RESEARCH   **Open Access**

Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis

Quentin JM Huys[1,2,3,4*], Diego A Pizzagalli[5], Ryan Bogdan[6] and Peter Dayan[1]

Abstract

Background: Depression is characterised partly by blunted reactions to reward. However, tasks probing this deficiency have not distinguished insensitivity to reward from insensitivity to the prediction errors for reward that determine learning and are putatively reported by the phasic activity of dopamine neurons. We attempted to disentangle these factors with respect to anhedonia in the context of stress, Major Depressive Disorder (MDD), Bipolar Disorder (BPD) and a dopaminergic challenge.

Methods: Six behavioural datasets involving 392 experimental sessions were subjected to a model-based, Bayesian meta-analysis. Participants across all six studies performed a probabilistic reward task that used an asymmetric reinforcement schedule to assess reward learning. Healthy controls were tested under baseline conditions, stress or after receiving the dopamine D_2 agonist pramipexole. In addition, participants with current or past MDD or BPD were evaluated. Reinforcement learning models isolated the contributions of variation in reward sensitivity and learning rate.

Results: MDD and anhedonia reduced reward sensitivity more than they affected the learning rate, while a low dose of the dopamine D_2 agonist pramipexole showed the opposite pattern. Stress led to a pattern consistent with a mixed effect on reward sensitivity and learning rate.

Conclusion: Reward-related learning reflected at least two partially separable contributions. The first related to phasic prediction error signalling, and was preferentially modulated by a low dose of the dopamine agonist pramipexole. The second related directly to reward sensitivity, and was preferentially reduced in MDD and anhedonia. Stress altered both components. Collectively, these findings highlight the contribution of model-based reinforcement learning meta-analysis for dissecting anhedonic behavior.

Keywords: Anhedonia, Major depressive disorder, Depression, Reinforcement learning, Reward learning, Prediction error, Computational, Meta-analysis, Reward sensitivity, Learning rate

Background

Anhedonia is one of the cardinal symptoms for a clinical diagnosis of major depressive disorder (MDD; [1-3]) and refers to an inability to experience pleasure or a diminished reactivity to pleasurable stimuli. It is typically measured by verbal reports. In people subjectively reporting anhedonia, reward feedback objectively has less impact in a variety of behavioural tasks [4-15]. However, modern accounts of decision-making distinguish structurally different ways in which this reduction might be realized,

and precisely which of these is associated with anhedonia remains unclear.

Here, we attempt to distinguish two critical factors. The first factor is a reduction in the *primary sensitivity* to rewards. This is possibly the closest behavioural equivalent to the notion of a reduction in consummatory pleasure. Instruments measuring anhedonia, such as the relevant subscores of the Beck Depression Inventory [16] or the Mood and Anxiety Symptom Questionnaire (MASQ) [17] typically focus on this factor [18]. The second factor is an alteration in participants' ability to *learn* from reward feedback. This is emphasized by preclinical animal models of depression [19-22] and, because of the

*Correspondence: qhuys@cantab.net
[1]Gatsby Computational Neuroscience Unit, UCL, London, UK
[2]Wellcome Trust Centre for Neuroimaging, UCL, London, UK
Full list of author information is available at the end of the article

close association between dopamine (DA) signalling and reward learning [23-27], by neurobiological accounts linking anhedonia to DA [11,14,28-35]. It is most important to separate these factors, since they are likely to be associated with radically different ætiologies and therapeutic routes.

The distinction between these two factors is sharp in the mathematical formulation of reward learning based on prediction errors that underpins the account of DA activity [23-27], and is itself based on principles articulated in psychological [36] and engineering [37,38] theories of learning. Consider an experiment in which a reward of a given magnitude is given stochastically on some trials: where $r_t = 1$ if the participant did receive the reward on trial t, and $r_t = 0$ if it did not. We write ρ for the value the participant assigns to this reward. The participants are assumed to build and maintain an expectation ($\mathcal{Q}_t$) of the average reward it might gain on this trial, by means of a prediction error on trial t which is the difference $\delta_t = \rho r_t - \mathcal{Q}_t$ between the obtained ρr_t and expected $\mathcal{Q}_t$ reward. This prediction error can be used to improve the expectations adaptively [39] by adding the error to the previous expectation $\mathcal{Q}_{t+1} = \mathcal{Q}_t + \epsilon \delta_t$, where $0 \leq \epsilon \leq 1$ is a learning rate. This reduces the prediction error over time, at least on average.

