# Experimentally induced and real-world anxiety have no demonstrable effect on goal-directed behaviour

C. M. Gillan[1,2,*], M. M. Vaghi[3,*,†] (ID), F. H. Hezemans[3], S. van Ghesel Grothe[4], J. Dafflon[5], A. B. Brühl[6], G. Savulich[3] and T. W. Robbins[3]

[1]Trinity College Dublin, Dublin, Ireland; [2]New York University, New York, USA; [3]University of Cambridge, Cambridge, UK; [4]University of Amsterdam, Amsterdam, Netherlands; [5]Kings College London, London, UK and [6]University Hospital of Psychiatry Zurich, Zurich, Switzerland

## Original Article

*Denotes equal first author contribution.

†Present address: Max Planck UCL Centre for Computational Psychiatry and Ageing Research, 10–12 Russell Square, London, United Kingdom.

**Author for correspondence:**
M. M. Vaghi,
E-mail: matilde.vaghi@gmail.com

## Abstract

**Background.** Goal-directed control guides optimal decision-making and it is an important cognitive faculty that protects against developing habits. Previous studies have found some evidence of goal-directed deficits when healthy individuals are stressed, and in psychiatric conditions characterised by compulsive behaviours and anxiety. Here, we tested if goal-directed control is affected by state anxiety, which might explain the former results.
**Methods.** We carried out a causal test of this hypothesis in two experiments (between-subject $N = 88$; within-subject $N = 50$) that used the inhalation of hypercapnic gas (7.5% $CO_2$) to induce an acute state of anxiety in healthy volunteers. In a third experiment (N = 1413), we used a correlational design to test if real-life anxiety-provoking events (panic attacks, stressful events) are associated with impaired goal-directed control.
**Results.** In the former two causal experiments, we induced a profoundly anxious state, both physiologically and psychologically, but this did not affect goal-directed performance. In the third, correlational, study, we found no evidence for an association between goal-directed control, panic attacks or stressful life eventsover and above variance accounted for by trait differences in compulsivity.
**Conclusions.** In sum, three complementary experiments found no evidence that anxiety impairs goal-directed control in human subjects.

## Background

Two well-established systems contribute to everyday decision making and behaviour, the goal-directed and the habitual system (Dickinson, 1985). Goal-directed behaviour is characterised by actions that are appropriate to the current desire for a given outcome and informed by the knowledge of the causal relationship between an action and the associated outcome (Dickinson & Balleine, 1994). More recently goal-directed control has been formalised as model-based planning, within a reinforcement learning framework (Daw, Gershman, Seymour, Dayan, & Dolan, 2011).

Though no previous study has examined whether experimentally induced state anxiety impairs goal-directed planning, a related literature on stress-induction offers a basis for this suggestion. Specifically, acute stress has been shown to induce deficits in goal-directed planning (Park, Lee, & Chey, 2017; Schwabe & Wolf, 2009, 2010), albeit inconsistently (null results: Heller, Ezie, Otto, & Timpano, 2018; Otto, Raio, Chiang, Phelps, & Daw, 2013; Radenbach et al. 2015) in healthy individuals. Acute anxiety and stress manipulations produce similar cardiovascular changes, and induce negative affect, but anxiety induction differs from stress in terms of the specific psychological experience (e.g. increased vigilance, panic, fear) and other aspects of the physiological response (Bailey, Argyropoulos, Kendrick, & Nutt, 2005; Shin & Liberzon, 2010).

Physiological and psychological stress has been likened to anxiety, and it is generally thought to impair several forms of deliberative and reflective processes, in favour of more automatic and reflexive ones (Shields, Sazma, & Yonelinas, 2016). From a neurobiological perspective, there is evidence that this mechanism is regulated by catecholamines, which act on prefrontal functioning under stress (Arnsten 1998). It has been suggested that reliance on faster, habitual mechanisms might be an evolutionary advantage in stressful situations (Arnsten 1998). Similarly, in the case of anxiety, the attentional control theory (Eysenck et al., 2007) suggests that anxiety impairs cognitive performance of top-down, executive tasks by giving greater influence to the bottom-up attentional system.

In addition, anxiety is a prominent feature of pathological manifestations characterised by an impoverished goal-directed system. For example, a fragile goal-directed system is hypothesised to lead one to get stuck in habits (Gillan, Otto, Phelps, & Daw, 2015) and typifies not

only Obsessive-Compulsive Disorder (OCD) (Gillan et al., 2011; Gillan & Robbins, 2014; Vaghi et al., 2018) but also several other psychiatric conditions on the compulsivity spectrum such as eating disorder, drug abuse and alcohol addiction (Sjoerds et al., 2013; Voon et al., 2014). Accordingly, it has been suggested that goal-directed deficits constitute a trans-diagnostic trait (Gillan, Kosinski, Whelan, Phelps, & Daw, 2016; Robbins, Gillan, Smith, de Wit, & Ersche, 2012). One potential issue with this model is its specificity. Compulsivity is highly comorbid with anxiety (Nestadt et al., 2009), which is unsurprising, as OCD has only recently moved out of the Diagnostic and Statistical Manual category of anxiety disorders into its own classification (Stein et al., 2010). Accordingly, this raises the possibility that elevated anxiety levels in OCD might account for failures in goal-directed planning and consequent overreliance on habits.

In support of this idea, social anxiety patients appear to show similar deficits in goal-directed planning to OCD patients, despite the fact that they do not have a compulsive phenotype (Alvares, Balleine, & Guastella, 2014). Cross-sectional, correlational work has started to address this issue, finding that when a range of psychopathology measures are taken (and controlled for) within the same individuals, there is no meaningful contribution of trait anxiety to goal-directed deficits, while the association with compulsivity is robust (Gillan et al., 2016; Robbins et al., 2012). However, these studies are limited not just by their correlational nature, but because they assess trait anxiety, which does not speak to acute states of anxiety that are experienced by patients more transiently, often in response to their own obsessive and compulsive symptoms (Mataix-Cols et al., 2003).

Here, we aimed to characterise the relationship between increased anxiety and the functioning of the goal-directed system. We used a combination of causal and correlational approaches to investigate this in three experiments spanning laboratory and real-life settings.

Firstly, we used hypercapnic gas (i.e. with increased $CO_2$ level) to experimentally induce state anxiety and test its impact on goal-directed control, operationalised as sensitivity to contingency degradation (Vaghi et al., 2018). Hypercapnic gas is a well-validated method for experimentally inducing a transitory state of acute anxiety in healthy volunteers (Woods, Charney, Goodman, & Heninger, 1988). At very high doses (35% $CO_2$) it generates symptoms similar to those of panic disorder, with increased blood pressure and bradycardia (Argyropoulos et al., 2002; Griez, Zandbergen, Pols, & de Loof, 1990; Perna, Barbini, Cocchi, Bertani, & Gasperini, 1995), especially in subjects with panic disorder or susceptibility to it (Perna, Bertani, Caldirola, & Bellodi, 1996; Perna et al., 1994). We used lower doses (7.5% $CO_2$) which are reported to be sufficient to induce physiological and psychological symptoms of anxiety and sustained arousal associated with an anxiety state (Bailey et al., 2005). Subjects had profound physiological and subjective psychological responses to the anxiety induction procedure including changes in the heart rate, blood pressure and self-reported anxiety, but it failed to induce deficits in goal-directed control over behaviour.

To attempt to rule out the possibility that this null effect was an issue of the sensitivity of our study design, we repeated this experiment using a within-subjects design and a more commonly used, and potentially more sensitive, measure of goal-directed control – a 'model-based planning' measure derived from the two-step reinforcement learning task described above (Daw et al., 2011). Again, the $CO_2$ manipulation resulted in substantial physiological and psychological effects consistent with the

induction of an acute state of anxiety, but this had no demonstrable detrimental effect on goal-directed behaviour.

In a third and final experiment, we tested this hypothesis in a naturalistic, real-world setting using a large-scale correlational design ($N = 1413$) (Gillan et al., 2016). We investigated if goal-directed (model-based) control is impaired in individuals who suffered recent 'real life' acute anxiety, specifically known to be associated with the experience of a recent panic attack (Aronson & Logue, 1988) and/or major life-stressors (Vyas, Pillai, & Chattarji, 2004). We found that the frequency of panic attacks in the past week and higher levels of stress in the past year were both modestly associated with deficits in goal-directed planning. Crucially, neither survived controlling for a correlated psychiatric trait, compulsive behaviour and intrusive thought, which we previously showed have a strong association with goal-directed planning using these same data (Gillan et al., 2016).

## Methods

### Experiment 1

#### Subjects

A total of 88 participants was recruited through university mailing lists, departmental research panels and posted flyers within the University of Cambridge and the wider community. Participants were randomly assigned to either the $CO_2$-induced anxiety group ($n = 43$, 20 females; mean age = 27.55, S.D. = 11.04) or the normal air 'placebo' group ($n = 45$, 24 females; mean age = 27.40, S.D. = 10.03) (online Supplementary Material for further details on recruitment and inclusion and exclusion criteria).

