in high vs. low working memory groups, respectively. High and low working memory groups were defined based on a median split on the n-back performance. The performance of the high- and low working memory participants in the stress and control groups is shown in Fig. 8A. Although Figs. 8A and B suggest that the sensitivity for the change in reward contingencies was particularly affected in stressed participants with low baseline working memory capacity (t(39) = 2.88, p = 0.0065), while the high-working memory stress group and the control group did not differ (t(40) = 1.09, p = 0.282), the respective working memory × stress interaction was not statistically significant (F(52) = 0.89, p = 0.35, η^2_{ges} = 0.01).

Again, we tested whether our model's predictions matched the pattern found in the behavioral data around the reversals and therefore generated 50 simulations for each participant's individual set of parameters. These simulations showed a pattern that strongly resembled the actual data, except for the stress/low working memory group (Fig. 8C). For this group, the correspondence between the simulated and the actual data was lower. In the actual data, the behavior is hardly influenced by the contingency changes, while the simulations show a decrease of the advantageous decisions after a reversal. However, the order of the four groups in the posterior predictive behavior is broadly consistent with the measured data, i.e. also in the simulated data the stress/low working memory group shows the smallest difference from pre- to post-reversal. However, the difference between pre- and post-reversal can still be clearly seen in the simulations, which is not reflected in the actual data. This can be explained by the much smaller sub-sample size in the measured data (data n = {10,12,13,19} vs. 50 in the simulations). On the other hand, this indicates that there are other sources of noise in the measured data that cannot be mapped with the learning model of the Markov decision task.

After analyzing the influence of working memory capacity on the performance in the Markov decision task, we investigated whether working memory was associated with the model parameters. Scores in the n-back task were overall positively correlated with the temperature parameter in stage 1 (r = 0.3, p = 0.03) and tended to be associated with a higher temperature parameter in the second stage (r = 0.26, p = 0.06). Given that the temperature parameter determines to which extent the learned information is used to guide subsequent choices, the observed link to working memory processes is not surprising and might also point to general cognitive capacities that contribute to both flexible learning and working memory. Moreover, high n- back scores tended to be associated with a lower learning rate in the second stage (r = −0.26, p = 0.06, supplemental table S3). However, there were no significant correlations between working memory and model parameters in the stress and control groups (all p > 0.177) and there were no significant main or interaction effects including the factor stress in our working memory × group ANOVA (all F(50) ⟨ 1.78, all p ⟩ 0.19).

4. Discussion

Successful adaptation to dynamic environments is crucial for survival, particularly under highly stressful or threatening conditions. Stress, however, is assumed to impede behavioral flexibility (Otto et al., 2013; Plessow et al., 2011; Raio et al., 2017; Schwabe and Wolf, 2011; Vogel et al., 2016). Here, we sought to shed light on the neurocomputational mechanisms involved in the stress-induced deficit in flexible learning. Our behavioral data show that stress indeed reduced participants' sensitivity to changes in outcome contingencies. In line with these data, our model-based analyses suggest that stress tended to favor rather explorative behavior, as reflected in the tendency of a reduced softmax temperature for the first stage decision. We assume that this is moderated by a reduced utilization of value signals negotiated by model-based and model-free processes. Most importantly, our model-based fMRI analyses revealed that stress reduced the contributions of structures implicated in model-based control and those involved in model-free control of learning.

To tackle specifically the flexibility of learning, we modified the original Markov decision task (Daw et al., 2011) by including several reversals in reward contingencies. This modification increased the task difficulty and made it more demanding to establish a valid model of the task structure, thus favoring, irrespective of stress, model-free over model-based learning. Indeed, although we obtained clear evidence for model-based contributions, model-free elements prevailed during learning. This overall bias towards more model-free learning, reflected in participants' stay probabilities and the weighing parameter w, corroborates recent research suggesting that task complexity facilitates an increased reliance on model-free learning (Kim et al., 2018). The task-related bias towards more model-free learning may explain why we did not observe a further stress-induced shift towards model-free learning that has been suggested before (Park et al., 2017). Accordingly, the proposed bias towards model-free learning associated with the modified task might be considered a limitation, although it is to be noted that participants' choice behavior and the computational modeling parameters provided evidence for both model-based and model-free learning mechanisms. Our behavioral data point to an impairment of flexible learning that is not owing to an altered balance of model-based and model-free processes but rather to a reduced contribution of both model-based and model-free processes to behavior, in contrast to earlier findings suggesting mainly a stress-induced impairment of model-based learning (Otto et al., 2013). The observed impairment seemed to be most pronounced in individuals with low working memory capacity. Although the respective interaction effect did not reach statistical significance, this pattern is generally in line with evidence suggesting that a high working memory capacity may prevent stress effects on model-based learning (Otto et al., 2013).