The critical factors that might be associated with anhedonia are the two parameters ρ and ϵ in the above expressions. First, ρ is a measure of primary reward sensitivity – the larger ρ, the more sensitive the participant is to a reward of a given magnitude, or the greater the internal worth of an external reward. We comment on the difference between liking and wanting rewards later [40]. Second, the term ϵ governs how reward prediction errors influence learning. Evidence suggests that the phasic activity of DA cells is roughly proportional to δ_t; thus, alterations in the amount of DA released per spike, or in the sensitivity of postsynaptic receptors should behave like a change in the *learning rate* ϵ and affect the speed at which reward affects behaviour [41].

These quantities are formal parameters of a reinforcement learning rule. The question thus arises whether they can actually be distinguished in experimental practice. In this paper, we focus on objective measures of learning behavior. Crudely, since ρ only affects one term in δ_t, whereas ϵ controls the effect of the whole of δ_t on $\mathcal{Q}_{t+1}$, these quantities are theoretically distinguishable. A change in reward sensitivity leads to the asymptotic average value of $\mathcal{Q}_t$ being different. However, a change in learning rate alters the speed at which this asymptote is reached. More particularly, if the behavioural impacts of anhedonia were due to a DA deficit, the extent of anhedonic symptoms might correlate preferentially with the learning rate ϵ. If instead its effects were due to a change in primary reward sensitivity, then one would

expect the impact to concentrate on the reward sensitivity ρ. To examine this, we re-analyzed what is perhaps the most substantial body of data on reward learning in anhedonia, namely the probabilistic reward task of (Figure 1A-B). Behaviour in this task has previously been quantified by dividing it into three blocks and examining the evolution of a response bias across the blocks (Figure 1C). However, such measures cannot easily disentangle ρ and ϵ: Figure 1D-E shows that varying both parameters can roughly qualitatively explain the observed patterns. However, there are subtle differences to do with the asymptote of learning that provide us with a window of opportunity.