#### Anxiety manipulation

The anxiety induction consisted of the inhalation of air enriched with 7.5% $CO_2$ (7.5% $CO_2$, 20% $O_2$, 71.5% $N_2$, pre-mixed, BOC Special Gases, Guildford, UK). As the experimenter had to manually switch a lever to activate the delivery of one of the two air preparations, $CO_2$ was administered in a single-blind manner while measuring goal-directed/habit behaviour via controlled tasks, and was designed to induce a physiological state of acute anxiety in a reliable and controlled manner (Bailey et al., 2005). Participants inhaled the assigned air preparation as long as they were doing the task. To measure the effectiveness of this procedure at inducing acute anxiety, we recorded physiological measurements comprising heart rate, diastolic and systolic blood pressure and psychological measurements comprising the 17-item Acute Panic Inventory (Liebowitz et al., 1984), 10-item Positive and Negative Affective Scale (Watson, Clark, & Tellegen, 1988) and three Visual Analogue Scales assessing anxiety, fear and happiness. Physiological measures were collected 10 min before, during and 15 min after the experimental manipulation. Psychological measures of subjective feeling due to the experimental manipulation were concomitantly collected, the only difference being that they were not interrogated during the performance of the task but immediately after and retrospectively on how they were feeling.

#### Contingency degradation paradigm

In this between-subjects design, subjects in each group performed a contingency degradation task described previously and further detailed in the online Supplementary Material (Vaghi et al., 2018) (Fig. 1a). In short, the task was a free operant, self-paced procedure which tests subjects' ability to detect action-outcome instrumental contingencies (Vaghi et al., 2018), one of the earliest
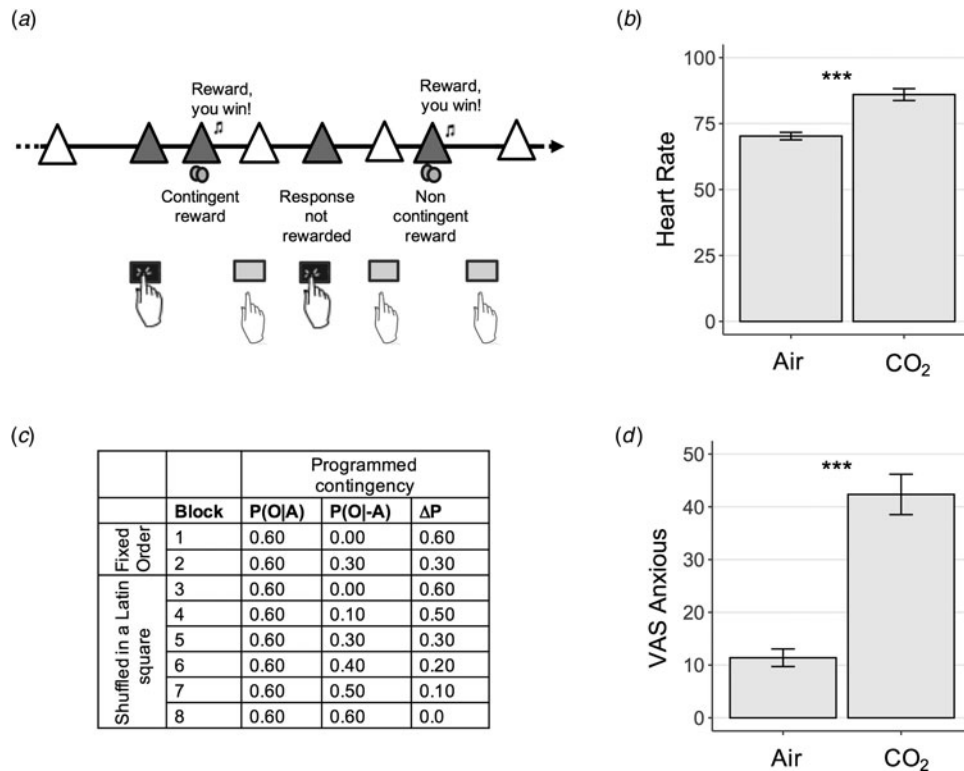
**Fig. 1.** Experiment 1 Study Design – Contingency Degradation Task. (*a*) Contingency degradation task design. In each block, subjects were presented with a white triangle, signalling that they had the opportunity to press or to not press the space bar, in a free-operant, self-paced procedure (Vaghi et al., 2018). The triangle turned yellow (here pictured in grey) when a response was recorded. Rewards (a 25 pence image) were delivered according to a probability of outcome given action, P(O|A), on trials when a response was made, and a probability of outcome given no action, P(O|-A), when a response was not made. (*b*) Physiological response to anxiety induction. Subjects' heart rate was elevated significantly during the gas condition, $p < 0.001$. Error bars represent s.e.. (*c*) Programmed contingencies. Each participant completed eight blocks where contingency was systematically varied through changes to P(O|-A). The first two blocks were considered training blocks and appeared in a fixed order as denoted in the table. The six remaining test blocks were presented in a counterbalanced order across subjects. (*d*) Psychological response to anxiety induction. Anxiety scores measures using a visual analogue scale (VAS) were also significantly elevated during the inhalation of gas compared with air, $p < 0.001$. Error bars represent s.e. ***, $p < 0.001$.

operationalisations of goal-directed learning from the animal literature (Dickinson, Nicholas, & Adams, 1983).

### Experienced contingency

As expected, for normal and $CO_2$-enriched air condition, experienced contingencies (based on experienced event frequencies, see online Supplementary Material, Table S1) matched the a priori programmed ones ($CO_2$: $r = 1.00$, $p < 0.001$; air: $r = 1.00$, $p < 0.001$). Therefore, programmed contingencies were used for subsequent analysis. Our findings were not confounded by between-group differences in experienced contingencies, as no main effect of group ($F_{(1, 63)} = 0.80$, $p = 0.37$, $\eta_G^2 = 0.003$) nor interaction between group and block ($F_{(2.57, 161.61)} = 0.17$, $p = 0.89$, $\eta_G^2 = 0.002$) was found.

### Data analysis

We first performed analyses of variance (ANOVA) to determine whether there was a between-group difference in sensitivity to instrumental contingency as measured by the response rate and causality judgment. The response rate was computed by dividing the number of bins for which a response was made by the total number of bins within each block. For each dependent variable, programmed contingency was used as a within-subject factor, and group was used as a between-subject factor. Analyses were conducted separately for the initial learning blocks and the test blocks. For the test blocks, we also investigated the relationship between the response rate and contingency judgments, using a linear mixed-effects model. Specifically, we used contingency judgement and group as fixed effects, and we allowed the intercept and slope to vary between participants as random effects. We obtained $p$ values for the fixed effects using the Kenward–Roger method. Bayes factor analysis was used in case of failure to reject the null hypothesis, to examine the relative evidence for the null with default JZS priors for ANOVA (Rouder, Speckman, Sun, Morey, and Iverson, 2009) and (Rouder, Morey, Speckman, & Province, 2012). Previous research (Schwabe, Tegenthoff, Höffken, & Wolf, 2010) found a between-subjects effect size of stress on habitual performance for which default JSZ priors are suitable as specified in Rouder et al., 2009 and Rouder et al., 2012. Analyses were performed in R version 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org/) using the 'afex' package for ANOVA and linear mixed models, the 'Bayes Factor' and 'brms' package for Bayes factor analysis and the 'tidyverse' packages for data organisation and visualisation.

## Experiment 2

### Subjects

A total of 61 healthy volunteers was recruited from the local community in the same manner as described in Experiment 1. Screening and exclusion criteria were identical to Experiment 1.

Further exclusion criteria were applied contingent on the experimental task employed here (online Supplementary Material). The final sample size for analysis was 50 (26 female) with ages ranging from 18–62.

### Reinforcement learning task

Participants completed a reinforcement-learning task that quantifies individual differences in goal-directed ('model-based') learning, which is operationalised as a parameter estimate from a logistic regression analysis predicting choices in the task (Daw, Niv, & Dayan, 2005). The task (Fig. 3a) has been extensively used and described elsewhere (37) and further detailed in the online Supplementary Material.

### Anxiety induction

The anxiety induction procedure as well as collection of physiological and psychological measures was identical to Experiment 1, except for the within-subjects design. Participants attended a single test session during which they completed two versions of the Reinforcement Learning Task during 20 min inhalation of air enriched with 7.5% $CO_2$ and normal air. Gas was administered in a single-blind manner and the order of $CO_2$ v. normal air was counterbalanced.