Although our behavioral data may be interpreted as an indication of impaired flexible learning after stress due to reduced sensitivity for reversals, an alternative view would be that stress encourages more explorative choices at the first stage. More specifically, stressed participants may have learned the stimulus-action-reward-associations in the same way as controls but nevertheless tend not to use this information to guide their behavior. This is indicated by the trend towards a stress-induced reduction of the first stage temperature parameter and further supported by the positive correlation between the first stage sensitivity index and both stage's temperature parameters. At first glance, these findings might seem to be in conflict with previous findings suggesting that stress leads to rather exploitative decisions (Lenow et al., 2017; Luksys and Sandi, 2011). However, these previous studies used classical foraging tasks and such tasks require a different type of decision-making in which the overall environment is used as a proxy for the value of future unknown options, compared to current prospects. Thereby, the focus is on reward calculations which usually determine the switch to a new option below a certain threshold, while the focus in the present task is to maintain probabilistic rules to guide actions. Therefore, another possible explanation is that working memory mediates the explorative choice behavior in the first stage, given that exploration could also be due to an inability to maintain the relevant information to guide upcoming decisions. In line with this idea, performance appeared to be particularly explorative after stress in participants with low working memory performance. Increased explorative behavior in this task can be both advantageous and disadvantageous: it prevents the reliable repetition (exploitation) of a learned contingency but protects against a performance drop when contingencies change. This could explain why the proportion of advantageous decisions did not differ between the groups overall, while there were group differences in the trials around the reversals.

Our data provided initial evidence that stressed participants use value information less for their decision in the first, but not the second stage, as indicated by the softmax temperature parameter and the sensitivity index. This view is further supported by a significantly reduced sensitivity index in the stressed participants with low working memory capacity, given the fact that working memory holds behaviorally relevant information to guide action. The stress-related impairment in first stage choices was accompanied by reduced activity in the ilPFC in the stress group during first stage onset compared to the control group. Thus, the reduced behavioral sensitivity to reversals may be owing to detrimental stress effects on the ilPFC, which has previously been linked to the arbitration between model-based and model-free learning (Lee et al., 2014).

In support of the view that stress interfered with both model-based and model-free control, our imaging findings showed that stress affected the neural underpinnings of both model-free and model-based learning. More specifically, stress reduced the activity associated with model-free prediction errors in the ilPFC. At the same time, the stress group showed an increase in ilPFC activity during model-based learning. The ilPFC has been associated with an arbitrator signal that determines whether behavior is guided by model-based or model-free learning systems (Lee et al., 2014). It is assumed that this arbitrator reduces activity in brain areas implicated in model-free learning when the arbitrator deems that behavior should be guided by the model-based system (Lee et al., 2014). Accordingly, a stress-induced increase in ilPFC activity related to model-based learning processes may be paralleled by a decrease or suppression of the model-free system, as observed here. At the same time, stress decreased activity during model-based prediction error computations in the putamen and posterior hippocampus. In particular the hippocampus has very recently been implicated in model-based planning (Schuck and Niv, 2019; Vikbladh et al., 2019). In fact, the idea of a cognitive map stored in the hippocampus has been proposed already several decades ago (O'Kneefe and Nadel, 1978). For long, however, this idea was limited to spatial references. A recent integrative approach suggests that the hippocampus may also encode cognitive maps that capture complex relationships between cues, actions, results and other characteristics of the environment, enabling flexible, goal-directed decision making (Wikenheiser and Schoenbaum, 2016). Importantly, however, the hippocampus may not as whole be involved in model-based learning. Accumulating evidence from human neuroimaging and rodent lesion studies suggests a functional dissociation within the hippocampus along its anterior (ventral) – posterior (dorsal) axis (Poppenk et al., 2013; Zeidman and Maguire, 2016). In a recent rodent study, specifically the dorsal (posterior) hippocampus was linked to model-based planning behavior (Miller et al., 2017). This finding dovetails with the present data showing that stress reduced specifically the posterior hippocampal activity associated with model-based prediction errors.

