nature human behaviour

ARTICLES
https://doi.org/10.1038/s41562-021-01180-y

Check for updates

Increased and biased deliberation in social anxiety

Lindsay E. Hunter[1] ✉, Elana A. Meer[2,8], Claire M. Gillan[3,4,5], Ming Hsu[6,7] and Nathaniel D. Daw[1,2] ✉

A goal of computational psychiatry is to ground symptoms in basic mechanisms. Theory suggests that avoidance in anxiety disorders may reflect dysregulated mental simulation, a process for evaluating candidate actions. If so, these covert processes should have observable consequences: choices reflecting increased and biased deliberation. In two online general population samples, we examined how self-report symptoms of social anxiety disorder predict choices in a socially framed reinforcement learning task, the patent race, in which the pattern of choices reflects the content of deliberation. Using a computational model to assess learning strategy, we found that self-report social anxiety was indeed associated with increased deliberative evaluation. This effect was stronger for a particular subset of feedback ('upward counterfactual') in one of the experiments, broadly matching the biased content of rumination in social anxiety disorder, and robust to controlling for other psychiatric symptoms. These results suggest a grounding of symptoms of social anxiety disorder in more basic neuro-computational mechanisms.

Unique among areas of medicine, psychiatry has no laboratory diagnostic tests. This is largely due to a lack of understanding regarding how mental health symptoms arise from dysfunction in underlying brain mechanisms. Recent research has attempted to fill this gap by connecting these disorders to relatively well-characterized neuro-computational systems, notably those that support reinforcement learning (RL)[1–5]. In this respect, a key feature of RL in the brain is that it arises from a combination of at least two evaluative mechanisms, more deliberative versus automatic, which have been formalized in terms of model-based and model-free learning[6]. Model-based learning evaluates actions by iteratively simulating their consequences using a learned representation, or 'model' of the task's contingencies, such as a spatial map. Model-free learning skirts this computation by learning actions' long-run values directly from experience when they are chosen; these values permit quick but inflexible decisions and are a potential substrate for habits. A line of research describes the biological substrates for these functions, such as representations of future spatial trajectories in hippocampus that may support mental simulations of candidate routes for model-based evaluation[7,8] and dopaminergic temporal-difference prediction error signals suited for model-free learning[9]. Although adaptive behaviour relies on the ability to flexibly recruit both strategies, people also vary greatly in their tendency to do so. Accordingly, by comparing RL models with trial-by-trial choices in people learning sequential choice tasks (for example, two-step Markov decision processes (MDPs)[10]), the degree to which subjects utilize model-based RL has been shown to vary both situationally (for example, under dual-task interference[11]) and between individuals (for example, with genotypic variations that affect prefrontal dopamine[12]).

Abnormal imbalance between these mechanisms has also been the focus of persistent investigation regarding mental illness. In particular, it has long been suggested that symptoms related to compulsion (a dimension that cuts across illnesses including obsessive–compulsive disorder (OCD) and drug abuse) might arise from an imbalance that favours automaticity[13–17]. Indeed, for a variety of disorders involving compulsion, both patients' diagnoses[18] and self-reported symptoms in a large general-population sample[16] are associated with deficient model-based learning in a two-step MDP.

More tentatively, theorists have suggested that some aspects of other disorders (notably, worry and overthinking in anxiety and also rumination in depression) might relate to a converse imbalance, favouring excess deliberation[3,19–21]. In turn, such aberrant covert evaluation might drive other, more behavioural symptoms of these disorders, including avoidance in anxiety[20,22]. There is as yet less evidence relating symptoms of anxiety or depression to increased mental simulation in value-based learning (and a negative result in a small depression sample[23]). However, some analyses in ref. [16] revealed a small trend of increased model-based evaluation specifically for social anxiety, rather than other depressive and anxious symptoms. Social anxiety is also an interesting test case for this hypothesis, both practically (because clinically significant levels of it are frequent in the population tested online[24]) and substantively, because it involves enhanced mentalizing and counterfactual thinking (see ref. [25] for a review), both of which are psychological constructs closely related to model-based evaluation.

We thus sought to investigate the hypothesis that social anxiety is associated with increased deliberation. To better probe their relationship, we turned to another well-studied type of learning task with a more social framing: a competitive economic game[26–28], the patent race, in which subjects must learn to adjust their behaviour to their competitors' moves. As with MDPs, choices on this sort of task are well characterized by a subject-specific combination of two strategies, paralleling model-free and model-based RL: direct (model-free) learning about which moves are successful versus deriving moves' values indirectly based on which moves the opponent prefers (equivalent to a model of the opponent[29]). Moreover, the patterns of neuroimaging correlates and individual differences (for example, with dopaminergic genes) related to this dichotomy parallel those reported for MDPs[27,28].

The patent race game also enables us to investigate anxiety's effects not just on overall model usage but also on more granular

[1]Department of Psychology, Princeton University, Princeton, NJ, USA. [2]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. [3]School of Psychology, Trinity College Dublin, Dublin, Ireland. [4]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [5]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland. [6]Haas School of Business, University of California, Berkeley, Berkeley, CA, USA. [7]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [8]Present address: Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: lehunter@princeton.edu; ndaw@princeton.edu

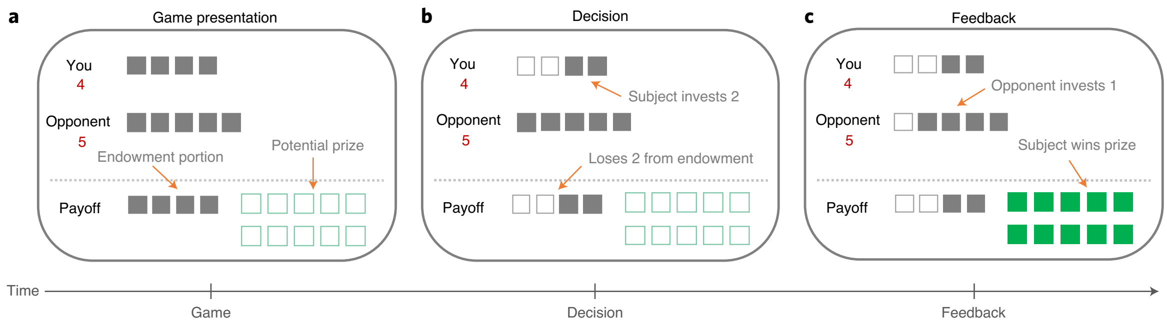

Fig. 1 | Schematic of patent race game. Prior to each round, a fixation screen appeared for a random duration between 4 and 8 s. **a**, Subjects were then presented with information regarding their endowment ($4), the endowment of the opponent ($5) and the potential prize ($10). **b**, The arrow keys allowed subjects to select how much to invest (indicated by the number of white boxes), and the space bar was then used to submit the selected investment amount. **c**, The opponent's choice was revealed 2–6 s later. If the subject's investment was strictly more than those of the opponent, the subject won the prize; in either case, the subject kept the portion of the endowment not invested.

operations of planning. Although classic models have viewed planning as unitary (that is, evaluating all possible trajectories of action in the task at once), this is clearly unrealistic, since there are typically too many possibilities to consider practically. Accordingly, recent research aims to decompose model-based evaluation into a series of steps, such as simulating individual candidate actions or accessing memories for particular events[8,23,30,31]. These more granular models suggest a mechanism for overall variation in planning (resulting from fewer or more steps, overall) but also speak to the possibility of not just quantitative but qualitative variation across individuals: bias in which actions or events are likely to be simulated. If present, such bias predicts measurable patterns of effects on choices, for example, a stronger tendency to update values for some options than others. Again, anxiety provides an intriguing test case and application for these more process-level learning models, since bias in mental simulation has been hypothesized to connect covert symptoms (such as narrow preoccupation on certain events in rumination or worry) to behavioural ones such as avoidance[20,22].

In social anxiety, studies of the content of rumination in post-event processing suggest it includes an excess of 'upward counterfactual' thoughts about previous social interactions: that is, 'if only' thoughts about how the events could have gone better, which is thought to fuel anticipatory anxiety and avoidance of future social interactions[32,33]. This suggests a testable (albeit speculative, since we did not assess rumination or its content in the current study) hypothesis about behaviour in the patent race game. Model-based evaluation in such tasks can be decomposed into steps of computing the updated value of individual candidate actions, given individual outcomes. These steps each amount to counterfactual updates: computing the value of moves not taken, in light of the opponent's move at each step[29]. A bias toward upward counterfactuals would predict a tendency toward model-based updating for a subset of the six moves available, differing from trial to trial depending on its outcome. If such a bias is observed in choices, this would connect a specific cognitive operation (biased processing of certain events) and a behavioural consequence (later choices) in anxiety.