The fact that varying either parameter can lead to similar qualitative patterns shows that the two parameters play partially replaceable roles and may not be fully separable [42]. Any separation is likely to require substantial amounts of data. We therefore here fitted trial-by-trial individual reinforcement learning (RL) models in a meta-analytic manner to a series of six experiments involving 392 experimental sessions across 313 different participants. Model fitting allows for comprehensive tests of the ability of each hypothesis to account for the entire dataset. Bayesian model comparison ensures that the conclusions are based on parsimonious accounts of the data, avoiding overfitting and overly complex explanations [42-46].
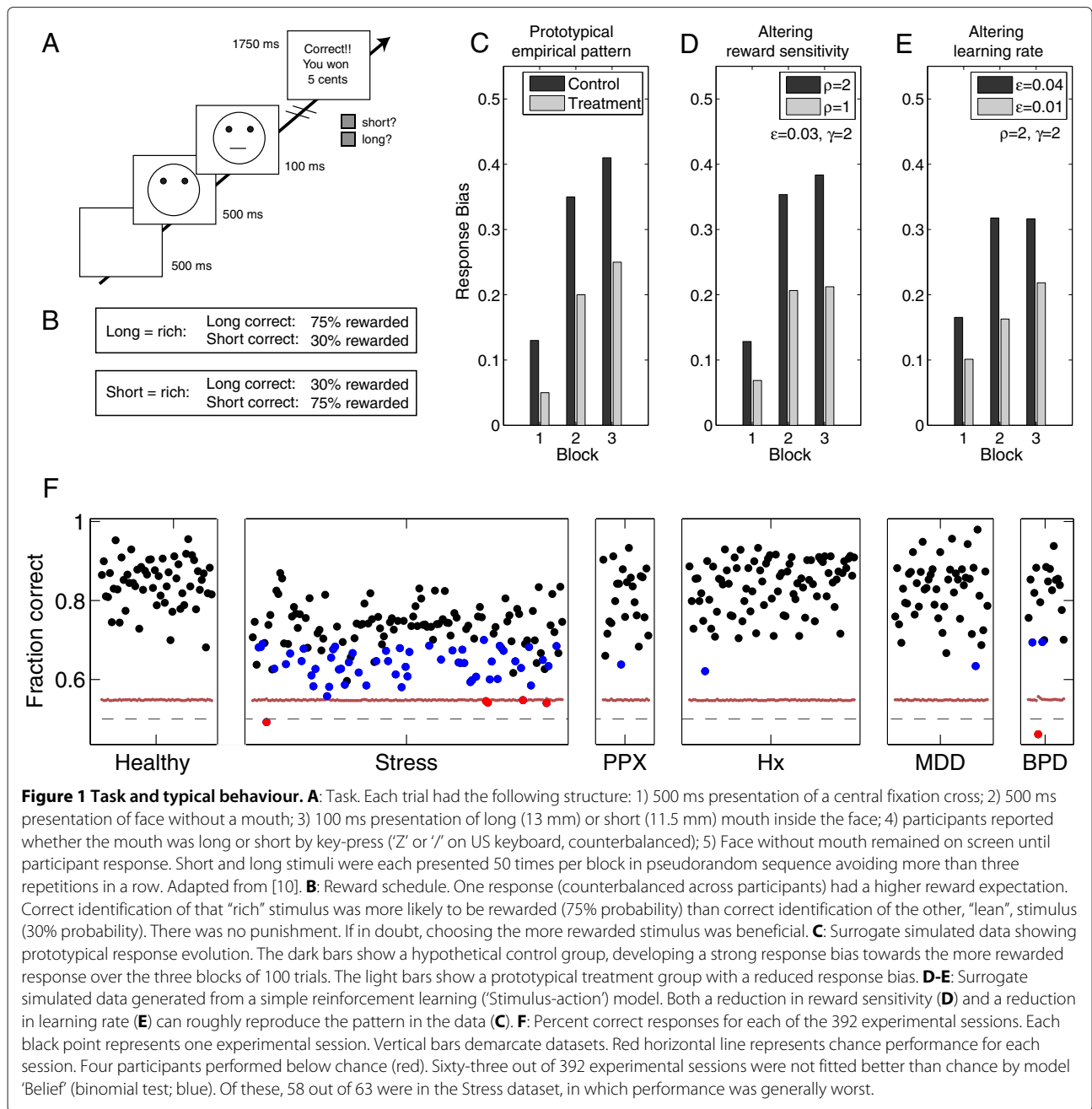
To maximise the chance of identifying specific contributions of learning rate and reward sensitivity, we jointly analysed a series of datasets that are likely to differentially affect the two parameters (Table 1). Three of these probed anhedonia in the context of depression; one in bipolar disorder. A further dataset probed the effect of a dopaminergic manipulation, which we expected to primarily affect ϵ. A final dataset probed the effect of stress, which is prominently involved in the pathogenesis of depression, and which we have previously suggested may reduce phasic DA bursts via an increase of tonic DA release [47].

Our main result is that measures of anhedonia and depression preferentially affected the reward sensitivity ρ, while a dopamine manipulation by pramipexole mainly affected the learning rate ϵ. Stress, however, affected both reward sensitivity and the learning rate. There was no difference in these two parameters in euthymic bipolar individuals, or those with a past history of depression.

Methods
Task and data
In this paper, we re-analyse 392 sessions of behavioural data derived from a probabilistic reward task ([10]; adapted from [55]), which is displayed schematically in Figure 1A. Central to the task is that an asymmetrical reinforcement scheme was used to induce a response bias:

Figure 1 Task and typical behaviour. A: Task. Each trial had the following structure: 1) 500 ms presentation of a central fixation cross; 2) 500 ms presentation of face without a mouth; 3) 100 ms presentation of long (13 mm) or short (11.5 mm) mouth inside the face; 4) participants reported whether the mouth was long or short by key-press ('Z' or '/' on US keyboard, counterbalanced); 5) Face without mouth remained on screen until participant response. Short and long stimuli were each presented 50 times per block in pseudorandom sequence avoiding more than three repetitions in a row. Adapted from [10]. **B**: Reward schedule. One response (counterbalanced across participants) had a higher reward expectation. Correct identification of that "rich" stimulus was more likely to be rewarded (75% probability) than correct identification of the other, "lean", stimulus (30% probability). There was no punishment. If in doubt, choosing the more rewarded stimulus was beneficial. **C**: Surrogate simulated data showing prototypical response evolution. The dark bars show a hypothetical control group, developing a strong response bias towards the more rewarded response over the three blocks of 100 trials. The light bars show a prototypical treatment group with a reduced response bias. **D-E**: Surrogate simulated data generated from a simple reinforcement learning ('Stimulus-action') model. Both a reduction in reward sensitivity (**D**) and a reduction in learning rate (**E**) can roughly reproduce the pattern in the data (**C**). **F**: Percent correct responses for each of the 392 experimental sessions. Each black point represents one experimental session. Vertical bars demarcate datasets. Red horizontal line represents chance performance for each session. Four participants performed below chance (red). Sixty-three out of 392 experimental sessions were not fitted better than chance by model 'Belief' (binomial test; blue). Of these, 58 out of 63 were in the Stress dataset, in which performance was generally worst.

Correct responses to one stimulus, designated "rich", were more likely to be rewarded than correct responses to the other stimulus, designated "lean" (Figure 1B). No feedback was given on other trials, including incorrect trials, and no explicit information about the asymmetry was provided. Participants were explicitly encouraged to win as much money as possible, and so could benefit from reporting the rich, rather than the lean, stimulus when in doubt. One measure of the tendency to do this is the response bias [10]:

$$\frac{1}{2} \log \left(\frac{n(a_1|s_r) \, n(a_1|s_l)}{n(a_2|s_r) \, n(a_2|s_l)} \right) \tag{1}$$

where s_r and s_l indicate presentation of the rich and lean stimulus, respectively, a_1 and a_2 are the two possible key presses, and $n(a|s)$ is the number of times a particular choice was made in response to that stimulus. Each count n was augmented by $\frac{1}{2}$ to avoid numerical instabilities. Outlier trials with very short (< 150 ms) or very

Table 1 Datasets used

Name	Manipulation	Reference	Participants #
Healthy	High *vs* low BDI scores	[10]	57
	Healthy volunteers. Payment: US$5 and course credit.		
MDD	MDD *vs* controls	[48]	48
	23 participants during an episode of MDD and 25 controls matched for age, sex, education, ethnicity and marital status. MDD participants met DSM-IV criteria for MDD, had Hamilton Rating Scale for depression scores $\geq$ 17, and no other axis I comorbidities except for anxiety. Inclusion required a minimal drug-free period of 2 weeks. Payment: US$10/hour plus 5US$ average task earnings.		
Hx	History of MDD *vs* no history	[49]	85
	Currently healthy participants with and without a history of major depressive disorder (MDD). Participants received US$15/hr in compensation for their time, as well as their task "earnings" (on average, US$5).		
BPD	BPD *vs* controls	[50]	19
	Euthymic outpatients (matched to the same 25 controls as in dataset 'MDD'). The outpatients had a long-standing diagnosis of Bipolar Disorder, currently satisfying criteria on the Affective Disorder Evaluation, which contains modified SCID mood and psychosis modules. Patients were classified as euthymic if they currently Young Mania Rating Scale [51] score <12, and if their Hamilton Rating Scale for Depression [52] score was below 8. Exclusion criteria included other axis I disorders, a past history of MDD, substance abuse and ECT in the past 6 months. Participants were paid US$25 for participation and task earnings.		
PPX	Pramipexole *vs* placebo	[53]	24
	Healthy volunteers received either placebo or a single dose of the D2/D3 agonist pramipexole hydrochloride (PPX) 0.5 mg 2 hours prior to the task. At this low dose, PPX is thought to reduce phasic DA release through autoreceptor stimulation. Payment: US$ 40 for the pharmacological session and US$24.60 for the task session.		
Stress	Threat-of-shock acute *vs* no stressor	[54]	79(x2) +1
	Healthy volunteers took part in the task twice (one missing session), once in a no-stress condition and once in a stress condition. Participants were told that poor performance on the task might lead to a shock being delivered through electrodes attached to the back of their neck. In the stress condition, they were told that this was quite likely, whereas they were told that no shock would be delived in the no-stress condition. No shocks were actually delivered. Notably, the version of the task used in this study was more difficult, with the difference in size between long and short mouth being smaller. This resulted in fewer correct discriminations (see Figure 1F). Payment: either course credit or US$10/hour as well as money "won" during the task (US$10.60 on average).		