### Data analysis

Data were analysed using mixed-effects logistic regression in the *lme4* package in R 3.5.1 (http://cran.us.r-project.org). In line with previous studies (Daw et al., 2011), we tested the extent to which subjects tend to repeat actions performed on the previous trial or explore a new one ('Stay': coded switch = 0; stay = 1), and whether these choices were influenced by whether or not their previous action was rewarded ('Reward': coded as rewarded = 1; unrewarded = −1), was followed by a rare or common transition ('Transition': coded as common = 1, rare = −1) and their interaction ('Reward × Transition'). The intercept reflects tendencies to repeat the same action from one trial to the next, the main effect of reward reflects the contribution of model-free learning to subjects' choices, while an interaction between reward and transition is the hallmark of model-based (goal-directed) behaviour. We included anxiety induction as a within-subjects factor (coded $CO_2 = 1$, Air = −1). We used Bound Optimization by Quadratic Approximation with 1e5 functional evaluations. The model was specified as follows: Stay ~ Reward × Transition × $CO_2$ + (Reward × Transition × $CO_2$ + 1|Subject). Bayes factor analysis was used in case of failure to reject the null hypothesis using the anovaBF function in the BayesFactor package in R, with default JZS priors for ANOVA from (Rouder et al., 2012). To avoid the issues with nested interactions from the logistic model, we extracted estimates for model-based planning separately for each subject in each condition and used these to compare an ANOVA model with a within-subjects effect of gas to an intercept-only model.

### Computational modelling

A more elaborated form of this analysis is presented in the online supplement. In brief, this method allows for analysis of a greater number of potential behavioural confounds, including separating the distinct role of the learning rate and choice randomness from that of model-based, model-free and choice repetition estimates from the simpler analysis. These results largely recapitulate the main findings of the paper, with slight differences flagged as appropriate.

## Experiment 3

### Participants

Data were collected online using Amazon's Mechanical Turk. Details of the experimental procedure can be found elsewhere (Gillan et al., 2016), but in brief, data were analysed from 1413 individuals (823 female) with ages ranging from 18 to 76 (M = 33, S.D. = 11), who were based in the USA, had a history of good performance on the platform (i.e. were paid in full on at least 95% of their previous tasks), passed a comprehension test, and did not fall victim to a 'catch' question (online Supplementary Material).

### Reinforcement learning task

The task employed in this study was the same as that described in Experiment 2. The only difference was that subjects completed it remotely, and therefore a more rigorous quality control procedure was implemented (detailed in Supplement).

### Panic attacks and life stress

The occurrence of recent panic attacks was assessed using item 1 on the self-report version of the Panic Disorder Severity Scale (PDSS; Shear et al., 1997). Life stress was assessed using the Social Readjustment Scale (Holmes & Rahe, 1967), which presents an inventory of common stressful life events to participants and asks them to select those that applied to them in the previous 12 months. The present sample had a mean score of 159 (S.D. = 120). Scores lower than 150 are considered evidence of 'no significant stress' (N = 775), while scores in excess of 300 are considered signs of major stress (N = 179 in this sample) (Fig. 5b) (see also online Supplementary Material). Control variables were also included as detailed in the online Supplementary Material.

### Data analysis

We performed the same analysis as in Experiment 2, but here we additionally controlled for variables that have been previously linked to model-based planning, namely: IQ, age, gender and a trans-diagnostic psychiatric trait 'Compulsive Behaviour and Intrusive Thought'. This covariate was derived from previous published work (Gillan et al., 2016; Rouault, Seow, Gillan, & Fleming, 2018) that applied factor analysis to a series of questionnaires linked to self-reported measures of psychopathology. Factors were labelled based on items that loaded most strongly on each of the identified factors. Accordingly, items from questionnaires related to 'compulsive' disorders such as OCD and addiction most strongly loaded on the factor named 'Compulsive Behaviour and Intrusive thought'. Scores of each subject on this factor were used as a covariate in the present analysis to rule out the possibility that any association between stress, panic attacks and goal-directed control could be better explained by compulsivity. Bayes factor analysis was conducted on a linear model where residuals for model-based planning was the dependent measure and life stress or panic symptoms were the experimental models compared to an intercept-only model. As in experiment 2, we complemented our regression analysis with a computational model, details of which are available in the online supplement.

## Results

### Anxiety induction and contingency degradation (Experiment 1)

Here we tested if experimentally induced anxiety would affect subjects' ability to detect action-outcome instrumental
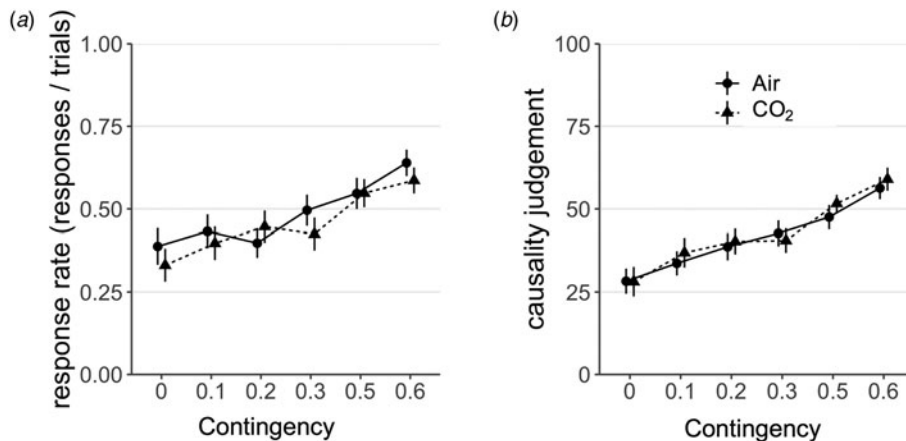
**Fig. 2.** Results from Experiment 1. (*a*) There was no effect of $CO_2$-induced anxiety on subjects' sensitivity to instrumental contingency as measured by choice responses, $F_{(3.73, 320.59)} = 1.74$, $p = 0.15$. Error bars represent S.E. (*b*). There was similarly no effect of group on the extent to which causality judgements scaled with instrumental contingency, $F_{(2.99, 256.89)} = 0.33$, $p = 0.81$. Error bars represent S.E.

contingencies. In a between-subjects design, one group was assigned to inhale hypercapnic gas (7.5% $CO_2$) during the performance on the contingency degradation task, while the other inhaled normal air. Psychological and physiological measures confirmed that anxiety induction was successful and of a magnitude similar to that observed in prior studies (Cooper et al., 2013; Garner, Attwood, Baldwin, & Munafò, 2012; Garner, Attwood, Baldwin, James, & Munafò, 2011): participants in the $CO_2$ condition experienced greater self-reported anxiety ($F_{(1.97, 159.61)} = 35.57$, $p < 0.001$) and had a higher heart rate ($F_{(1.96, 152.92)} = 36.64$, $p < 0.001$) than those assigned to the air condition (Fig. 1*b* and *d*; online Supplementary materials).

Participants learnt the contingencies in the training phase ($F_{(1, 86)} = 26.48$, $p < 0.001$, $\eta_G^2 = 0.03$). Experimentally induced anxiety did not affect subjects' behavioural sensitivity to instrumental contingency. Participants overall adjusted their response rate in line with the underlying contingency, as evidenced by a main effect of contingency on the response rate in the test blocks ($F_{(3.73, 320.59)} = 29.95$, $p < 0.001$, $\eta_G^2 = 0.07$). In the test blocks, there was no between-group difference ($F_{(1, 86)} = 0.22$, $p = 0.64$, $\eta_G^2 = 0.002$) and no group by contingency interaction ($F_{(3.73, 320.59)} = 1.74$, $p = 0.15$, $\eta_G^2 = 0.004$) (Fig. 2*a*). Bayes factor analysis further confirmed these findings. Specifically, the null model was strongly preferred over the alternative model with the main effect of anxiety and interaction effect of anxiety by contingency (BF$_{01} = 16.81$ and online Supplementary Fig. S1*a*).

The same was true of participants' *subjective* assessments of instrumental contingency (i.e. their explicit model of the environment). Subjects accurately tracked the underlying contingency of the task (training blocks, $F_{(1, 86)} = 30.46$, $p < 0.001$, $\eta_G^2 = 0.12$; test blocks, $F_{(2.99, 256.89)} = 26.22$, $p < 0.001$, $\eta_G^2 = 0.13$) and the experimental manipulation did not affect this. There was no between-group difference ($F_{(1, 86)} = 0.16$, $p = 0.69$, $\eta_G^2 = 0.001$) and no group by contingency interaction ($F_{(2.99, 256.89)} = 0.33$, $p = 0.81$, $\eta_G^2 = 0.002$) (Fig. 2*b*) on causality judgements. Bayes factor analysis further confirmed these findings. Specifically, the null model was strongly preferred over the alternative model with a main effect of anxiety and interaction effect of anxiety by contingency (BF$_{01} = 386.15$ and online Supplementary Fig. S1*b*). Mirroring the findings on choice responses, experimentally induced anxiety did not affect subjective judgments of instrumental contingency – adding weight to the suggestion that state

anxiety may not have an appreciable effect on goal-directed control over action.