These neural changes are most-likely driven by the many hormones and neurotransmitters that are released in response to stressful encounters. Receptors for these stress mediators are abundantly expressed in those regions involved in model-based and model-free learning, in particular, in prefrontal and limbic areas (Herman et al., 2003). Accordingly, it has been shown across tasks and species that these stress mediators, including catecholamines and glucocorticoids, may affect prefrontal and limbic activity and function (Arnsten, 2009; J. J. Kim and Diamond, 2002). Most interestingly with respect to the present findings, it has been shown that glucocorticoids may reduce specifically posterior medial temporal activity during a declarative memory task (de Quervain et al., 2003), exactly that region that was reduced by stress during model-based processing.

Finally, one might argue that the modification of the original Markov decision task impacts the assessment of model-based and model-free processes in our task. While the present task modification, which was required to probe flexible learning in a highly volatile environment, might complicate the direct comparison to studies using the classical Markov decision task to some extent, we assume that also the modified task version allows the assessment of model-based and model-free processes. First, participants' choice behavior and our modeling parameters provided evidence for the involvement of model-based and model-free learning mechanisms. Furthermore, our neuroimaging data revealed neural activity patterns that are well in line with the previously reported neural signatures of model-based and model-free learning, respectively (Daw et al., 2011; Gläscher et al., 2010; Vikbladh et al., 2019). Moreover, the contingency reversals required participants to learn the new transitions, whereas these transitions were assumed to be known by our model. To test whether this affected the performance of our model, we implemented an enhanced model that included state space learning as used in Gläscher et al. (2010) (see Methods for details). Model simulations showed only slightly worse model performance that this enhanced model and our winning model were very similar in their capacity to fit the experimental data. Further analyses also revealed that the transition probabilities in the Markov decision task are learned within the first 10 trials. This supports our original model choice, in which state space learning was omitted.

Together, our data show that stress reduces both model-free and model-based computations during learning in a highly volatile environment. These findings provide novel insights into the neurocomputational mechanisms through which stress hampers the cognitive adaptation to highly volatile environments. A better understanding of these mechanisms may aid the development of new approaches to prevent such stress-induced deficits, with considerable implications, for instance, for educational settings (Vogel and Schwabe, 2016) and stress-related psychopathologies characterized by a deficit in the flexible adaptation to dynamic environments (Koob and Kreek 2007; LaGarde et al., 2010; de Quervain et al. 2017).

Author contributions

L.S. and J.G. conceived and designed the experiment, A.C. performed research, A.C. and F.K. analyzed the data, J.G. fit the computational models, L.S. and J.G. supervised research and analysis, A.C. and L.S. drafted the manuscript, all authors contributed to the manuscript.

Data availability statement

The data that support the findings of this study are openly available on OSF at https://osf.io/tm7ez/?view_only=9d620c297db8461283f01b17a3fa4574.

Declaration of Competing Interest

The authors declare no competing financial interests.

Acknowledgments

This work was supported by the Landesforschungsförderung Hamburg (LFF FV38). We thank Anna Neuhäußer, Theresa Ruder, Leonie el Khafif, Lucie Lilienthal and Rosann Stocker for their assistance with data collection and Carlo Hiller for programming the task.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.117747.

References

Aarts, E., Roelofs, A., van Turennout, M., 2008. Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. J. Neurosci. 28 (18), 4671–4678. doi:10.1523/JNEUROSCI.4400-07.2008.