Accordingly, to examine the relationship between social anxiety and model-based deliberation, we conducted two large-scale, online experiments. In each experiment, 500 subjects were recruited to complete the Liebowitz Social Anxiety Scale (LSAS) and play 80 rounds of a patent race game (Fig. 1) against a computerized opponent. Experiment 2 was conducted to replicate the first, and to extend it by assessing symptoms for a broader range of psychiatric symptoms, allowing us to probe the specificity of our findings. Below, we report both experiments' results in parallel, but broken out by experiment where applicable.

Results

In line with recommendations for studies conducted using Amazon Mechanical Turk (AMT), a priori exclusion criteria were applied to ensure data quality by eliminating low-effort participants[34]. Of the 966 participants (experiment 1, $N=489$; experiment 2, $N=477$) who completed the task, we eliminated participants who shirked either the Raven's matrix test (experiment 1, $N=36$; experiment 2, $N=49$) or the patent race task (experiment 1, $N=41$; experiment 2, $N=97$). (Thus, in total, 77 and 146 were removed from experiment 1 and 2, respectively, where the larger number in the second study is likely due to the longer session.) Specifically, we removed from consideration subjects who got zero or one items correct on Raven's matrices or who chose the same move on more than 95% of trials (that is >76/80 rounds) in the patent race[27]. The remaining analyses concern the data of 743 participants (experiment 1, $N=412$, 58% male, age mean 35.2 years, s.d. 10.4 years; experiment 2, $N=331$, 55% male, age mean 35.8 years, s.d. 10.1 years).

Our primary covariate of interest was self-report anxiety symptoms, measured using the LSAS[35] (Fig. 2). Consistent with previous reports for the AMT population[24], LSAS scores were high (experiment 1: mean 54.52; s.d. 29.29 out of 144; experiment 2: mean 45.34; s.d. 29.38; Fig. 2). Indeed, although dimensional self-report symptom scores as studied here do not substitute for a formal clinical diagnosis, the average participant (and 72% of all participants) scored above the threshold (30) previously shown to predict a diagnosis of social anxiety disorder[36]. We also measured fluid intelligence using an abbreviated Raven's matrix scale[37]. On average, subjects answered 5.5 (experiment 1, s.d. 1.8) and 5.4 (experiment 2, s.d. 1.7) of the nine Raven's matrix problems correctly, corresponding to a predicted mean of 46 and 44 correct on the full 60-question set (using the weighted prediction from ref. [37]; Fig. 2).

Although the patent race game in principle admits of a mixed strategy Nash equilibrium, our participants' aggregate behaviour was off it (similar to previous studies; Supplementary Table 1 (ref. [27])). Research in behavioural economics has instead emphasized the more dynamic question of how subjects adjust their choices from trial to trial in light of experience with opponents' strategies or outcomes[26–28], parallel to other studies of reinforcement learning in single-player choice tasks[10,16]. Accordingly, we characterized such learning by fitting trial-by-trial choice sequences to the experience-weighted attraction (EWA[29]) model, which is also algebraically equivalent to the hybrid reinforcement learning model[10] that is widely used to study the trade-off between model-based and model-free learning in single-player tasks in psychology and neuroscience (Methods). The fitting procedure provides, for each participant, estimates of free parameters that

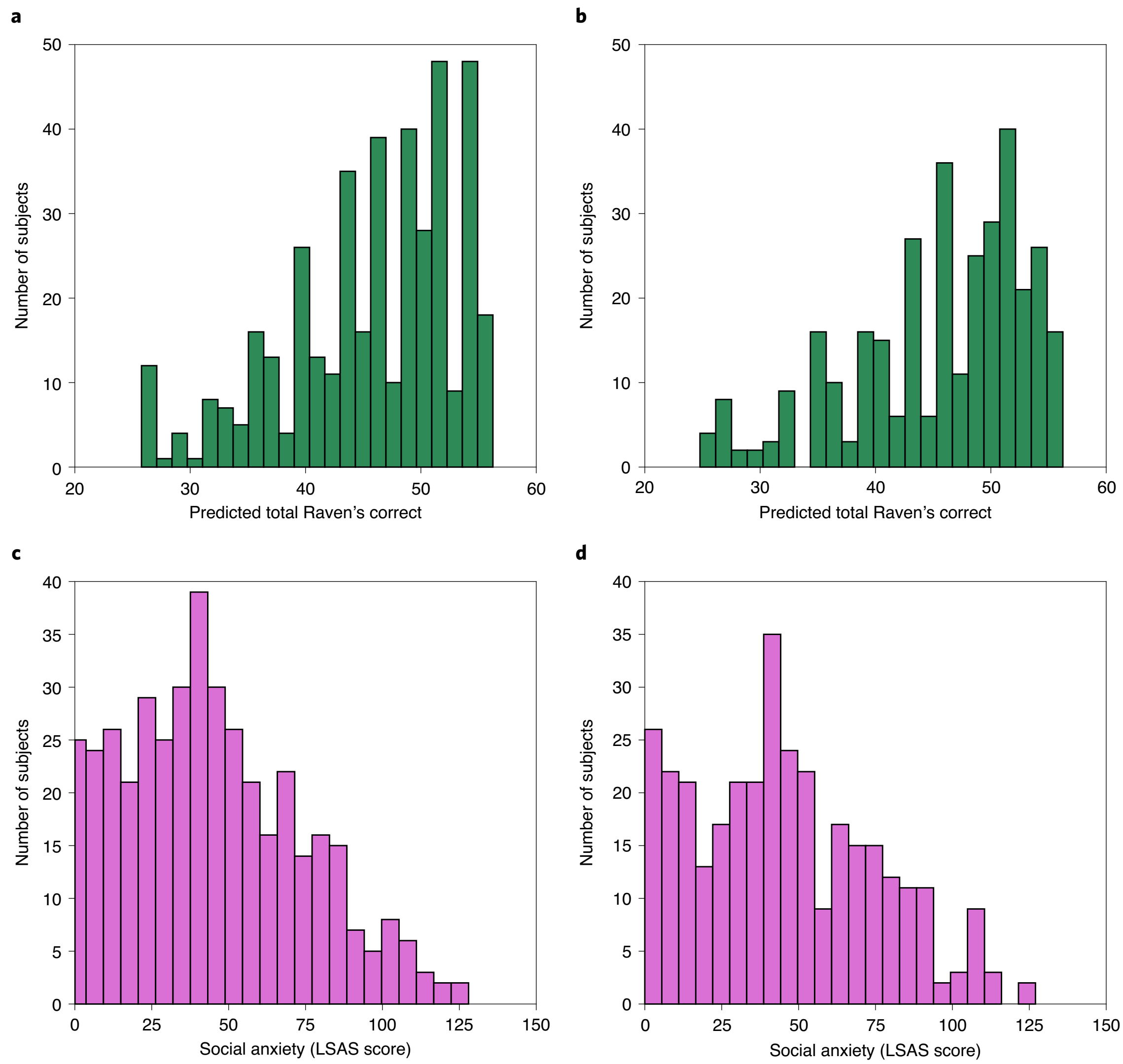

Fig. 2 | Distributions of scores. a–d, Raven's scores based on subjects' responses on the nine-item abbreviated version of the RSPM (**a**,**b**) and social anxiety scores based on LSAS (**c**,**d**) in experiment 1 (**a**,**c**, $N = 412$) and experiment 2 (**b**,**d**, $N = 331$).

best characterize the observed choices. The group-level distributions of these are shown in Supplementary Table 2, and are similar across the two experiments.