### Individual differences

Prior work showed that individual differences might be important in revealing the effect of stress on goal-directed behaviour (Heller et al., 2018; Otto et al., 2013; Radenbach et al., 2015; Schwabe & Wolf, 2010). Therefore, we tested if the detrimental effect of $CO_2$ on goal-directed behaviour might depend on individual differences in sensitivity to the $CO_2$ manipulation, assessed in terms of change in both self-reported and physiological measures of anxiety. For the former, we ran the model explained above with programmed contingency as a within-subject factor, introducing a change in self-report anxiety as a between-subject covariate. The change in self-report anxiety was computed as the difference between VAS-anxious before inhaling the gas and after inhaling the gas. As above, there was a significant effect of programmed contingency on the response rate ($F_{(3.73, 309.85)} = 25.42$, $p < 0.001$), but there was no main effect of subjectively reported change in self-report anxiety ($F_{(1, 83)} = 0.28$, $p = 0.60$) nor an interaction effect with programmed contingency ($F_{(3.73, 309.85)} = 0.20$, $p = 0.42$). Similar findings were obtained on subjective causality ratings. Accordingly, programmed contingency significantly predicted causality ratings ($F_{(3.18, 264.24)} = 33.10$, $p < 0.001$), but there was not a main effect ($F_{(1, 83)} = 0.00$, $p = 0.96$) nor a significant interaction with subjectively reported change in self-report anxiety ($F_{(3.18, 264.24)} = 0.20$, $p = 0.90$). Therefore, individual differences in anxiety, as self-reported by subjects upon $CO_2$ challenge, did not affect goal-directed planning.

We conducted the same analyses by using physiological changes in the heart rate as a putatively more objective measure of the change in anxiety arising from our manipulation. The physiological index for change in the heart rate was computed as above, i.e. the difference between the heart rate before inhaling the gas and after inhaling the gas. Changes in the heart rate did not have a main effect on the response rate ($F_{(1, 80)} = 0.1$, $p = 0.75$), but there was a trend for an interaction between changes in the heart rate and programmed contingency ($F_{(3.80, 303.94)} = 2.00$, $p = 0.10$). Individuals with higher changes in the heart rate tended to show slightly *greater* sensitivity to instrumental contingency, as their response rate depended more strongly on programmed contingency. Thus, if any moderating effect of anxiety sensitivity exists,

**Fig. 3.** Experiment 2 Study Design – Model-Based Learning Task. (*a*) On each trial, subjects chose between two fractals, which probabilistically transition to either an orange or blue state (pictured here in greyscale) where they must make another choice. In this schematic, the fractal on the left had a 70% chance of transitioning to the blue state, what is called a 'common' transition, and a 30% chance of transitioning to the orange state, i.e. a 'rare' transition. In the second orange or blue state, subjects again chose between two fractals, each of which was associated with a probability of reward (a pound coin). Unlike the transition structure, these reward probabilities drifted slowly over time ($0.25 < p < 0.75$). This meant that subjects were required to dynamically track which of the fractals in the orange and blue states were currently best. The reward probabilities depicted (34%, 68%, 72% and 67%) refer to the probability of reward for each of the 4 options presented in an example trial at a certain point along the reward probability drifts. Model-based planning on this task is operationalised as the extent to which subjects' decision to repeat an action at the first stage, depend on (i) whether this action was rewarded on the previous trial and (ii) and whether the path from action to outcome was expected ('common'). (*b*) Physiological response to anxiety induction. Heart rate was elevated significantly during the gas condition, $F_{(1,49)} = 10.72$, $p = 0.002$. Error bars represent S.E. (*c*) Psychological response to anxiety induction. Self-reported anxiety levels were also significantly elevated during the inhalation of gas compared with air, $F_{(1,49)} = 57.47$, $p < 0.001$. Error bars represent S.E. ***, $p < 0.001$.
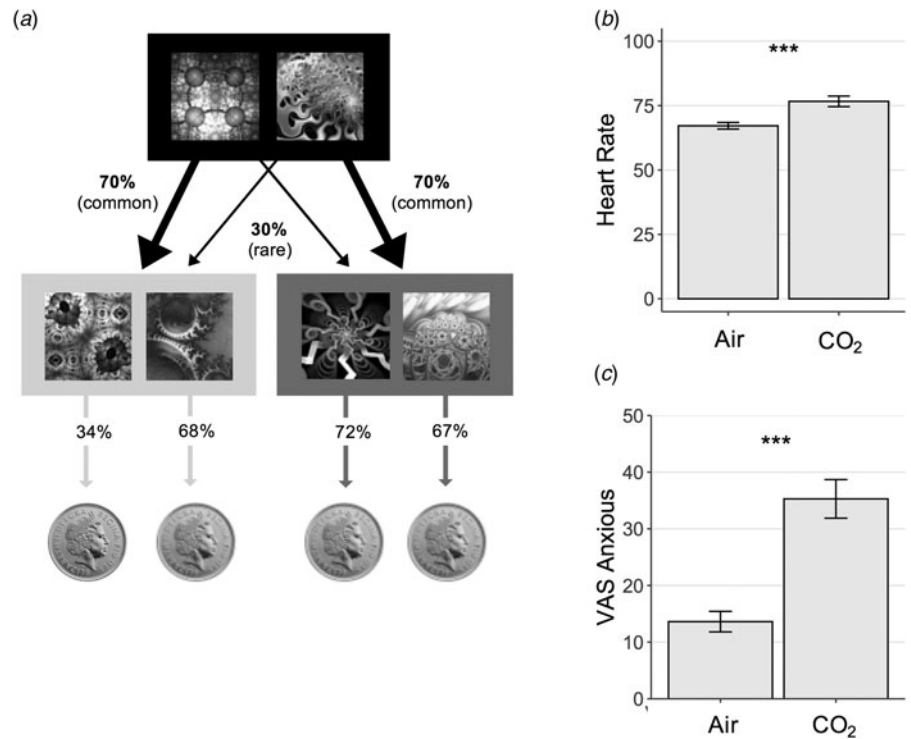
it goes in the opposite direction to what has been shown in individual difference research with stress and goal-directed control (e.g. Otto et al., 2013; Radenbach et al., 2015). Changes in the heart rate did not affect subjective causality ratings ($F_{(1, 80)} = 0.08$, $p = 0.78$). Similarly there was not a significant interaction between changes in the heart rate and programmed contingency ($F_{(3,29, 263.45)} = 0.44$, $p = 0.74$) in predicting subjective causality ratings.

### *Anxiety induction and model-based planning (Experiment 2)*

We adopted a complementary approach to Experiment 1 to test if anxiety induction would affect goal-directed planning. We employed a 'model-based' learning task (Fig. 3) (Daw et al., 2005, 2011) in the context of a within-subjects design. The task is more commonly used in the literature and findings consistent with those in Experiment 1 would thus speak to the generalisability and validity of the findings. The within-subjects design overcomes the potential problem that individual differences in goal-directed control [e.g. associated with compulsiveness, IQ, age (Gillan et al., 2016)] may have hindered our ability to detect changes resulting from anxiety-induction in Experiment 1.

As in Experiment 1, the $CO_2$ manipulation was effective in inducing anxiety in subjects (Fig. 3*b* and *c*), with a significant increase in self-reported anxiety $F_{(1,49)} = 57.47$, $p < 0.001$ and heart-rate, $F_{(1,49)} = 10.72$, $p = 0.002$. However, consistent with the results of Experiment 1, this did not alter goal-directed performance; $CO_2$ had no effect on model-based planning ($\beta = -0.03$, S.E. $= 0.04$, $p = 0.44$). Bayes factor analysis indicated that there was moderate evidence in favour of the null model over the alternative model that included the acute anxiety manipulation ($BF_{01} = 3.5$).

The regression model overall fit subjects' behaviour as expected; 'model-free' behaviour was evident in the sample ($\beta = 0.55$, S.E. = 0.08, $p < 0.001$) which refers to how much subjects tend to repeat actions that were recently rewarded. Model-based learning was also overall significant ($\beta = 0.28$, S.E. = .06, $p < 0.001$), such that subjects took environmental contingency into account when deciding whether or not to repeat a rewarded choice. Finally, subjects showed an overall biased tendency to repeat choices from one trial to the next, regardless of reward or transition information ($\beta = 1.59$, S.E. = 0.12, $p < 0.001$). Much like model-based learning, there was no effect of anxiety on model-free learning ($\beta = -0.02$, SE = 0.03, $p = 0.52$), or action repetition ($\beta = -0.08$, S.E. = 0.04, $p = 0.06$; Fig. 4, online Supplementary Table S5). Although the latter approached significance such that subjects had a slight tendency to switch choices more while under $CO_2$. These analyses were complemented with a full computational model (online Supplementary Material), with the only difference being that the effect of $CO_2$ on choice switching was significant in this more comprehensive computational analysis (online Supplementary Table S8). Thus, it appears there may be a modest association between acute anxiety and an increased tendency to explore new options from trial to trial.