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. 14 (10), 1338–1344. doi:10.1038/nn.2921.

Arnsten, A.F.T, 2009. Stress signalling pathways that impair prefrontal cortex structure and function. Nat. Rev. Neurosci. 10 (6), 410–422. doi:10.1038/nrn2648.

Balleine, B.W., O'doherty, J.P., 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35 (1), 48.

Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47 (1), 129–141. doi:10.1016/j.neuron.2005.05.020.

Beck, A.T., Ward, C., Mendelson, M., Mock, J., Erbaugh, J., 1961. Beck depression inventory (BDI). Arch. Gen. Psychiatry 4 (6), 561–571.

Braun, S., Hauber, W., 2013. Acute stressor effects on goal-directed action in rats. Learn. Mem. 20 (12), 700–709. doi:10.1101/lm.032987.113.

Brown, J.W., 2009. Conflict effects without conflict in anterior cingulate cortex: multiple response effects and context specific representations. Neuroimage 47 (1), 334–341. doi:10.1016/j.neuroimage.2009.04.034.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76 (1). doi:10.18637/jss.v076.i01.

Collin, S.H.P., Milivojevic, B., Doeller, C.F, 2015. Memory hierarchies map onto the hippocampal long axis in humans. Nat. Neurosci. 18 (11), 1562–1564. doi:10.1038/nn.4138.

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., Rushworth, M.F.S, 2009. Effort-based cost-benefit valuation and the human brain. J. Neurosci. 29 (14), 4531–4541. doi:10.1523/JNEUROSCI.4515-08.2009.

Dandolo, L.C., Schwabe, L., 2018. Time-dependent memory transformation along the hippocampal anterior–posterior axis. Nat. Commun. 9 (1). doi:10.1038/s41467-018-03661-7.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron 69 (6), 1204–1215. doi:10.1016/j.neuron.2011.02.027.

Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8 (12), 1704–1711. doi:10.1038/nn1560.

de Quervain, D., Henke, K., Aerni, A., Treyer, V., McGaugh, J.L., Berthold, T., Nitsch, R.M., Buck, A., Roozendaal, B., Hock, C., 2003. Glucocorticoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe: glucocorticoids impair memory retrieval: a PET-study. Eur. J. Neurosci. 17 (6), 1296–1302. doi:10.1046/j.1460-9568.2003.02542.x.

de Quervain, D., Schwabe, L., Roozendaal, B., 2017. Stress, glucocorticoids and memory: implications for treating fear-related disorders. Nat. Rev. Neurosci. 18 (1), 7–19. doi:10.1038/nrn.2016.155.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19 (2), 430–441. doi:10.1016/S1053-8119(03)00073-9.

Diamond, D.M., Campbell, A.M., Park, C.R., Halonen, J., Zoladz, P.R., 2007. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson Law. Neural Plast. 2007, 1–33. doi:10.1155/2007/60803.

Dolan, R.J., Dayan, P., 2013. Goals and habits in the brain. Neuron 80 (2), 312–325. doi:10.1016/j.neuron.2013.09.007.

Fanselow, M.S., Dong, H.-.W., 2010. Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65 (1), 7–19. doi:10.1016/j.neuron.2009.11.031.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-.G, 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Garvert, M.M., Dolan, R.J., Behrens, T.E., 2017. A Map of Abstract Relational Knowledge in the Human Hippocampal–Entorhinal Cortex, 6. ELife doi:10.7554/eLife.17086.

Gershman, S.J., Uchida, N., 2019. Believing in dopamine. Nat. Rev. Neurosci. 20 (11), 703–714. doi:10.1038/s41583-019-0220-7.

Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66 (4), 585–595. doi:10.1016/j.neuron.2010.04.016.

Goldfarb, E.V., Sinha, R., 2018. Drug-induced glucocorticoids and memory for substance use. Trends Neurosci. 41 (11), 853–868. doi:10.1016/j.tins.2018.08.005.

Goodman, J., Leong, K.-.C., Packard, M.G., 2012. Emotional modulation of multiple memory systems: implications for the neurobiology of post-traumatic stress disorder. Rev. Neurosci. 23, 5–6. doi:10.1515/revneuro-2012-0049.