Our main hypothesis was that higher self-reported LSAS scores would be associated with increased reliance on model-based, counterfactual updating, reflected in larger estimated values of the parameter w. Indeed, the two variables were positively related (Fig. 3 and Supplementary Table 3), that is, higher self-reported social anxiety (LSAS) predicted greater use of counterfactual updating (w) in the patent race task in experiment 1 ($t(409) = 2.59$, $P = 0.01$, $\beta = 0.026$, 95% confidence interval 0.006 to 0.045) and experiment 2 ($t(328) = 2.5$, $P = 0.01$, $\beta = 0.034$, 95% confidence interval 0.007 to 0.060). These results also control for any effects of abstract reasoning as approximated by each participant's Raven's matrix score, which is included in the same regression as an additional explanatory variable. We also formally tested whether the effect of LSAS on w differed across experiments and found no significant effect ($t(737) = 0.50$, $P = 0.62$, $\beta = 0.008$, 95% confidence interval −0.024 to 0.040; Supplementary Table 4). No correlations for other model parameters, with either LSAS or Raven's score, were significant across both experiments, although a consistent trend was seen for a positive relationship between Raven's score and inverse temperature β (Supplementary Table 3; experiment 1, $t(409)$, $P = 0.078$, $\beta = 0.039$, 95% confidence interval −0.004 to 0.083; experiment 2, $P < 0.001$, $\beta = 0.12$, 95% confidence interval 0.053 to 0.187). Since that parameter captures the overall noisiness of responses, this suggests that higher intelligence is accompanied by a generalized improvement in task performance, less specific than the shift towards higher w associated with LSAS.

We next sought to unpack the intuition behind the main finding using a simpler, more theory-agnostic analysis[10]. The foregoing model characterizes the choices as determined by expected values that are learned incrementally by an error-driven running average over the outcomes (real versus counterfactual) of the series of earlier trials. By omitting the long-run dependencies over multiple trials so as to focus only on how each choice is affected by the single most recent trial's outcomes (real versus counterfactual), we can produce a more concrete, simplified approximation to the model[10]. Specifically, we performed a multinomial logit regression that explains each trial's move in terms only of the vector of payoffs possible given the opponent's move on the previous trial, comprising the payoff actually received and the counterfactual payoffs that would have been received had the played made other choices (Methods). Such a regression corresponds to a parametric

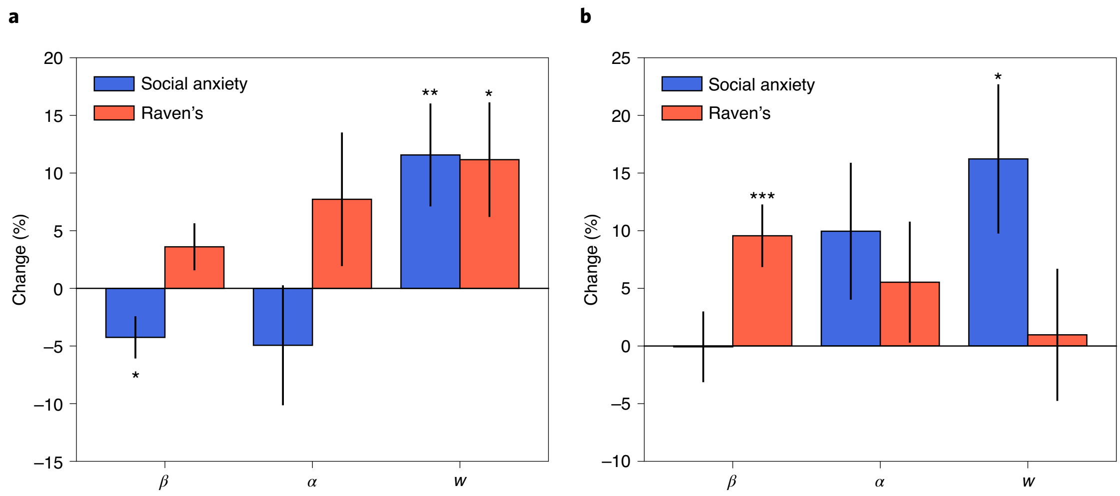

Fig. 3 | Change in EWA parameters as a function of LSAS and abbreviated nine-item Raven's matrices. a,b, Per cent change in EWA parameters in experiment 1 (**a**, $N = 412$) and 2 (**b**, $N = 331$). β represents the inverse temperature, α is the learning rate and w dictates the rate of counterfactual updating. The *y* axes indicate the per cent change in the parameter for each change of 1 s.d. in the predictor. Error bars indicate 1 s.e.; $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$.

limit of the full model (where the learning rate $\alpha=1$). Then, the free parameters are the overall inverse temperature $\hat{\beta}$ and the fractional weight of counterfactual to actual payoffs, $\hat{w}$. For $\hat{w}$ of 0, a choice is determined only by the (model-free, actually received) payoff for the previously chosen move; as $\hat{w}$ approaches 1, it becomes equally dependent on the (model-based, counterfactual) payoffs that would have been received for the other moves, given the opponent's previous move (Methods). Using this simplified approach, we found that LSAS was associated with the same direction of change in the counterfactual sensitivity $\hat{w}$. However, (as expected because the simplified analysis omits the effects of trials preceding the most recent one) the effect was not statistically significant in either dataset considered alone (experiment 1, $t(409) = 1.31$, $P=0.19$, $\beta=0.009$, 95% confidence interval −0.004 to 0.022; experiment 2, $t(328) = 1.18$, $P=0.24$, $\beta=0.007$, 95% confidence interval −0.004 to 0.018). The same effect was trending with both datasets pooled ($t(740) = 1.95$, $P=0.052$, $\beta=0.007$, 95% confidence interval -8×10^{-4} to 0.015; Supplementary Table 5).

Returning to the full EWA model, we next used an elaborated variant of the model to investigate the hypothesis that the effects of social anxiety would be driven by counterfactual updating about a subset of options. This model subdivides the parameter w into two, w^{+} and w^{-}, which govern counterfactual updating separately for options that would have been better or worse, respectively, than the one taken. We first verified that the elaborated model fitted choices better than the original one, using the Bayesian information criterion to correct for overfitting from the additional free parameters. Indeed, the integrated Bayesian information criterion (iBIC) score[3] for the extended variant was lower (indicating a better fit, correcting for overfitting due to additional free parameters) than that of the standard EWA model for each experiment (experiment 1, $\mathrm{iBIC}^{w}=30{,}656$, $\mathrm{iBIC}^{w\pm}=30{,}126$; experiment 2, $\mathrm{iBIC}^{w}=26{,}409$, $\mathrm{iBIC}^{w\pm}=26{,}159$).

If increased deliberation in social anxiety is driven by rumination about the events in the task, then the behavioural effects should reflect the biases of that rumination. In particular, post-event processing in people with high social anxiety involves an excess of upward counterfactual thoughts ('if only' thoughts about how the situation could have gone better[33,38]). Although we did not explicitly assess the content of counterfactual thought in this study, in the context of the value learning model hypothesized here, such a bias predicts a specific consequence for choices: the link between social anxiety and w should be due, more strongly, to a relationship between social anxiety and w^{+} (upwards counterfactual updating) more so than w^{-}. Indeed, social anxiety predicted a robust increase in upwards counterfactual updating, indexed by w^{+} (experiment 1: $t(409) = 2.66$, $P=0.008$, $\beta=0.031$, 95% confidence interval 0.008 to 0.053; experiment 2: $t(328) = 3.44$, $P<0.001$, $\beta=0.047$, 95% confidence interval 0.020 to 0.074), and had no significant relationship with w^{-} (Fig. 4 and Supplementary Table 6). To formally compare these effects, we tested whether the association between social anxiety and w^{+} was significantly greater than that for w^{-}; this difference was significant for experiment 2 (Supplementary Table 7; $t(163.64) = 2.36$, $P=0.019$, $\beta=0.040$, 95% confidence interval 0.007 to 0.074) and estimated in the same direction but not significant for experiment 1 ($t(362.17) = 1.54$, $P=0.124$, $\beta=0.022$, 95% confidence interval −0.006 to 0.050). Even at baseline (that is, at an average level of social anxiety), upwards counterfactual updating (indexed by w^{+}) was significantly higher than downwards (w^{-}) in both experiments (Supplementary Table 7; experiment 1: $t(397) = 13.2$, $P<0.001$, $\beta=0.187$, 95% confidence interval 0.159 to 0.214; experiment 2: $t(310) = 5.92$, $P<0.001$, $\beta=0.099$, 95% confidence interval 0.066 to 0.132).

We next examined whether our results were specific to social anxiety by controlling for additional psychopathological symptoms. In general, there are complex patterns of comorbidity among different mental illnesses, and the effects we observed might in principle be subserved by other factors. Notably, many of the symptoms we assess are common in other disorders, including depression and generalized anxiety. Moreover, a number of studies have linked deficits in goal-directed choice to compulsivity[16,18]. We used participants' responses to a larger battery of self-report symptom questionnaires included only in experiment 2 to examine how counterfactual reasoning in the patent race task related to symptoms of psychiatric conditions other than social anxiety. We summarized these using the transdiagnostic dimensions identified in ref. [16] using factor analysis on the same battery studied there. Using this method, we computed scores for each subject along three dimensions: 'anxious–depression', 'compulsive behaviour and intrusive thought' and 'social withdrawal' (the last corresponding largely to LSAS; see Supplementary Figs. 1 and 2 for factor loadings).