Full details of all the patient and control groups are provided in the original publications.

long (> 1500 ms) reaction times are excluded (see [10] for a full description of the 2-step procedure used to exclude trials with outlier responses). Figure 1F shows the fraction of correct responses for each of the 392 individual experimental sessions. In addition to the computer task, participants completed self-report questionnaires (see Table 2). The datasets and manipulations are shown in Table 1. Briefly, the studies examined the effect of i) depression (categorical diagnosis according to DSM-IV; continuous quantification based on self-report measures of depressive features and anhedonia; and past history of MDD); ii) bipolar disorder, currently euthymic (categorical diagnosis according to DSM-IV); iii) stress; and iv) low-dose D_2 agonist pramipexole. The low dose (0.5 mg) of pramipexole was assumed to reduce phasic DA bursts to unexpected rewards due to presynaptic (autoreceptor) activation [53,56]. We note that the dataset 'Stress' differs

from the others because a more difficult version of the task was used [54].

Reinforcement learning models

Reinforcement learning models account for every choice on every trial for every participant individually. Here, we describe the model for one particular participant. 'Weights' for emitting a particular choice are updated after every trial to predict the next choice. We consider a set of factors that might affect the weights, and use complexity-sensitive model comparison methods to try to identify the importance of each. Briefly, write a_t for the participant's choice on trial t (key '/' or 'z'), and $\bar{a}_t$ for the choice not taken on that trial (key 'z' or '/'). If stimulus s_t (long or short mouth) was presented, the model assigns to a_t a probability $p(a_t|s_t)$. This probability depends on the 'weights' $\mathcal{W}_t(a_t|s_t)$ and $\mathcal{W}_t(\bar{a}_t|s_t)$ assigned to each

Table 2 Sample characteristics: psychometric measures

Measure	N	mean	median	1st quartile	3rd quartile
BDI	366	8.3	6	1	11
BDA	281	1.9	0	1	3
BDI\A	281	6.9	1	5	10
Generalized distress depression (GDD)	276	23.2	16	20	30
Generalized distress anxiety (GDA)	276	18.8	15	18	24
Anxious anxiety (AA)	276	53.7	45	53.5	65
Anhedonic Depression (AD)	276	21.1	17	20.5	25
Full (BDI + BDA + MASQ subscores)	255				

BDI is the total Beck Depression Inventory II [16]. BDA stands for the anhedonic subscore of the BDI consisting of the sum of items 4 (loss of pleasure), 12 (loss of interest), 15 (loss of energy) and 21 (interest in sex). BDI\A is the total BDI score without the anhedonic component (BDI\A = BDI - BDA). Mood and Anxiety Symptom Questionnaire (MASQ) subscores [17] were anxious arousal (AA), anhedonic depression (AD), general distress anxiety (GDA) and generalised distress depression (GDD).

choice when presented with stimulus s_t. The mapping from weight to probability is made via a 'softmax' function so that a choice a_t will be expected to be emitted more frequently the bigger the difference between its weight and the weight of the alternative choice, or more specifically:

$$p(a_t|s_t) = \frac{1}{1 + \exp\left[-(\mathcal{W}_t(a_t, s_t) - \mathcal{W}_t(\bar{a}_t, s_t))\right]} \quad (2)$$

The choice weights themselves change over time (hence the subscript on $\mathcal{W}$) and are composed of several terms, whose contributions differ for the different models. The models are variants on an underlying full model called 'Belief', for which

$$\mathcal{W}_t(a_t, s_t) = \gamma \mathcal{I}(a_t, s_t) + \zeta \, \mathcal{Q}_t(a_t, s_t) + (1-\zeta) \, \mathcal{Q}_t(a_t, \bar{s}_t) \quad (3)$$

The first of these terms, $\gamma \mathcal{I}(a_t, s_t)$, depends on instructions: $\mathcal{I}(a_t, s_t) = 1$ if a_t is the instructed choice for stimulus s_t (for instance pressing 'z' for the long mouth) and is zero otherwise. The parameter γ thus determines the participants' ability to follow the instructions. The bigger γ, the larger the contribution from $\mathcal{I}(a_t, s_t)$, and hence the instructed response contributes more to choice. Importantly, this instructed choice is symmetric between rich and lean stimulus; thus this term leaves the asymmetry to the other terms.