Gourley, S.L., Swanson, A.M., Jacobs, A.M., Howell, J.L., Mo, M., DiLeone, R.J., Koleske, A.J., Taylor, J.R., 2012. Action control is mediated by prefrontal BDNF and glucocorticoid receptor binding. Proc. Natl. Acad. Sci. 109 (50), 20714–20719. doi:10.1073/pnas.1208342109.

Haruno, M., Kawato, M., 2006. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J. Neurophysiol. 95 (2), 948–959. doi:10.1152/jn.00382.2005.

Herman, J.P., Figueiredo, H., Mueller, N.K., Ulrich-Lai, Y., Ostrander, M.M., Choi, D.C., Cullinan, W.E., 2003. Central mechanisms of stress integration: hierarchical circuitry controlling hypothalamo–pituitary–adrenocortical responsiveness. Front. Neuroendocrinol. 24 (3), 151–180. doi:10.1016/j.yfrne.2003.07.001.

Kim, D., Park, G.Y., O'Doherty, J.P., Lee, S.W, 2018. Task complexity interacts with state-space uncertainty in the arbitration process between model-based and model-free reinforcement-learning at both behavioral and neural levels. BioRxiv doi:10.1101/393983.

Kim, J.J., Diamond, D.M., 2002. The stressed hippocampus, synaptic plasticity and lost memories. Nat. Rev. Neurosci. 3 (6), 453–462. doi:10.1038/nrn849.

Kim, J.J., Lee, H.J., Han, J.-.S., Packard, M.G., 2001. Amygdala is critical for stress-induced modulation of hippocampal long-term potentiation and learning. J. Neurosci. 21 (14), 5222–5228. doi:10.1523/JNEUROSCI.21-14-05222.2001.

Kirchner, W.K., 1958. Age differences in short-term retention of rapidly changing information. J. Exp. Psychol. 55 (4), 352–358. doi:10.1037/h0043688.

Kirschbaum, C., Kudielka, B.M., Gaab, J., Schommer, N.C., & Hellhammer, D.H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis: *psychosomatic medicine*, 61(2), 154–162. 10.1097/00006842-199903000-00006

Kirschbaum, C., Pirke, K.-.M., Hellhammer, D.H., 1993. The 'trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28 (1–2), 76–81. doi:10.1159/000119004.

Koob, G., Kreek, M.J., 2007. Stress, dysregulation of drug reward pathways, and the transition to drug dependence. Am. J. Psychiatry 164 (8), 1149–1159. doi:10.1176/appi.ajp.2007.05030503.

LaGarde, G., Doyon, J., Brunet, A., 2010. Memory and executive dysfunctions associated with acute posttraumatic stress disorder. Psychiatry Res. 177 (1–2), 144–149. doi:10.1016/j.psychres.2009.02.002.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach Atlas labels for functional brain mapping. Hum. Brain Mapp. 10 (3), 120–131 10.1002/1097-0193(200007)10:3<120::AID−HBM30>3.0.CO;2-8.

Lee, S.W., Shimojo, S., O'Doherty, J.P., 2014. Neural computations underlying arbitration between model-based and model-free learning. Neuron 81 (3), 687–699. doi:10.1016/j.neuron.2013.11.028.

Lenow, J.K., Constantino, S.M., Daw, N.D., Phelps, E.A., 2017. Chronic and acute stress promote overexploitation in serial decision making. J. Neurosci. 37 (23), 5681–5689. doi:10.1523/JNEUROSCI.3618-16.2017.

Luksys, G., Sandi, C., 2011. Neural mechanisms and computations underlying stress effects on learning and memory. Curr. Opin. Neurobiol. 21 (3), 502–508. doi:10.1016/j.conb.2011.03.003.

Lupien, S.J., McEwen, B.S., Gunnar, M.R., Heim, C., 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat. Rev. Neurosci. 10 (6), 434–445. doi:10.1038/nrn2639.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19 (3), 1233–1239. doi:10.1016/S1053-8119(03)00169-1.