Even controlling for these other factors, upwards counterfactual learning predicted social anxiety (now captured by the 'social withdrawal' factor; Fig. 5 and Supplementary Table 8; $t(326)=3.23$,

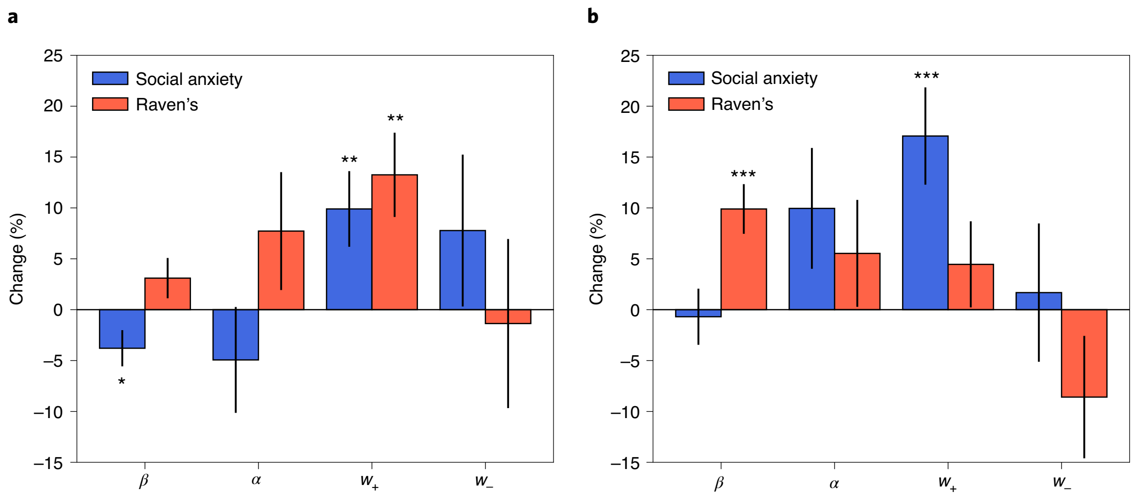

Fig. 4 | Change in EWA parameters as a function of LSAS and abbreviated nine-item Raven's matrices. a,**b**, Per cent change in EWA parameters for experiment 1 (**a**, $N=412$) and 2 (**b**, $N=331$). Here, β represents the inverse temperature, α is the learning rate and w, which dictates the rate of counterfactual updating, is split into w^{+} and w^{-}, which control upwards and downwards counterfactual learning, respectively. The y axes indicate the per cent change in the parameter for each change of 1 s.d. in the predictor. Error bars indicate 1 s.e.; $*P<0.05$; $**P<0.01$; $***P<0.001$.

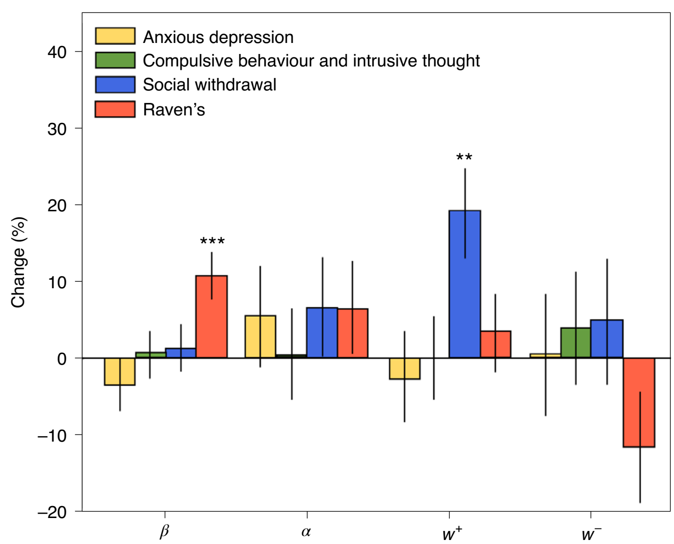

Fig. 5 | Per cent change in EWA parameters as a function of psychiatric symptom dimensions and Raven's (abbreviated nine-item Raven's matrices) for experiment 2 subjects ($N=331$). Here, β represents the inverse temperature, α is the learning rate and w, which dictates the rate of counterfactual updating, is split into w^{+} and w^{-}, which control upwards and downwards counterfactual learning, respectively. The y axes indicate the per cent change in the parameter for each change of 1 s.d. in the predictor. Error bars indicate 1 s.e.; $**P<0.01$; $***P<0.001$.

$P=0.001$, $\beta=0.049$, 95% confidence interval 0.019 to 0.078). The other psychiatric factors did not correlate significantly with any model parameters in this task (P values >0.1). Again, the relationship between 'social withdrawal' and counterfactual updating was significantly greater for upwards counterfactual updating w^{+} versus downwards counterfactual updating w^{-} (Supplementary Table 9; $t(138.37)=3.32$, $P=0.0012$, $\beta=0.061$, 95% confidence interval 0.025 to 0.097).

Last, we verified the robustness of our modelling approach by demonstrating the recoverability of the EWA parameters in simulated data for both the base and valenced model (Methods). Here, we simulated a dataset with known parameters and refit the model to the simulated dataset. We found that the ensemble of per-subject parameter values recovered from the simulated data tracked the true parameters for both models, with relatively small standard deviations in the per-subject errors around the parameters of interest (relative to ground truth) and little confusability across parameters (the latter quantified by correlation coefficients between the errors) (Supplementary Figs. 3–5).

Discussion

We investigated the relationship between social anxiety and model-based learning in two large-scale behavioural experiments. By fitting the parameters of the EWA model to subjects' choices in a patent race task, we derived indices reflecting the extent to which each participant's behaviour reflected valuation of actions via direct, model-free reinforcement versus model-based learning about other, counterfactual moves (and which ones). In line with our hypothesis, self-report social anxiety (LSAS) predicted a significant increase in overall model usage, as indexed by the EWA parameter w. Decomposed further, in one experiment, this effect was driven principally by increased updating of upwards counterfactual actions, suggesting a bias in the content of planning. The results were robust to the inclusion of additional dimensions of psychiatric symptoms as well as control for a measure of fluid intelligence. The latter was associated with overall choice consistency (parameter), perhaps reflecting a more generic contribution of motivation or task engagement.

Our results are consistent with longstanding suggestions that anxiety and other mental illnesses might be associated with excessive and/or biased deliberative (model-based) processing, such as uncontrolled forward search for action valuation[3,19–22]. For instance, theoretical work on value-based computation in depression has analogized rumination to mental simulation, for computing the value of potential actions given their anticipated consequences. Although in principle this would be expected to give rise to other, behavioural effects, there has so far been limited evidence connecting these hypothesized computations to actual choices or learning behaviour in illness[3,23]. Results of this sort are tantalizing in part because they point the way toward grounding the symptoms of mental illness, here potentially including overthinking in and avoidance of social situations, in the dysfunction of well-characterized, more basic neuro-computational mechanisms of evaluation and learning. Such a mechanistic understanding of these illnesses, known in medicine as an aetiology, is currently lacking, a situation that hampers both diagnosis and treatment, and ultimately contributes to poor patient outcomes.

Indeed, computational modelling proposes that many seemingly disparate aspects of anxiety disorders, including both cognitive symptoms such as worry and behavioural ones such as avoidance, may reflect a common deficit in action evaluation, which in turn may be due to dysregulated or biased mental simulation[20,22]. Thus, though operating at a different level of analysis, these theories are similar in spirit to the classic cognitive programme in psychiatry, which aims to identify misbeliefs or schemas[39,40] that mediate symptoms, including behavioural and somatic ones[41]. Although we do not directly assess many of the real-world symptoms that these processes hypothetically underlie in social anxiety – an important question for future work – a strength of the theories' formal grounding is that it allows us rigorously to test an abstracted laboratory operationalization of the core hypothesized cognitive–behavioural relationship: that in anxiety, altered choice behaviour reflects enhanced processing of certain events (counterfactual action outcomes).