### *Individual differences*

Following the same logic as Experiment 1 – that individual differences in sensitivity to $CO_2$ might be important in revealing the effect of stress on goal-directed behaviour and switching (Otto et al., 2013; Radenbach et al., 2015; Schwabe & Wolf, 2010) – we tested if the effects of $CO_2$ on model-based planning might be detectable when we take into account how strongly subjects reacted to the $CO_2$ manipulation. As we were not powered to construct a model with a 4-way interaction (which requires the inclusion of all subordinate interactions), we extracted individual coefficients for the effect of $CO_2$ on model-based planning and switching and tested for correlation with subjects' change in self-
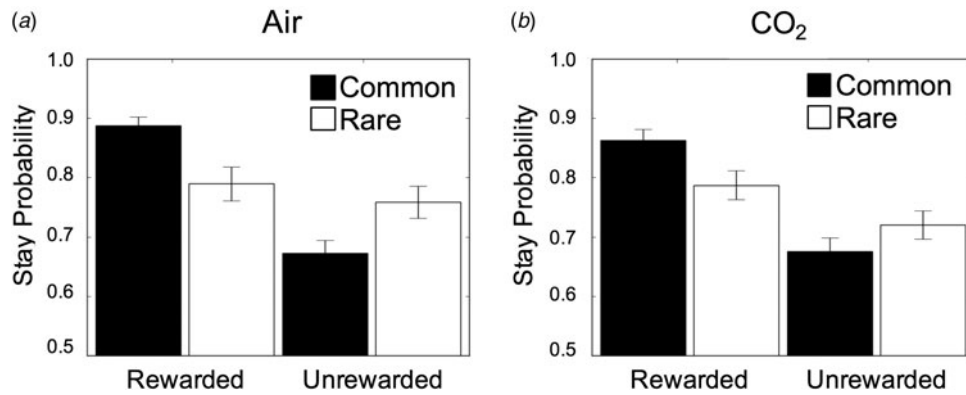
**Fig. 4.** Results from Experiment 2. (*a*) Stay/switch behaviour for subjects in air condition as a function of whether or not the last trial was rewarded/unrewarded and followed a rare/common transition. Error bars represent s.e. (*b*) The same plot, showing the group average behaviour under $CO_2$. In both plots, subjects showed the classic signatures of both model-based and model-free planning, indexed by a significant reward × transition interaction ($\beta = 0.28$, s.e. = 0.06, $p < 0.001$) and a main effect of reward ($\beta = 0.55$, s.e. = 0.08, $p < 0.001$). Error bars represent s.e.

reported anxiety and heart rate under $CO_2$. There was no significant correlation between the effect of $CO_2$ on model-based planning and change in anxiety, $r = -0.20$, $p = 0.16$, but there was a marginal association with a change in the heart rate under $CO_2$, $r = -0.29$, $p = 0.05$. The analogous analysis from the computational model provided less support, where the correlation between change in self-reported anxiety was not significant, $r = -0.20$, $p = 0.18$, and nor was the correlation with a change in the heart rate, $r = -0.15$, $p = 0.30$. The Bayes factor indicated there was anecdotal evidence for the null with respect to the correlation between changes in self-reported anxiety and model-based planning (regression: $BF_{01} = 1.3$; computational model: $BF_{01} = 1.4$). For a change in the heart-rate, there was anecdotal evidence in favour of a relationship with the change in model-based planning in regression analysis ($BF_{10} = 1.74$), but anecdotal evidence in favour of the null from the computational analysis ($BF_{01} = 1.9$). Nonetheless, the direction of these trends, on the whole, suggested that those subjects whose model-based planning performance declined the most during $CO_2$ may have also had the biggest psychological and physiological reaction to $CO_2$. However, it is notable that (i) these results go in the *opposite* direction to those in Experiment 1 and (ii) if they exist, they are *very small*. To contextualise these findings in terms of the effect size, a sample of $N = 258$ would be needed for future studies to have 90% power to detect an association between change in anxiety and changes in model-based planning under $CO_2$ using either the regression or computational model. For the heart-rate, $N = 462$ would be needed to have 90% power to detect an association with a change in the computational modelling parameterisation of model-based planning, and $N = 119$ to detect changes in the regression-defined model-model-based planning.

In contrast to model-based planning, there was a significant main effect of $CO_2$ on switching. Though not the focus of the present study, we thus repeated the individual difference analysis for switching in an exploratory fashion. We found mixed evidence. There was an association with a change in self-report anxiety, where those individuals who were most anxious under $CO_2$ tended to switch more under $CO_2$ (regression: $r = -0.43$, $p = 0.001$; computational model: $r = -0.29$, $p = 0.04$). However, the same was not true for a change in the heart rate (regression: $r = -0.13$, $p = 0.37$, computational model: $r = -0.09$, $p = 0.54$). There was strong evidence that a change in self-reported anxiety correlated with a change in switching behaviour under

$CO_2$ in the regression ($BF_{10} = 25.4$), but only anecdotal evidence for this from the full computational model ($BF_{10} = 2.17$). For the heart rate, there was anecdotal evidence in favour of the *null* from both analyses ($BF_{01} = 2.13$; $BF_{01} = 2.6$).

### Real life anxiety (Experiment 3)

In two independent studies (Experiments 1 and 2) we found no effect of an acute anxiety induction on goal-directed planning. In a final experiment, we tested if anxiety in a real-life, more ecologically valid, setting might be necessary to reveal the hypothesised detrimental effect of anxiety on goal-directed behaviour. We tested 1413 subjects online using Amazon's Mechanical Turk on the model-based learning task described above. Findings relating to the association between compulsivity and model-based planning have been published elsewhere (Gillan et al., 2016), but in data not previously published, we enquired about whether subjects had a panic attack in the past week, which is known to induce a temporary state of acute anxiety. We chose to examine panic attacks, rather than using a questionnaire probing state anxiety, because state anxiety has an unacceptably high correlation with trait anxiety when measured in the absence of an acute stressor [e.g. $r = 0.71$ (Grös, Antony, Simms, & McCabe, 2007)]. As our prior work has already demonstrated that trait anxiety is not related to goal-directed planning (Gillan et al., 2016), we wanted to ensure that our measure of acute anxiety was not in large part confounded by trait anxiety. Measuring the occurrence of recent panic attacks is an attractive alternative (although not without limitation), because they represent an acute anxiety provoking event (Aronson & Logue, 1988) and as such is more comparable to our lab-based anxiety induction. Criteria for a panic attack were from item 1 of a validated instrument [PDSS (Shear et al., 1997)] and in brief required subjects to have experienced 4 of 17 symptoms (e.g. rapid or pounding heartbeat, feeling of choking, nausea, chills or hot flushes, fear of dying) and that the panic attack must have been a 'sudden rush of fear or discomfort', peaking within 10 min. Episodes like panic attacks that have fewer than four symptoms were defined as limited symptom attacks, but also contributed to subjects' score. Specifically, subjects indicated the frequency of panic or limited symptom attacks in the past week on item 1 of the PDSS and this served as our measure for subsequent analyses.
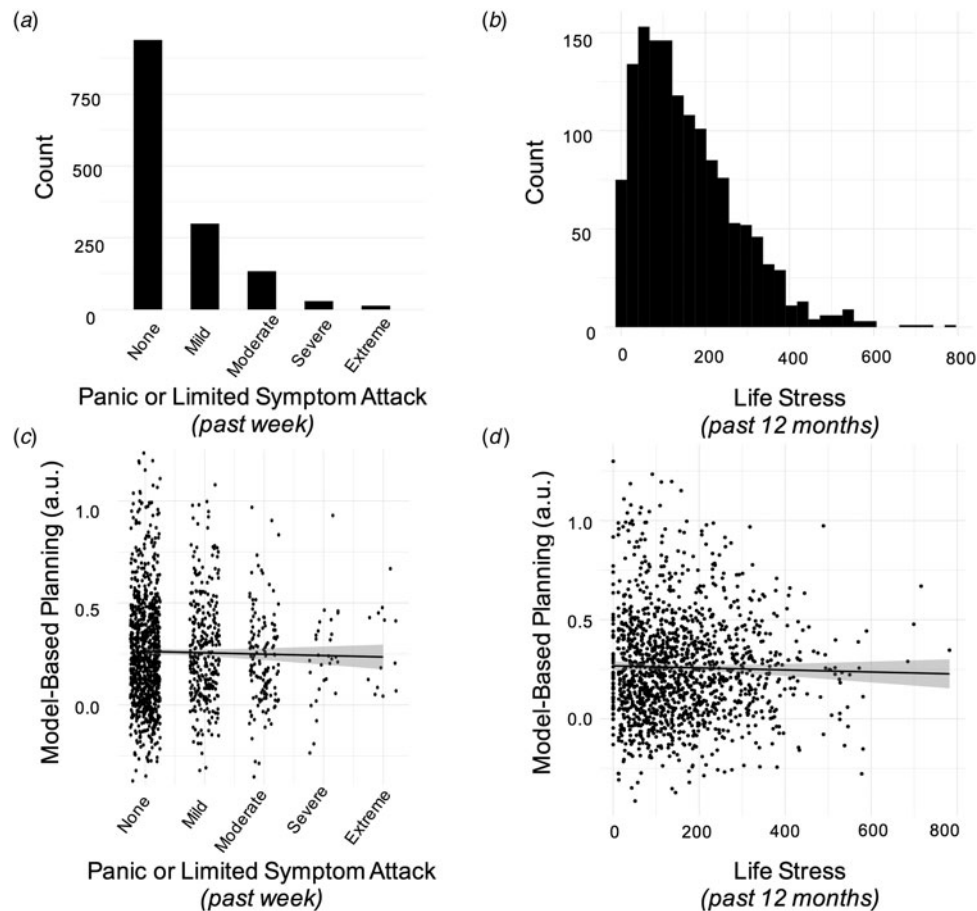
**Fig. 5.** Results from Experiment 3. (*a*) Histogram displaying the number of individuals endorsing the various levels of frequency and severity of panic/limitd symptom attacks in the past week. Scores were coded as follows: none ('no panic or limited symptom attacks'), mild (no full panic attacks and no more than 1 limited symptom attack/day), moderate ('1 or 2 full panic attacks and/or multiple limited symptom attacks/day'), severe (severe: more than 2 full attacks but not more than 1/day on average) and extreme ('full panic attacks occurred more than once a day, more days than not'). (*b*) Histogram displaying the distribution of life stress scores in the sample. (*c*) There was no association between model-based planning and the occurrence of panic attacks in the past week, after controlling for age, gender, IQ and compulsive symptomatology, $\beta = -0.01$, s.e. $= 0.01$, $p = 0.33$. The *Y*-axis displays residuals for model-based planning after these features are taken into account. (*d*) There was no association between model-based planning and life stress experienced over the past year, after controlling for age, gender, IQ and compulsive symptomatology, $\beta = -0.01$, s.e. $= 0.01$, $p = 0.33$. As above, the *Y*-axis displays residuals for model-based planning.