McClure, S.M., Berns, G.S., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron 38 (2), 339–346. doi:10.1016/S0896-6273(03)00154-5.

Miller, K.J., Botvinick, M.M., Brody, C.D., 2017. Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. 20 (9), 1269–1276. doi:10.1038/nn.4613.

Nasca, C., Bigio, B., Zelli, D., Nicoletti, F., McEwen, B.S., 2015. Mind the gap: glucocorticoids modulate hippocampal glutamate tone underlying individual differences in stress susceptibility. Mol. Psychiatry 20 (6), 755–763. doi:10.1038/mp.2014.96.

Nee, D.E., Kastner, S., Brown, J.W., 2011. Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. Neuroimage 54 (1), 528–540. doi:10.1016/j.neuroimage.2010.08.027.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38 (2), 329–337.

O'Kneefe, J., Nadel, L., 1978. The Hippocampus as a Cognitive Map. Oxford University Press.

Oliveira, F.T.P., McDonald, J.J., & Goodman, D. (2007). *Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action–outcome associations*. 19(12), 11.

Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A., Daw, N.D., 2013. Working-memory capacity protects model-based learning from stress. Proc. Natl. Acad. Sci. 110 (52), 20941–20946. doi:10.1073/pnas.1312011110.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25 (1), 46–59. doi:10.1002/hbm.20131.

Park, H., Lee, D., Chey, J., 2017. Stress enhances model-free reinforcement learning only after negative outcome. PLoS ONE 12 (7), 1–12.

Pfeiffer, B.E., Foster, D.J., 2013. Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497 (7447), 74–79. doi:10.1038/nature12112.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer.

Plessow, F., Fischer, R., Kirschbaum, C., Goschke, T., 2011. Inflexibly focused under stress: acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. J. Cogn. Neurosci. 23 (11), 3218–3227.

Poppenk, J., Evensmoen, H.R., Moscovitch, M., Nadel, L., 2013. Long-axis specialization of the human hippocampus. Trends Cogn. Sci. 17 (5), 230–240. doi:10.1016/j.tics.2013.03.005.

R Core Team, 2019. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing https://www.R-project.org/.

Radenbach, C., Reiter, A.M.F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., Deserno, L., Schlagenhauf, F, 2015. The interaction of acute and chronic stress impairs model-based behavioral control. Psychoneuroendocrinology 53, 268–280. doi:10.1016/j.psyneuen.2014.12.017.

Raio, C.M., Hartley, C.A., Orederu, T.A., Li, J., Phelps, E.A., 2017. Stress attenuates the flexible updating of aversive value. Proc. Natl. Acad. Sci. 201702565.

Rohleder, N., Kirschbaum, C., 2006. The hypothalamic–pituitary–adrenal (HPA) axis in habitual smokers. Int. J. Psychophysiol. 59 (3), 236–243. doi:10.1016/j.ijpsycho.2005.10.012.

Roozendaal, B., McEwen, B.S., Chattarji, S., 2009. Stress, memory and the amygdala. Nat. Rev. Neurosci. 10 (6), 423–433. doi:10.1038/nrn2651.

Schuck, N.W., Niv, Y., 2019. Sequential replay of nonspatial task states in the human hippocampus. Science 364 (6447). doi:10.1126/science.aaw5181, eaaw5181.

Schulz, P., Schlotz, W., 1999. Trierer Inventar zur Erfassung von chronischem Sre (TICS): skalenkonstruktion, teststatistische Überprüfung und Validierung der Skala Arbeitsüberlastung. [The Trier Inventory for the Assessment of Chronic Stress (TICS). Scale construction, statistical testing, and validation of the scale work overload.]. Diagnostica 45 (1), 8–19. doi:10.1026//0012-1924.45.1.8.

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O.T., Oitzl, M.S., 2012a. Stress effects on memory: an update and integration. Neurosci. Biobehav. Rev. 36 (7), 1740–1749. doi:10.1016/j.neubiorev.2011.07.002.