While these considerations have been argued to relate to anxiety disorders broadly (and aspects of them even to depression as well), the present study concerns social anxiety as assessed with the LSAS questionnaire (which primarily measures avoidance of social situations). Which aspects of our results, if any, bear on social anxiety specifically? Some evidence that the effects might be specific to social anxiety comes from our finding, in experiment 2, of no similar associations of task behaviour with a factor comprising other depression and anxiety symptoms. This is not determinative, because we chose a socially framed RL task to highlight any effect of social anxiety specifically. Social anxiety is also enriched in the sampled population (giving us relatively better power to observe its effects), and simpler and more homogeneous as a disorder compared with more complex syndromes such as depression. All that said, substantively differential results for social anxiety versus other disorders might also arise insofar as overthinking in social contexts is, in some sense, goal directed and task focused, as opposed to worry and repetitive thought in other disorders which may be more idle and distracting (M. Paulus & M. Stein, personal communication[42]). Future work comparing multiple tasks within the same cohort, and manipulating task framing while holding task structure fixed, will be required to fully address these issues.

Our finding also complements previous reports (albeit using a different, single-player MDP task) that symptoms of compulsive disorders are associated with the opposite imbalance: declines in model-based learning[16,18]. However, in the current task we found no evidence that compulsivity is associated with reduced model-based learning, in contrast to Gillan's setting. The negative finding regarding compulsion is surprising given that OCD in particular has been associated with reduced model-based or goal-directed learning across a range of other tasks, beyond two-step Markov decision tasks.[14,15] Presumably, something about how decision-making is operationalized in the current task accounts for the difference. Other than the social framing, differences from the two-step task include that the task is nonsequential and differences in the parameterization of the model, such that the parameter w here isolates the strength of model-based updating, whereas the analogous parameter most strongly affected in two-step tasks, β^{MB}, also incorporates an element of choice consistency, like β in the current model.

One advantage of the current task, relative to others such as two-step MDPs, is that it allows us to compare the strength of model-based learning for different actions and outcomes. This allows us more finely to investigate individual steps of model-based planning, with implications both for testing process-level theories of planning and for characterizing in more detail how these operations change with illness. Much earlier research has viewed 'model-based' learning as all-or-nothing, exhaustive recomputation of action values over a tree of future states[6] and characterized individual differences in the overall tendency to deploy it[16]. But recently, more realistic process-level accounts are emerging that decompose planning into component steps that update individual actions in the light of individual outcomes[8,30,31,43]. Such theories emphasize that, even for healthy choice, since each step takes time and occupies resources such as working memory, they must be judiciously prioritized. One such prediction is that 'upwards counterfactual' information should be prioritized for learning, for example because it signals actionable choice policy improvements[8,44,45]. Our results here support this prediction, in that, even at baseline levels of social anxiety, planning is stronger for upwards than downwards counterfactuals (that is, w^+ is greater than w^-; see also refs. [46–48] for related results in different settings).

Finally, this type of process-level decomposition of planning may provide a useful and mechanistic level of description for differentiating patterns of biased event processing that are characteristic of different illnesses. Our finding that, in one experiment, social anxiety is associated more strongly with increased processing of upwards counterfactuals (over and above the baseline difference) seems the most likely of our results to be a feature of social versus other anxiety: for instance, it echoes the tendency for post-event rumination in social anxiety to also focus on such events[33]. That (in one of the experiments) the social anxiety effect more strongly enhances the bias also seen at baseline (which is in turn arguably rational), rather than being more sloppily associated with both types of counterfactuals, may also suggest that it reflects an extreme case along an otherwise adaptive spectrum. Although we do not assess rumination or worry directly (and this is an important direction for future work), its narrow focus makes it an appealing counterpart to theories of selective or biased model-based simulation. If these two constructs do coincide, then rumination would also lead to behavioural effects (such as avoidance); and the patterns of both could vary across illnesses. For instance, it seems likely that (a different pattern) of biased contemplation of negative events occurs in post-traumatic stress disorder, and in turn that that, if this were in fact planning, it would result in other symptoms such as overgeneralization of fear and avoidance[20,22]. This line of investigation points toward the promise of a better basic understanding of the microstructure of planning (which events are contemplated, when and why) and its relevance for understanding in a new computational and functional context many psychiatric phenomena, including not just compulsivity and rumination but also craving, obsession and hallucination.

Methods

Participants and procedures. Overall, 1,000 participants (500 per experiment) were recruited online using Amazon Mechanical Turk (AMT) in two successive experiments, of whom 966 ($N_1 = 489$, $N_2 = 477$) provided a complete dataset. No statistical methods were used to pre-determine sample sizes, but sample size was chosen to be similar to previous studies, specifically Gillan et al.'s[16] experiment 1, which used a similar online design, and Set et al.'s[28], which used the same task. Subjects were required to meet the following conditions: 18 years or older, based in the United States (that is, had a US billing address with an associated US credit card, debit card or bank account), and had successfully completed most previous tasks on the Amazon platform (95% of previous tasks approved). All participants provided informed consent and were paid a base rate in addition to a bonus proportional to their (nominal) earnings during the reinforcement learning task. All procedures were pre-approved by Princeton University's Institutional Review Board and were in compliance with all relevant ethical regulations.

Via their web browser, all participants completed the LSAS (Liebowitz, 1987), an abbreviated nine-item version of the Raven's Standard Progressive Matrices (RSPM) test to capture fluid intelligence[37,49], and 80 rounds of a patent race game[27]. Procedures for the two experiments differed only in that subjects in experiment 2 completed a more comprehensive psychopathological assessment (in addition to LSAS) before proceeding to the Raven's test and patent race game. In particular, experiment 2 included a battery of 209 multiple-choice questions which gauged symptom severity across a range of disorders and constructs (the same battery used by ref. [16]): alcoholism (Alcohol Use Disorders Identification Test), apathy (Apathy Evaluation Scale), impulsivity (Barratt Impulsiveness Scale 11), eating disorders (Eating Attitudes Test), social anxiety (LSAS), obsessive–compulsive disorder (OCD) (Obsessive-Compulsive Inventory-Revised (OCI-R)), schizotypy (Short Scales for Measuring Schizotypy), depression (Self-Rating Depression Scale) and generalized anxiety (State Trait Anxiety Inventory). Item 13 on the Self-Rating Depression Scale was administered following the erroneous wording also used in ref. [16] where participants rated "I am restless and can't sleep" instead of the original item, "I am restless and can't keep still".

Patent race game. Subjects played 80 rounds of an asymmetric patent race task, a competitive, simultaneous move game in which a 'strong' player with more resources competes with a 'weak' player with fewer resources[26–28]. In each round of the game, subjects (who played the 'weak' role) were endowed with $4 and chose how much to invest (in integer dollars, $0–4) to obtain a $10 prize. The computerized opponent was endowed with $5 and thus held a stronger position. The rules of the game, including each player's endowments and the payoffs conditional on different moves, were common knowledge and remained fixed throughout the game. When a player invested strictly more than their opponent, they won the $10 prize on that round. (In case of a tie, no one received the prize.) Regardless of the outcome, players kept the uninvested portion of their endowment.

Subjects were advised that the computerized opponent's choices on each round were drawn randomly from a pool of choices made by previous human participants at that round of the game. Thus, although opponents were anonymous and unlikely to be encountered more than once in a row, they represented people at the same stage of progression through the task as the subject. Importantly, participants were not instructed as to the strategy mixture played by the opponents but could only discover it by trial and error.

Learning models. Since the distribution of opponents' moves was unknown to the subjects (and potentially non-stationary), the game presents a learning problem: finding which moves are most effective. Two leading models for this process in behavioural game theory correspond to model-free and model-based RL (though model-free RL is known in this literature simply as 'reinforcement learning' while model-based RL is referred to as 'belief learning'); a third model, known as EWA[29] characterizes behaviour by a weighted combination of these two strategies and has previously been used in a series of studies with this task[27,28,50].

The model-free rule is simple Q-learning. It maintains an expected value for each possible move, updated whenever a move is chosen according to the received payoff[51]. In its original formulation[52,53], belief learning turns on learning the opponent's move distribution (a model about the opponent's preferences or 'beliefs', updated each time their move is observed). With this and the payoff matrix, the expected payoffs for each of the players' responses can be computed. In fact, marginalizing the beliefs, the same payoff estimates for the player's moves can be updated in place at each time step, by updating each of them according to the reward that would have been received had the player chosen that move, given the opponent's move[29]. This approach can be viewed either as an algebraic trick for conveniently implementing the predictions of the model-based rule, or as a substantive hypothesis for how these computations might actually be implemented in the brain using counterfactual updates in place of the belief model. Similar approaches, which substitute replayed experience for a world model, have also been examined for other RL tasks such as spatial navigation[8,54].