Consistent with other general population samples (Barrera, Wilson, & Norton, 2010), approximately a third ($N = 474$) of our online sample indicated they had experienced a panic or limited symptom attack in the past week (Fig. 5*a*). The frequency of panic attacks in the past week was correlated with reductions in model-based planning ($\beta = -0.03$, s.e. $= 0.01$, $p = 0.012$), but this did not survive controlling for 'Compulsive Behaviour and Intrusive Thought', a transdiagnostic psychiatric dimension that is negatively correlated with model-based planning ($\beta = -0.04$, s.e. $= 0.01$, $p < 0.001$); note this finding was previously published (Gillan et al., 2016). This compulsive factor presents a confound to interpretation of the panic results, because it is positively correlated with frequency of panic attacks ($r = 0.42$, $p < 0.001$). When compulsivity was accounted for, the effect of panic attacks on model-based planning was reduced to $\beta = -0.01$, s.e. $= 0.01$, $p = 0.33$ (Fig. 5*c*). Moreover, results from the more elaborate computational model showed that the effect of panic attacks on goal-directed planning approached zero and went in the opposite direction ($\beta = 0.003$, s.e. $= 0.01$, $p = 0.81$) after compulsivity was controlled for (online Supplementary Table S9).

We observed an association between the frequency of panic attacks and choice switching ($p = 0.012$), mirroring our causal result from Experiment 2. However, the effect of panic attacks on increased switching did not survive inclusion of compulsivity in the model for the one-trial-back regression ($p = 0.23$), or in the computational model (online Supplementary Table S9; $p = 0.06$).

Finally, we tested if life stress in the past year was associated with deficits in model-based planning. This was assessed using the Social Readjustment Scale (Holmes & Rahe, 1967), which presents an inventory of common stressful life events to participants and asks them to select those that applied to them in the previous 12 months (e.g. at the extreme end these include the death of a spouse or divorce) (Fig. 5*b*). Much like a recent panic attack, we found that life stress scores were linked to failures in model-based planning ($\beta = -0.02$, s.e. $= 0.01$, $p = 0.04$). However, as was the case for panic attacks, life stress was also correlated with the compulsive factor ($r = 0.29$, $p < 0.001$), and indeed the relationship to model-based planning did not survive inclusion of the compulsive factor in the analysis. Specifically, the effect of life stress on model-based planning was reduced to $\beta = -0.01$, s.e. $= 0.01$, $p = 0.33$ (Fig. 5*d*) in the regression analysis

and $\beta = -0.01$, S.D. $= 0.01$, $p = 0.24$ in the full computational model (online Supplementary Table S10).

## Discussion

Across three independent studies, we found little or no evidence that anxiety has a detrimental effect on goal-directed planning. The first two experiments employed an extensively validated causal manipulation for inducing an acute state of anxiety, inhalation of air enriched with $CO_2$ (Argyropoulos et al., 2002; Bailey et al., 2005). Using both between- and within-subject designs, and two well-validated tests for goal-directed behaviour, neither study found evidence that the causal manipulation had an effect on model-based planning. A third study took a correlational, but the larger scale ($N = 1413$), approach and tested if individuals who had panic/limited symptom attacks in the past week, which are associated with an increase in acute state anxiety (Aronson & Logue, 1988), had poorer goal-directed performance. Unlike most clinical studies, this design incorporated a comprehensive range of clinical assessments and could thus control for clinical confounds such as trait differences in compulsivity. While we found that those who experienced more panic attacks in the past week had poorer goal-directed planning, this did not survive controlling for compulsivity, a correlated trait that has been extensively studied in the content of goal-directed control failures. Together, these data contribute to a larger literature suggesting that trait (Gillan et al., 2016), and now state, anxiety do not have a clear detrimental effect on goal-directed planning in human subjects.

The most consistent cognitive changes that have been linked to trait anxiety are an increased attentional bias to threat or 'hypervigilance' (Mogg, Bradley, de Bono, & Painter, 1997) and the tendency to interpret ambiguous stimuli as threatening (Eysenck, Mogg, May, Richards, & Mathews, 1991). Results from studies using the 7.5% $CO_2$ challenge closely mirror these findings – with the manipulation increasing alerting and orienting (Garner et al., 2012), threat processing (e.g. hypervigilance) (Garner et al., 2011) and negative interpretations of neutral events (Cooper et al., 2013), suggesting that 7.5% hypercapnic gas manipulation in the lab can mirror cognitive changes observed in association with anxiety. While the putative role that anxiety plays in more complex forms of decision-making is of broad interest (Paulus & Yu, 2012), there is a dearth of evidence suggesting it has effects that are not explained by increases in threat-sensitivity and vigilance. For example, while there is some evidence to suggest that clinically anxious individuals tend to make better long-term choices e.g. on the Iowa Gambling Task (IGT), this appears to result from a bias to avoid losses, which in the context of this task is confounded with the choice of 'advantageous' decks (Mueller, Nguyen, Ray, & Borkovec, 2010). Even this, however, has been inconsistently shown, with another study finding that high trait anxiety leads to *impaired* IGT performance (Miu, Heilman, & Houser, 2008). One potential explanation for inconsistent results in this area is that studies have been largely cross-sectional and correlational – something we overcame here by using causal manipulation of anxiety in studies 1 and 2.

Prior studies have suggested that, in the absence of a main effect of stress on goal-directed control, individual differences in sensitivity to the stressor itself may be important to consider (Otto et al., 2013; Radenbach et al., 2015). We repeated this general analytic approach here to facilitate comparison across studies, but the data were equivocal. There was no evidence that

self-report anxiety or physiological sensitivity (i.e. heart rate) to the $CO_2$ manipulation was associated with effects of the stressor on goal-directed behaviour in Experiment 1. One analysis of Experiment 2 data showed a marginal effect in the expected direction– diminished performance on the model-based learning task was observed in individuals that had the biggest change in heart rate under $CO_2$. But notably, evidence was anecdotal and sometimes in favour of the null, depending on the analysis in question. Results varied depending on the measure of goal-directed planning was computational *v.* regression-based and whether the individual difference measure was self-report or physiological. More generally, it is difficult to interpret these effects in any causal framework given the absence of a main effect, such that these associations are driven, in part, by individuals who actually performed nominally *better* under $CO_2$ ($N = 22/50$ in Experiment 2). Moreover, individual differences in sensitivity to $CO_2$ is a somewhat problematic measure because it is itself a marker of mental health difficulties (Perna et al., 1996), presenting a confound to interpretation.