Schwabe, L., Tegenthoff, M., Hoffken, O., Wolf, O.T., 2012b. Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. J. Neurosci. 32 (30), 10146–10155. doi:10.1523/JNEUROSCI.1304-12.2012.

Schwabe, L., Wolf, O.T., 2009. Stress prompts habit behavior in humans. J. Neurosci. 29 (22), 7191–7198. doi:10.1523/JNEUROSCI.0979-09.2009.

Schwabe, L., Wolf, O.T., 2012. Stress modulates the engagement of multiple memory systems in classification learning. J. Neurosci. 32 (32), 11042–11049. doi:10.1523/JNEUROSCI.1484-12.2012.

Schwabe, L., Schächinger, H., de Kloet, E.R., Oitzl, M.S., 2010. Corticosteroids operate as a switch between memory systems. J. Cogn. Neurosci. 22 (7), 1362–1372. doi:10.1162/jocn.2009.21278.

Schwabe, L., Tegenthoff, M., Höffken, O., Wolf, O.T., 2013. Mineralocorticoid receptor blockade prevents stress-induced modulation of multiple memory systems in the human brain. Biol. Psychiatry 74 (11), 801–808. doi:10.1016/j.biopsych.2013.06.001.

Schwabe, L., Wolf, O.T., 2011. Stress-induced modulation of instrumental behavior: from goal-directed to habitual control of action. Behav. Brain Res. 219 (2), 321–328. doi:10.1016/j.bbr.2010.12.038.

Schwabe, L., Wolf, O.T., 2013. Stress and multiple memory systems: from 'thinking' to 'doing.'. Trends Cogn. Sci. 17 (2), 60–68. doi:10.1016/j.tics.2012.12.001.

Sloman, S.A., 1996. The empirical case for two systems of reasoning. Psychol. Bull. 119 (1), 3–22.

Spielberger, C.D., Gorsuch, R.L., Lushene, R.E., 1970. STAI Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press.

Stachenfeld, K.L., Botvinick, M.M., Gershman, S.J., 2017. The Hippocampus as a Predictive Map. BioRxiv doi:10.1101/097170.

Strange, B.A., Witter, M.P., Lein, E.S., Moser, E.I., 2014. Functional organization of the hippocampal longitudinal axis. Nat. Rev. Neurosci. 15, 655–669.

Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An introduction. MIT Press.

Vikbladh, O.M., Meager, M.R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., Daw, N.D., 2019. Hippocampal contributions to model-based planning and spatial memory. Neuron 102 (3), 683–693. doi:10.1016/j.neuron.2019.02.014.

Vogel, S., Fernández, G., Joëls, M., Schwabe, L., 2016. Cognitive adaptation under stress: a case for the mineralocorticoid receptor. Trends Cogn. Sci. 20 (3), 192–203. doi:10.1016/j.tics.2015.12.003.

Vogel, S., Schwabe, L., 2016. Learning and memory under stress: implications for the classroom. Npj Sci. Learn. (1) 1. doi:10.1038/npjscilearn.2016.11.

Watanabe, S., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11, 3571–3594.

Weger, M., Sandi, C., 2018. High anxiety trait: a vulnerable phenotype for stress-induced depression. Neurosci. Biobehav. Rev. 87, 27–37. doi:10.1016/j.neubiorev.2018.01.012.

Wikenheiser, A.M., Schoenbaum, G., 2016. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. Nat. Rev. Neurosci. 17 (8), 513–523. doi:10.1038/nrn.2016.56.

Wirz, L., Bogdanov, M., Schwabe, L., 2018. Habits under stress: mechanistic insights across different types of learning. Curr. Opin. Behav. Sci 20, 9–16. doi:10.1016/j.cobeha.2017.08.009.

Zarr, N., Brown, J.W., 2016. Hierarchical error representation in medial prefrontal cortex. Neuroimage 124, 238–247. doi:10.1016/j.neuroimage.2015.08.063.

Zeidman, P., Maguire, E.A., 2016. Anterior hippocampus: the anatomy of perception, imagination and episodic memory. Nat. Rev. Neurosci. 17 (3), 173–182. doi:10.1038/nrn.2015.24.