In the context of this game, model-free and model-based learning make different predictions about how the subjective value of each strategy (that is, each possible investment amount) is updated with experience at each round. Consider a round in which the subject invests $4 and the opponent invests $2. Here, the subject's choice to invest $4 results in a total return of $10. Model-free learning would update the expected value of investing $4 by moving it closer to $10. In belief learning (implemented via counterfactual updating), the subject further updates the value of each other move by calculating the return it would have yielded on the previous trial given the observed investment made by the opponent. For example, in this case, the subject will update the value of investing $2 toward $2 and the value of $3 toward $11, since these are the amounts the subject would have won given that the opponent invested $2.

Computational modelling. Following previous work[26,27,55], we modelled subjects' learning on this task using Camerer and Ho's[29] EWA learning model. This constitutes a weighted combination of the two learning strategies discussed above, analogous to hybrid model-based/free models previously used for human choices in MDPs[10].

To highlight this relationship, we re-derive the EWA model here starting from the hybrid model[10] and using that model's notation. At each round t, we assume that participants estimate the action value $Q_t(a)$ for each of the five moves ($a \in$ invest $0–4). They then choose their move softmax in these estimates, $P(a_t = a) \propto \exp(\beta Q_t(a))$, with inverse temperature β.

The core assumption of the hybrid model is that these net action values themselves arise from the combination of two estimators, model-based and model-free: $Q_t(a) = wQ_t^{\mathrm{MB}}(a) + (1-w)Q_t^{\mathrm{MF}}(a)$, weighted by a weighting parameter w.

Model-free learning is accomplished by error-driven updating of the chosen action a_t's value according to its obtained payoff r_t, with learning rate parameter α: $Q_{t+1}^{\mathrm{MF}}(a_t) = Q_{t+1}^{\mathrm{MF}}(a_t) + \alpha\left(r_t - Q_t^{\mathrm{MF}}(a_t)\right)$. Rearranging, and incorporating decay for the unchosen options[16,56], we have for all moves a,

$$Q_{t+1}^{\mathrm{MF}}(a) = (1-\alpha)\,Q_{t+1}^{\mathrm{MF}}(a) + I_{a,a_t}\alpha r_t \tag{1}$$

The indicator I_{a,a_t} is 1 for $a = a_t$; 0 otherwise.

Instead of learning actions' payoffs directly, model-based methods learn the distribution of the actions' more proximal consequences. Here, these are the opponent's moves a^{opp}, played with probability $\pi(a^{\mathrm{opp}})$, and the associated rewards, that is the matrix $R(a, a^{\mathrm{opp}})$ of the participant's payoffs for each move pair. Action values are then computed in expectation over these quantities. Here, for each for each a,

$$Q_t^{\mathrm{MB}}(a) = \sum_{a^{\mathrm{opp}}} \pi_t\left(a^{\mathrm{opp}}\right) R\left(a, a^{\mathrm{opp}}\right) \tag{2}$$

The opponent move distribution is also learned by error-driven update, $\pi_{t+1}(a^{\mathrm{opp}}) = (1-\alpha)\,\pi_t(a^{\mathrm{opp}}) + \alpha I_{a^{\mathrm{opp}},a_t^{\mathrm{opp}}}$, for all a^{opp}. (R is taken as given, since it is instructed.)

As written, equation (3) implies full recomputation of each model-based action value $Q^{\mathrm{MB}}(a)$ by exhaustive enumeration at every round, which seems unrealistically cumbersome (for example, if computing a value requires time or working memory resources[57]). Recent attempts to produce a more plausible process-level theory with a clearer neural implementation have noted that Q^{MB} can equivalently be carried over between rounds, by maintaining the single set of net values Q and incorporating updates to them in-place from (potentially limited) steps of model-based evaluation, also mixed in-place with direct experiential updates from model-free learning[8,54].

In this task, substituting the update for π into the expression for Q^{MB}, the net update on each round for each a is $Q_{t+1}^{\mathrm{MB}}(a) = Q_t^{\mathrm{MB}}(a) + \alpha\left(R\left(a, a_t^{\mathrm{opp}}\right) - Q_t^{\mathrm{MB}}(a)\right)$ (Camerer and Ho, 1999). The rule is thus an error-driven update on each move's value, but according to the reward $R\left(a, a_t^{\mathrm{opp}}\right)$ that would have been obtained had that move been selected. For the move that was actually selected, a_t, this equals the obtained reward r_t and the update is the same as for Q^{MF}; for the others, it is the same learning rule but using the 'counterfactual reward'. Further substituting both MF and MB updates into Q,

$$Q_{t+1}(a) = (1-\alpha)\,Q_t(a) + \alpha\theta R\left(a, a_t^{\mathrm{opp}}\right) \tag{3}$$

with

$$\theta = 1,\ \ a = a_t$$

$$\theta = w,\ \ a \neq a_t$$

for all a. Here $\theta = 1$ for the chosen move, and $\theta = w$ for all other moves. Thus, the hybrid model is equivalent to model-free error-driven experiential learning (equation (1)) augmented with additional counterfactual updates weighted by w.

To obtain the full EWA model, we first omit the learning rate α from the second term of equation (3). (This rescales the values Q by a factor of $1/\alpha$, which does not change the model's behavioural predictions since the free softmax temperature β can rescale to cancel it; this change of variables improves model identifiability by reducing collinearity between β and α.) Next we introduce an experience counter N, incremented and decayed at each step as $N_{t+1} = \rho N_t + 1$. This is used to scale the learning updates dynamically:

$$Q_{t+1}(a) = \frac{(1-\alpha)\,N_t Q_t(a) + \theta R\left(a, a_t^{\mathrm{opp}}\right)}{N_{t+1}} \tag{4}$$

Note that, for $\rho = 0$, $N(t) = 1$ for all t, and the new term has no effect. For $\rho > 0$, N accumulates and drives a type of learning rate decay. For simplicity, and following previous work[28], we took the two decay parameters as equal, that is $\rho = (1-\alpha)$. Finally, we introduce free parameters for the initial values $Q_0(a)$ (for $a = 0–4$), capturing any a priori preferences for the actions.

A second advantage of the counterfactual updating approach, viewed as a realizable process-level account of model-based learning, is that it provides a plausible process-level grounding for the parameter w: if counterfactual updates are limited, such that only a subset occur on any particular trial with some overall fractional update rate w, then equation (3) produces the expected amount of updating. In turn, this approach extends to the possibility that updates for some options are prioritized, resulting in different effective w (refs. [8,31]). Here, we considered the possibility that 'upward' and 'downward' counterfactual learning might be differentially prioritized, by using an elaboration of the EWA model in which, w, the degree of model-based updating, is valence dependent. This parameterization splits the free parameter w into w^+ and w^-, which control upwards and downwards counterfactual learning, respectively. The reward received on a given round, $R\left(a_t, a_t^{\mathrm{opp}}\right)$, serves as the reference point. Moves that would have resulted in more reward than what was won in reality (based on the opponents' move on that round) are updated in proportion to w^+, whereas actions that would have resulted in less reward are weighted instead by w^-. Accordingly, the update rule for values is the same as before, but with

$$\begin{aligned} &\theta = 1,\ a = a_t \\ &\theta = w^+,\ a \neq a_t \text{ and } R\left(a, a_t^{\mathrm{opp}}\right) > R\left(a_t, a_t^{\mathrm{opp}}\right) \\ &\theta = w^-,\ a \neq a_t \text{ and } R\left(a, a_t^{\mathrm{opp}}\right) < R\left(a_t, a_t^{\mathrm{opp}}\right) \end{aligned} \tag{5}$$

Note that we recover the original model when $w^+ = w^-$. In total, the original model contains eight free parameters and the valence-dependent model contains nine free parameters, summarized in Supplementary Table 10.

We estimated the free parameters of the model (equation (4), using the two different definitions of θ) per-subject using an expectation-maximization optimization algorithm[3] implemented in the Julia language[58]. This procedure maximizes the likelihood of each individual's sequence of choices, where each individual's parameter estimates are random effects drawn from group-level Gaussian parameter distributions, whose means and variances are also estimated jointly with the individual estimates. (For w and α, which range between 0 and 1, the model assumed that the Gaussian-distributed variable was transformed using a sigmoidal function, the unit normal cumulative distribution function, to obtain a parameter in the appropriate range.) We estimated the full hierarchical model separately for each experiment.