Although no previous studies have examined the effect of acute experimentally induced state anxiety on goal-directed control, several studies examined the impact of stress (Dias-Ferreira et al., 2009; Heller et al., 2018; Otto et al., 2013; Radenbach et al., 2015; Schwabe & Wolf, 2009, 2010) in healthy volunteers. Three studies found that stress-induced goal-directed deficits (Park et al., 2017; Schwabe & Wolf, 2009, 2010), mirroring findings in rodents following 21 days of unpredictable stress exposure (Dias-Ferreira et al., 2009). Three other studies, however, found no such effect (Heller et al., 2018; Otto et al., 2013; Radenbach et al., 2015). One key point of departure between studies reporting positive and negative results was the type of stressor used. Those that found significant effects used a socially-evaluated cold pressor test, and those that did not use either the cold pressor in isolation (Otto et al., 2013), or a social stress test in isolation (Heller et al., 2018; Radenbach et al., 2015). This distinction is important as the socially evaluated cold pressor test has been shown to induce a much stronger increase in cortisol, compared to cold pressor test alone (Schwabe, Haddad, & Schachinger, 2008), with the procedures otherwise eliciting similar cardiovascular and subjective stress responses. The notion that cortisol might mediate stress effects on goal-directed planning is supported by the observation that changes in cortisol were linked to deficits in performance in studies that failed to otherwise show a main effect of stress (Otto et al., 2013; Radenbach et al., 2015). In other words, the largest increases in cortisol were linked to the largest task deficits. This ties in with pharmacological evidence showing that decrements in goal-directed performance cannot be induced through noradrenergic manipulation alone; concurrent glucocorticoid stimulation is also necessary (although not sufficient) (Schwabe et al., 2010; Schwabe, Tegenthoff, Höffken, & Wolf, 2012). Differential involvement of cortisol might explain why acute stress appears to have an impact on goal-directed planning, but anxiety induction does not. While acute stress and anxiety induction result in similar cardiovascular effects (i.e. increases in heart rate and blood pressure) (Bailey et al., 2005; Schwabe et al., 2008) and noradrenergic activation (Allen, Kennedy, Cryan, Dinan, & Clarke, 2014; Bailey, Argyropoulos, Lightman, & Nutt, 2003), anxiety induction via 7.5% $CO_2$ does not result in a reliable increase in cortisol (Oliveira, Chagas, Garcia, Crippa, & Zuardi, 2012; Woods et al., 1988). Hypercapnia causes more pronounced and specific increases in self-reported feelings of anxiousness, fear, panic and worry, which are reduced in

response to common treatments for generalised anxiety, including anxiolytics (Bailey, Kendrick, Diaper, Potokar, & Nutt, 2007; Diaper et al., 2012). Therefore, it is possible that our results are specific to anxiety induction rather than stress *per se*.

The extent to which more chronic forms of real-life stress impair goal-directed control is an open question and has only been partially addressed in one prior study with a relatively small sample ($N = 39$) (Radenbach et al., 2015). Subjects with high self-reported chronic stress levels had a larger effect of acute stress on model-based planning performance, than their low stress counterparts (Radenbach et al., 2015). This might suggest that goal-directed learning is in some sense more fragile in individuals who have high levels of chronic life stress, but this is difficult to assess as the authors did not report any test for the direct association between life stress and model-based planning. We tested this using a large sample ($N = 1413$) and did not find evidence for an association, after controlling for compulsivity. This suggests that the impact of real-life stress on goal-directed planning, if it exists, is certainly less pronounced than folk wisdom suggests. That said, here we studied goal-directed behaviour, rather than habit expression *per se*, which represents a point of departure from some of the prior research e.g. in rodents (Dias-Ferreira et al., 2009). Further work is needed in this direction as it is possible that any effect of anxiety is on habit expression, and not goal-directed control.

In Experiments 2 and 3, there was a suggestion that subjects' tendency to switch their choices from one trial to the next was increased following anxiety induction and the recent occurrence of a panic attack, respectively. These findings were not hypothesised and effect sizes were somewhat inconsistent across analysis methods, but given their consistency with a prior independent study (Radenbach et al., 2015), they warrant brief discussion. One possibility is that this increase in choice switching might reflect the enhanced uncertainty characteristic of anxious states (Grupe & Nitschke, 2013) and could arise as a result of activation of the noradrenergic system (Redmond & Huang, 1979; Yu & Dayan, 2005). Evidence for this comes from work suggesting that tonic noradrenaline release is linked to an increase in task irrelevant processing and a tendency to favour exploration over exploitation (Aston-Jones & Cohen, 2005), characterised by some as a network 'reset' (Bouret & Sara, 2005). This interpretation is limited by the absence of data on cortisol and noradrenaline response and the exploratory nature of the findings. Future research will be needed to test this more directly, using a cognitive test designed to explicitly separate exploration and exploitation.

This study had limitations. Firstly, null results are difficult to draw firm conclusions from. However, the findings of Experiment 3, which benefit from the inclusion of a previously published clinical effect size comparator (the effect of compulsivity on model-based planning), help to place these null findings into a meaningful context. It is unlikely that our manipulation was not strong enough to induce a robust anxiogenic effect because previous studies have demonstrated that the 7.5% $CO_2$ manipulation is powerful enough to elicit robust effects on behavioural performance relating to threat sensitivity and hyper-vigilance (Cooper et al., 2013; Garner et al., 2011, 2012), in addition to its well-documented physiological and psychological effects (Bailey et al., 2005, 2007). The magnitude of self-report and physiological changes in the present study was on-par with those observed in prior studies (Cooper et al., 2013; Garner et al., 2011, 2012). Finally, Bayesian analyses detail the extent to which evidence was in favour of the null, and this

was in most cases in the 'very strong' range. A second limitation is that using panic attacks to measure 'real world' state anxiety is an imperfect methodology. Although panic attacks are associated with an increase in state anxiety (Aronson & Logue, 1988), they are also associated with, and defined by, a much broader cascade of physical symptoms than the experience of state anxiety. However, this approach has two advantages over measuring self-reported state anxiety [e.g. using the STAI-state scale (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983)]. First, in the absence of an acute event (anxiety trigger), trait and state anxiety scores tend to be highly correlated [e.g. $r = 0.71$ (Grös et al., 2007)] and the STAI-scale is thus thought to be more reflective of trait than state anxiety. Second, leveraging naturally occurring panic attacks allowed us to mirror the acute and sudden onset of anxiety that our lab-based procedure achieved.

## Conclusions

Experimentally induced state anxiety failed to produce deficits in goal-directed behaviour as measured via two independent experiments using two well-validated probes. Such lack of effect was also observed in a more ecologically valid set-up, where we used recent panic attacks as a proxy for acute anxiety. While modest decreases in goal-directed planning were seen in individuals who had recent panic attacks in the past year, these effects did not survive when controlling for compulsivity. The same was true of the occurrence of major life stressors in the past year. In terms of clinical implications, these data suggest that state anxiety has little *specific* effect on goal-directed control, in contrast for example to compulsivity, which research has shown has a consistent association. This distinction may have important implications for the development of differential treatment approaches for patients who present with the same diagnosis, for example of OCD, but differ substantially in their levels of anxiety *v.* compulsivity. Dimensional approaches that seek to distinguish these dimensions and target them individually present a new frontier for psychiatry research aiming to develop more personalised treatment approaches. For future research studies more generally, these data highlight the necessity of using positive clinical control measures and causal manipulations to ascertain robust and specific associations given a deeply complex and highly inter-correlated mental health landscape.

institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

**Code availability statement.** Code to reproduce results is freely available on the Open Science Framework (https://osf.io/w4yfp/).

# References

Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2014). Biological and psychological markers of stress in humans: Focus on the trier social stress test. *Neuroscience Biobehavioral Review*, *38*, 94–124. doi: 10.1016/j.neubiorev.2013.11.005

Alvares, G. A., Balleine, B. W., & Guastella, A. J. (2014). Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PLoS One*, *9*(4), e94778. doi: 10.1371/journal.pone.0094778

Argyropoulos, S. V., Bailey, J. E., Hood, S. D., Kendrick, A. H., Rich, A. S., Laszlo, G., … Nutt, D. J. (2002). Inhalation of 35% CO(2) results in activation of the HPA axis in healthy volunteers. *Psychoneuroendocrinology*, *27*(6), 715–729.

Arnsten, AF. (1998). Catecholamine modulation of prefrontal cortical cognitive function. *Trends in cognitive sciences*, *2*(11), 436–447.

Aronson, T. A., & Logue, C. M. (1988). Phenomenology of panic attacks: A descriptive study of panic disorder patients' self-reports. *Journal of Clinical Psychiatry*, *49*(1), 8–13.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Bailey, J. E., Argyropoulos, S. V., Kendrick, A. H., & Nutt, D. J. (2005). Behavioral and cardiovascular effects of 7.5% $CO_2$ in human volunteers. *Depression and Anxiety*, *21*(1), 18–25. doi: 10.1002/da.20048

Bailey, J. E., Argyropoulos, S. V., Lightman, S. L., & Nutt, D. J. (2003). Does the brain noradrenaline network mediate the effects of the $CO_2$ challenge? *Journal of Psychopharmacology*, *17*(3), 252–259. doi: 10.1177/02698811030173002

Bailey, J. E., Kendrick, A., Diaper, A., Potokar, J. P., & Nutt, D. J. (2007). A validation of the 7.5% $CO_2$ model of GAD using paroxetine and lorazepam in healthy volunteers. *Journal of Psychopharmacology*, *21*(1), 42–49. doi: 10.1177/0269881106063889

Barrera, T. L., Wilson, K. P., & Norton, P. J. (2010). The experience of panic symptoms across racial groups in a student sample. *Journal of Anxiety Disorders*, *24*(8), 873–878. doi: 10.1016/j.janxdis.2010.06.010

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neuroscience*, *28*(11), 574–582. doi: 10.1016/j.tins.2005.09.002

Cooper, R., Howard, C. J., Attwood, A. S., Stirland, R., Rostant, V., Renton, L., … Munafò, M. R. (2013). Acutely induced anxiety increases negative interpretations of events in a closed-circuit television monitoring task. *Cognition and Emotion*, *27*(2), 273–282. doi: 10.1080/02699931.2012.704352

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). 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. *Nature Neuroscience*, *8*(12). doi: 10.1038/nn1560

Diaper, A., Papadopoulos, A., Rich, A. S., Dawson, G. R., Dourish, C. T., Nutt, D. J., & Bailey, J. E. (2012). The effect of a clinically effective and non-effective dose of lorazepam on 7.5% $CO_2$-induced anxiety. *Human Psychopharmacology*, *27*(6), 540–548. doi: 10.1002/hup.2261

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., … Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science (New York, N.Y.)*, *325* (5940), 621–625. doi: 325/5940/621 [pii] 10.1126/science.1171203

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1135), 67–78.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18. doi: 10.3758/BF03199951

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology*, *35*(FEB)

Eysenck, MW, Derakshan, N, Santos, R, & Calvo, MG. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychoogyl*, *100*(2), 144–150.