We next subjected the per-subject parameter estimates from these analyses to a set of multiple linear regressions, one for each parameter, to test whether LSAS or symptom factor scores (see below) predicted each of the free parameters from the EWA models. All tests are two-sided with an alpha value of 0.05. These models included an additional covariate to control for Raven's matrices scores, using the score predicted for the full 60-item set based on the nine problems given[37]. All predictors were standardized (z scored) for interpretability of coefficients. (Auxiliary analyses also considered age as an additional nuisance predictor, but this had no appreciable effects on the results.) To test whether the effects were different between parameters w^+ and w^-, we repeated the regression with both parameters as dependent variables (one each per participant, with an indicator variable for parameter type), and tested whether the slope on LSAS (or factor 3: social withdrawal) interacted with parameter type (valence). Here, we used a linear mixed-effects model (with random effects per participant on all of the parameters excluding the main effects of LSAS and Raven's) to capture the repeated-measure structure. We estimated this model using MATLAB's 'fitlme' function and the Satterthwaite approximation to the degrees of freedom.

To assess the relative fit of the elaborated versus standard variant of the EWA model while correcting for overfitting due to both group- and subject-level parameters, we computed iBIC scores[3]. This was defined as the marginal likelihood of the data given either model, aggregated across subjects, marginalizing per-subject parameters with the Laplace approximation, and penalizing for the group-level parameters using BIC.

One-trial-back EWA. We specified a simplified variant of the EWA model to complement the main findings with a more theory-agnostic analysis. Following prior work[10], we also considered a simpler, reduced model explaining the move on each trial based on the events on just the preceding trial. By eliminating the learning rate α (which otherwise enters non-linearly), the reduced model takes the form of a generalized linear model, here a multinomial logit regression onto the (five-valued multinomial) move. Moreover, the predictor variables of the simplified model have a more concrete, intuitive interpretation. In the one-trial-back version of the EWA model (taking $\alpha = 1$ and omitting initial Q values for the options, which here affect only the first trial), the only free parameters are the inverse temperature $\hat{\beta}$ and the rate of counterfactual updating $\hat{w}$. The predictors for choice a_t are $\hat{\beta}\theta R\left(a, a_{t-1}^{\mathrm{opp}}\right)$, where $\theta = 1$ for $a_t = a^{t-1}$ and $\theta = \hat{w}$ otherwise. In other words, the choice is logistic in the actual reward for the previous trial's choice and the vector of counterfactual rewards for the other choices, given the previous trial's choice, where $\hat{w}$ controls the relative regression weight for the counterfactual predictors. Thus, as $\hat{w}$ approaches 0, choice is based on the (model-free) received reward on the previous trial; as it approaches 1, the estimated value of each alternative investment amount is also given by the amount of reward the learner would have received on the previous trial given the opponent's observed strategy on the previous trial.

Parameter recovery. To assess the recoverability of the EWA model parameters, we use the model fitting procedure described above to recover the model parameters of artificial datasets for which the generative, ground-truth parameters are known.

To generate these artificial data, we simulated the process of collecting a new dataset corresponding as closely as possible in size and distribution to our existing dataset (comprising experiments 1 and 2 pooled). To determine the distribution of ground-truth per-subject parameters, we first estimated the group-level multivariate Gaussian parameter distributions for the pooled dataset ($N = 743$) using the model fitting procedure described above. We then drew parameters for each of a population ($N = 743$) of new subjects from this distribution and, for each, simulated 80 rounds of patent race investments. Next, we repeated the same model fitting technique that was applied to the real data to estimate the group- and subject-level parameter values for the synthetic dataset.

Finally, we compared ground truth versus recovered parameters per simulated subject, using Pearson's correlations and the variance and covariance of the recovery errors (ground truth − recovered).

Because our fits to the actual experimental dataset found that the parameter w^- in the valenced model was, on average, small, we repeated the simulation-and-recover analysis to verify that larger w^- would have been detectable if present. For this, we simulated another population of participants, the same as before but with the distribution of w^- modified to be uniform in (0, 1). The results (Supplementary Fig. 5) verify adequate recovery of larger w^- values in this case.

Factor analysis. For experiment 2, we used the factors identified in ref. [16] to reduce responses on the 209 items from the nine psychiatric symptom scales to scores on three dimensions that capture much of the intersubject variance. These were labelled by Gillan et al.[16] as 'anxious–depression', 'compulsive behaviour and intrusive thought' and 'social withdrawal' based on the items with the strongest loadings for each factor (Supplementary Fig. 1). We verified that the factor analysis procedure described in ref. [16], when applied to our data, produced substantially the same factor structure (correlations between factor loadings: factor 1: $R = 0.94$, $P < 0.001$; factor 2: $R = 0.91$, $P < 0.001$; factor 3: $R = 0.91$, $P < 0.001$). Because Gillan et al.'s study analysed the same battery of questionnaires using a larger sample ($N = 1{,}413$), we used the factor loadings estimated in that study to construct factor scores for each subject.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Processed data (per-participant estimated model parameters and covariates) supporting all of the statistical results of the study, and the raw choice data from which the model parameters were estimated, are available at https://github.com/ndawlab/patentrace. Raw psychometric data (questionnaire responses) are available from the corresponding authors upon request.

Code availability

Custom MATLAB code to reproduce all statistical results and tables is available at https://github.com/ndawlab/patentrace. Custom Julia code for estimating learning model parameters from raw choice data is available at https://github.com/ndawlab/em. Additional code (for figures and analyses of psychometric data) is available from the corresponding authors upon request.

Received: 16 January 2019; Accepted: 8 July 2021;
Published online: 16 August 2021

References

1. Maia, T. V. & Frank, M. J. From reinforcement learning models to psychiatric and neurological disorders. *Nat. Neurosci.* **14**, 154–162 (2011).
2. Huys, Q. J. M., Maia, T. V. & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).
3. Huys, Q. J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* **8**, e1002410 (2012).
4. Moutoussis, M., Eldar, E. & Dolan, R. J. Building a new field of computational psychiatry. *Biol. Psychiatry* **82**, 388–390 (2017).
5. Montague, P. R., Dolan, R. J. & Friston, K. J. Computational psychiatry. *Trends Cogn. Sci.* **16**, 72–80 (2012).
6. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
7. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007).
8. Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *Nat. Neurosci.* **21**, 1609–1617 (2018).
9. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
10. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–15 (2011).
11. Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. The curse of planning. *Psychol. Sci.* **24**, 751–761 (2013).
12. Doll, B. B., Bath, K. G., Daw, N. D. & Frank, M. J. Variability in dopamine genes dissociates model-based and model-free reinforcement learning. *J. Neurosci.* **36**, 1211–22 (2016).

13. Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481–1489 (2005).
14. Gillan, C. M. et al. Disruption in the balance between goal-directed behavior and habit learning in obsessive–compulsive disorder. *Am. J. Psychiatry* **168**, 718–726 (2011).
15. Gillan, C. M. et al. Counterfactual processing of economic action–outcome alternatives in obsessive–compulsive disorder: further evidence of impaired goal-directed behavior. *Biol. Psychiatry* **75**, 639–646 (2014).
16. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5**, e11305 (2016).
17. Reiter, A. M. F. et al. Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *J. Neurosci.* **36**, 10935–10948 (2016).
18. Voon, V. et al. Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* **20**, 345–52 (2015).
19. Huys, Q. J. M., Daw, N. D. & Dayan, P. Depression: a decision-theoretic analysis. *Annu. Rev. Neurosci.* **38**, 1–23 (2015).
20. Gagne, C., Dayan, P. & Bishop, S. J. When planning to survive goes wrong: predicting the future and replaying the past in anxiety and PTSD. *Curr. Opin. Behav. Sci.* **24**, 89–95 (2018).
21. Solway, A., Lohrenz, T. & Montague, P. R. Loss aversion correlates with the propensity to deploy model-based control. *Front. Neurosci.* **13**, 915 (2019).
22. Zorowitz, S., Momennejad, I. & Daw, N. D. Anxiety, avoidance, and sequential evaluation. *Comput. Psychiatry* **4**, 1 (2020).
23. Faulkner, P. et al. A comparison of "pruning" during multi-step planning in depressed and healthy individuals. *Psychol. Med.* https://doi.org/10.1017/S0033291721000799 (2021).
24. Shapiro, D. N., Chandler, J. & Mueller, P. A. Using Mechanical Turk to study clinical populations. *Clin. Psychol. Sci.* **1**, 213–220 (2013).
25. Norton, A. R. & Abbott, M. J. Self-focused cognition in social anxiety: a review of the theoretical and empirical literature. *Behav. Change* **33**, 44–64 (2015).
26. Rapoport, A. & Amaldoss, W. Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge. *J. Econ. Behav. Organ.* **42**, 483–521 (2000).
27. Zhu, L., Mathewson, K. E. & Hsu, M. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proc. Natl Acad. Sci. USA* **109**, 1419–1424 (2012).
28. Set, E. et al. Dissociable contribution of prefrontal and striatal dopaminergic genes to learning in economic games. *Proc. Natl Acad. Sci. USA* **111**, 9615–9620 (2014).
29. Camerer, C. & Ho, T. H. Experience-weighted attraction learning in normal form games. *Econometrica* **67**, 827–874 (1999).
30. Cushman, F. & Morris, A. Habitual control of goal selection in humans. *Proc. Natl Acad. Sci. USA* **112**, 13817–13822 (2015).
31. Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D. & Dolan, R. J. Experience replay is associated with efficient nonlocal learning. *Science* **372**, eabf1357 (2021).
32. Kocovski, N. L., Fleming, J. E., Hawley, L. L., Ho, M. H. R. & Antony, M. M. Mindfulness and acceptance-based group therapy and traditional cognitive behavioral group therapy for social anxiety disorder: mechanisms of change. *Behav. Res. Ther.* **70**, 11–22 (2015).
33. Kocovski, N. L., Endler, N. S., Rector, N. A. & Flett, G. L. Ruminative coping and post-event processing in social anxiety. *Behav. Res. Ther.* **43**, 971–984 (2005).
34. Crump, M. J. C., McDonnell, J. V. & Gureckis, T. M. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* **8**, e57410 (2013).
35. Liebowitz, M. R. Social phobia. *Mod. Probl. Pharmacopsychiatry* **22**, 141–73 (1987).
36. Mennin, D. S. et al. Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale. *J. Anxiety Disord.* **16**, 661–673 (2002).
37. Bilker, W. B. et al. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* **19**, 354–69 (2012).
38. Brozovich, F. & Heimberg, R. G. An analysis of post-event processing in social anxiety disorder. *Clin. Psychol. Rev.* **28**, 891–903 (2008).
39. Beck, A. T., Emery, G. & Greenberg, R. L. Anxiety disorders and phobias: a cognitive approach (Basic, 1985).
40. Clark, D. A. & Wells, A. in *Social Phobia: Diagnosis, Assessment, and Treatment* (eds Heimberg, R. G. et al.) (Guildford Press, 1995).
41. Hofmann, S. G., Carpenter, J. K. & Curtiss, J. in *Science and Practice in Cognitive Therapy: Foundations, Mechanisms, and Applications* (eds Leahy, R. L. et al.) 124–141 (Routledge, 2018).
42. Watkins, E. R. Constructive and unconstructive repetitive thought. *Psychol. Bull.* **134**, 163–206 (2008).
43. Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc. Natl Acad. Sci. USA* **113**, 12868–12873 (2016).
44. Icard, T., Cushman, F. & Knobe, J. On the instrumental value of hypothetical and counterfactual thought. *Proceedings of the 40th Annual Conference of the Cognitive Science Society* 517–522 (2018).
45. Caplin, A., Dean, M. & Leahy, J. *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy* Working paper no. 23652 (NBER, 2017).
46. Kahneman, D. & Miller, D. T. Norm theory. Comparing reality to its alternatives. *Psychol. Rev.* **93**, 136–153 (1986).
47. McCloy, R. & Byrne, R. M. J. Counterfactual thinking about controllable events. *Mem. Cogn.* **28**, 1071–1078 (2000).
48. Loomes, G. & Sugden, R. Regret theory: an alternative theory of rational choice under uncertainty. *Econ. J.* **92**, 805 (1982).
49. Raven, J. The Raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* **41**, 1–48 (2000).
50. Zhu, L., Jiang, Y., Scabini, D., Knight, R. T. & Hsu, M. Patients with basal ganglia damage show preserved learning in an economic game. *Nat. Commun.* **10**, 802 (2019).
51. Erev, I. & Roth, A. E. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **88**, 848–881 (1998).
52. Cheung, Y. W. & Friedman, D. Individual learning in normal form games: some laboratory results. *Games Econ. Behav.* **19**, 46–76 (1997).
53. Brown, G. W. Iterative solution of games by fictitious play. *Act. Anal. Prod. Allocation* **13**, 374–376 (1951).
54. Sutton, R. S. Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bull.* **2**, 160–163 (1991).
55. Amaldoss, W. & Jain, S. David vs. Goliath: an analysis of asymmetric mixed-strategy games and experimental evidence. *Manag. Sci.* **48**, 972–991 (2002).
56. Vikbladh, O. M. et al. Hippocampal contributions to model-based planning and spatial memory. *Neuron* **102**, 683–693.e4 (2019).
57. Keramati, M., Dezfouli, A. & Piray, P. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput. Biol.* **7**, e1002055 (2011).
58. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).

Acknowledgements
The authors thank Q. Huys, M. Paulus, M. Stein, A. Solway and S. Zorowitz for helpful conversations. This work was supported by NIMH grant R01MH121093, part of the CRNCS programme, and by a Scholar Award from the James S. McDonnell Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions
L.E.H., E.A.M., C.M.G., M.H. and N.D.D. contributed to the conception and design of the experiment L.E.H. and E.A.M. collected the data. L.E.H., E.A.M. and N.D.D. analysed the data. L.E.H. and N.D.D. prepared the initial draft of the manuscript, and all authors edited the manuscript and gave final approval of revisions.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01180-y.

Correspondence and requests for materials should be addressed to L.E.H. or N.D.D.

Peer review information *Nature Human Behaviour* thanks Camilla Nord and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Hunter, Daw

Last updated by author(s): 7/6/2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Participants were recruited via Amazon Mechanical Turk, filled out self-report psychiatric questionnaires via web forms, and played an economic game (implemented in JavaScript) in their browser.
Data analysis	Self-report data were analyzed using standard factor analysis libraries in R; game choices were fit with a learning model using expectation maximization code available at https://github.com/ndawlab/em implemented in Julia; fit parameters were regressed on self-report data using standard regression libraries in Matlab

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Both raw choice data and processed data (estimated subject parameters and covariates, sufficient to reproduce the statistical results and tables in the paper) are available at https://github.com/ndawlab/patentrace

Matlab code for reproducing the statistical analyses is included in the same repository.

Julia code for the EM model-estimation package used to estimate the subject parameters is available at https://github.com/ndawlab/em

Additional code and data are available from the authors upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a study comparing self report psychiatric symptom data to choice behavior in an economic game.
Research sample	Two general population samples (N=500 each) were recruited from Amazon Mechanical Turk. Subjects were required to be 18 years or older, based in the USA (i.e. had a US billing address with an associated US credit card, debit card or bank account), and to have successfully completed most previous tasks on the Amazon platform (95% of previous tasks approved). Expt. 1, 58% male, age mean=35.2 SD=10.4; Expt. 2, 55% male, age mean 35.8 SD=10.1.
Sampling strategy	A convenience sample was recruited via Amazon Mechanical Turk according to the above requirements. No formal power analyses were used to pre-determine sample sizes, but sample size was pre-chosen to be similar to previous studies, specifically Gillan et al's (2016) Expt. 1 which used a similar online design and Set et al. (2014) which used the same task.
Data collection	Data were obtained via forms and interactive games running in the participants' web browsers.
Timing	Experiment 1 was collected in July 2016. Experiment 2 was collected in April 2017.
Data exclusions	In line with recommendations for studies conducted using Amazon's Mechanical Turk (AMT), a priori exclusion criteria were applied to ensure data quality by eliminating low-effort participants. Of the 966 participants (Expt. 1, N = 489; Expt. 2, N = 477) who completed the task, we eliminated participants who shirked either the Ravens matrix test (Expt. 1, N = 36; Expt. 2, N = 49) or the patent race task (Expt. 1, N = 41; Expt. 2, N = 97). (Thus in total, 77 and 146 were removed from experiment 1 and experiment 2 respectively, where the larger number in the second study is likely due to the longer session). Specifically, we removed from consideration subjects who got 0 or 1 items correct on Raven's matrices or who chose the same move on more than 95% of trials (i.e. >76/80 rounds) in the patent race.
Non-participation	Of the 500 subjects recruited for each study, most completed (Expt. 1, N = 489; Expt. 2, N = 477). The remainder did not complete the web-browser tasks and return the "hit".
Randomization	n/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	See above.
Ethics oversight	The study was approved by the Princeton University IRB.

Note that full information on the approval of the study protocol must also be provided in the manuscript.