Garner, M., Attwood, A., Baldwin, D. S., James, A., & Munafò, M. R. (2011). Inhalation of 7.5% carbon dioxide increases threat processing in humans. *Neuropsychopharmacology*, *36*(8), 1557–1562. doi: 10.1038/npp.2011.15

Garner, M., Attwood, A., Baldwin, D. S., & Munafò, M. R. (2012). Inhalation of 7.5% carbon dioxide increases alerting and orienting attention network function. *Psychopharmacology* (*Berl*), *223*(1), 67–73. doi: 10.1007/s00213-012-2690-4

Gillan, C., Kosinski, M., Whelan, R., Phelps, E., & Daw, N. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife*, *5*(e11305). doi: http://dx.doi.org/10.7554/eLife.11305

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive. Affective and Behavioral Neuroscience*, *15*, 523–536.

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry*, *168*(7), 718–726. doi: appi.ajp.2011.10071062 [pii] 10.1176/appi.ajp.2011.10071062

Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive-compulsive disorder. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1655). doi: 10.1098/rstb.2013.0475

Griez, E., Zandbergen, J., Pols, H., & de Loof, C. (1990). Response to 35% $CO_2$ as a marker of panic in severe anxiety. *Americal Journal of Psychiatry*, *147* (6), 796–797. doi: 10.1176/ajp.147.6.796

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, *19*(4), 369–381. doi: 10.1037/1040-3590.19.4.369

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Review Neuroscience*, *14*(7), 488–501. doi: 10.1038/nrn3524

Heller, A. S., Ezie, C. E. C., Otto, A. R., & Timpano, K. R. (2018). Model-based learning and individual differences in depression: The moderating role of stress. *Behaviour Research and Therapy*, *111*, 19–26. doi: 10.1016/j.brat.2018.09.007

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, *11*(2), 213–218.

Liebowitz, M. R., Fyer, A. J., Gorman, J. M., & et al. (1984). Lactate provocation of panic attacks: I. clinical and behavioral findings. *Archives of General Psychiatry*, *41*(8), 764–770.

Mataix-Cols, D., Cullen, S., Lange, K., Zelaya, F., Andrew, C., Amaro, E., & … Phillips, M. L. (2003). Neural correlates of anxiety associated with obsessive-compulsive symptom dimensions in normal volunteers. *Biological Psychiatry*, *53*(6), 482–493. doi:10.1016/s0006-3223(03)01504-4.

Miu, A. C., Heilman, R. M., & Houser, D. (2008). Anxiety impairs decision-making: Psychophysiological evidence from an Iowa Gambling Task. *Biological Psychology*, *77*(3), 353–358. doi: 10.1016/j.biopsycho.2007.11.010

Mogg, K., Bradley, B. P., de Bono, J., & Painter, M. (1997). Time course of attentional bias for threat information in non-clinical anxiety. *Behaviour Research and Therapy*, *35*(4), 297–303.

Mueller, E. M., Nguyen, J., Ray, W. J., & Borkovec, T. D. (2010). Future-oriented decision-making in generalized anxiety disorder is evident across different versions of the Iowa Gambling Task. *The Journal of Behavior Therapy and Experimental Psychiatry*, *41*(2), 165–171. doi: 10.1016/j.jbtep.2009.12.002

Nestadt, G., Di, C. Z., Riddle, M. A., Grados, M. A., Greenberg, B. D., Fyer, A. J., … Roche, K. B. (2009). Obsessive-compulsive disorder: Subclassification based on co-morbidity. *Psychological Medicine*, *39*(9), 1491–1501. doi: 10.1017/S0033291708004753

Oliveira, D., Chagas, M., Garcia, L., Crippa, J., & Zuardi, A. (2012). Oxytocin interference in the effects induced by inhalation of 7.5% $CO_2$ in healthy volunteers. *Human Psychopharmacology Clinical and Experimental*, *27*, 378–385.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(52), 20941–20946. doi: 10.1073/pnas.1312011110

Park, H., Lee, D., & Chey, J. (2017). Stress enhances model-free reinforcement learning only after negative outcome. *PLoS One*, *12*(7), e0180588. doi: 10.1371/journal.pone.0180588

Paulus, M. P., & Yu, A. J. (2012). Emotion and decision-making: Affect-driven belief systems in anxiety and depression. *Trends in Cognitive Sciences*, *16*(9), 476–483. doi: 10.1016/j.tics.2012.07.009

Perna, G., Barbini, B., Cocchi, S., Bertani, A., & Gasperini, M. (1995). 35% $CO_2$ challenge in panic and mood disorders. *Journal of Affective Disorders*, *33*(3), 189–194.

Perna, G., Battaglia, M., Garberi, A., Arancio, C., Bertani, A., & Bellodi, L. (1994). Carbon dioxide/oxygen challenge test in panic disorder. *Psychiatry Research*, *52*(2), 159–171.

Perna, G., Bertani, A., Caldirola, D., & Bellodi, L. (1996). Family history of panic disorder and hypersensitivity to $CO_2$ in patients with panic disorder. *American Journal of Psychiatry*, *153*(8), 1060–1064. doi: 10.1176/ajp.153.8.1060

Radenbach, C., Reiter, A. M., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H. J., … 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

Redmond, D. E., & Huang, Y. H. (1979). Current concepts. II. New evidence for a locus coeruleus-norepinephrine connection with anxiety. *Life Sciences*, *25*(26), 2149–2162.

Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends in Cognitive Sciences*, *16*(1), 81–91. doi: 10.1016/j.tics.2011.11.009

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry*, *84*(6), 443–451. doi: 10.1016/j.biopsych.2017.12.017

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *T* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225

Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, *33*(6), 890–895. doi: 10.1016/j.psyneuen.2008.03.001

Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2010). Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *Journal of Neuroscience*, *30*(24), 8190–8196. doi: 10.1523/JNEUROSCI.0734-10.2010

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

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*(22), 7191–7198. doi: 29/22/7191 [pii] 10.1523/JNEUROSCI.0979-09.2009

Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, *35*, 977–986.

Shear, M. K., Brown, T. A., Barlow, D. H., Money, R., Sholomskas, D. E., Woods, S. W., … Papp, L. A. (1997). Multicenter collaborative panic disorder severity scale. *American Journal of Psychiatry*, *154*(11), 1571–1575. doi: 10.1176/ajp.154.11.1571

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience Biobehavioral Review*, *68*, 651–668. doi: 10.1016/j.neubiorev.2016.06.038

Shin, L. M., & Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology*, *35*(1), 169–191. doi: 10.1038/npp.2009.83

Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T., Penninx, B. W., & Veltman, D. J. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry*, *3*, e337. doi: 10.1038/tp.2013.107

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stein, D. J., Fineberg, N. A., Bienvenu, O. J., Denys, D., Lochner, C., Nestadt, G., … Phillips, K. A. (2010). Should OCD be classified as an anxiety disorder in DSM-V? *Depression and Anxiety*, *27*(6), 495–506. doi: 10.1002/da.20699

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2018). Action-outcome knowledge dissociates from behavior in obsessive-compulsive disorder following contingency degradation. *Biological Psychiatry Cogntive Neuroscience and Neuroimaging*, *4*, 200–209.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., … Bullmore, E. T. (2014). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*, 345–352.

Vyas, A., Pillai, A. G., & Chattarji, S. (2004). Recovery after chronic stress fails to reverse amygdaloid neuronal hypertrophy and enhanced anxiety-like behavior. *Neuroscience*, *128*(4), 667–673. doi: 10.1016/j.neuroscience.2004.07.013

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Woods, S. W., Charney, D. S., Goodman, W. K., & Heninger, G. R. (1988). Carbon dioxide-induced anxiety. Behavioral, physiologic, and biochemical effects of carbon dioxide in patients with panic disorders and healthy subjects. *Archives of General Psychiatry*, *45*(1), 43–52.

Yu, A., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4).