The second and the third term depend on the expected reward $\mathcal{Q}_t(a_t, s_t)$. This captures the effect of the experienced rewards on previous trials, just as described in the introduction (except allowing different predictions for the different actions and stimuli). $\mathcal{Q}_t(a_t, s_t)$ depends on four factors: the binary sequence r_t up to that point in time, which indicates whether a reward was delivered or not, an initial $\mathcal{Q}_0$ value, the learning rate ϵ and the subjective (i.e. internal to the participant as opposed to the external magnitude in a fixed number of cents) effect size of a reward ρ, which we identify with reward sensitivity.

After every choice, this $\mathcal{Q}$ value is updated according to the prediction error $\delta_t = \rho r_t - \mathcal{Q}_t(a_t, s_t)$ as follows:

$$\mathcal{Q}_{t+1}(a_t, s_t) = \mathcal{Q}_t(a_t, s_t) + \epsilon \delta_t \quad (4)$$

That is, after every trial, the expected reward $\mathcal{Q}(a, s)$ for choice a for stimulus s is increased towards the subjective reward size ρ if a reward is received ($r_t = 1$) but the expectation $\mathcal{Q}$ was lower than ρ, and it is decreased towards zero if no reward was received ($r_t = 0$). The larger ρ, the larger the effect of rewards on choice propensities. As the learning rate ϵ approaches 1, learning is so fast that the $\mathcal{Q}$ values are simply the last experienced outcome for each choice-action pair. For $0 < \epsilon < 1$, expectations represent exponentially weighted averages over the recent outcome history. A multiplicative change to δ is equivalent to a change in ϵ.

In the task, the mouth is only shown for a very short period of time. Thus participants cannot be sure which stimulus was actually presented, and, as experimenters, we cannot know what the participants perceived. This uncertainty has two consequences. First, it implies that the factor γ which governs the effect of the instructions, should be less than ∞. Second, the participants will not be sure which value $\mathcal{Q}_t(a_t, \text{long})$ or $\mathcal{Q}_t(a_t, \text{short})$ and instruction weight $\mathcal{I}(a_t, s_t)$ to employ in their choice, or which $\mathcal{Q}$ value to update using Equation 4. We capture this effect by assuming that they *know* which stimulus-choice pair to update in terms of learning (Equation 4), but that when choosing, they use a form of Bayesian decision theory [57] to combine estimates based on both possibilities. That is, we use a parameter $0 \le \zeta \le 1$ to represent participants' average uncertainty about which stimulus was actually presented. Assume participants expected .75 unit reward for pressing button 'z' given the long mouth ($\mathcal{Q}(z, \text{long}) = 0.75$), and 0 given the short stimulus ($\mathcal{Q}(z, \text{short}) = 0$). If they now believed with a probability ζ that they had seen stimulus s and with a probability $1 - \zeta$ that they had

seen stimulus $\bar{s}$, then their expectation for pressing button 'z' would be $\zeta \mathcal{Q}(z, \text{long}) + (1-\zeta)\mathcal{Q}(z, \text{short}) = .75\zeta$. This is the contribution of $\mathcal{Q}$ in Equation 3. We write ζ as applying to the $\mathcal{Q}$ term only. We could also apply it to the instruction term $\mathcal{I}(a,s)$, but this would be equivalent to rescaling γ (see Additional file 1).

Equations 3 and 4 comprise the full model 'Belief'. We also considered two simpler variants, both of which had one fewer free parameter, and one more complicated variant, with an extra parameter. First, at $\zeta = 1$, participants are certain, and use the correct stimulus-action value to guide their choice. The model in which this value is forced is called 'Stimulus-action'. At $\zeta = 0.5$, they are indiscriminate between the two stimuli. This is model 'Action' because they learn only about the values of choices, independently of stimuli. Finally, the more complex model 'Punishment' is based on the possibility that participants might treat a non-reward as a punishment, so making $\delta_t = \rho r_t + \rho^-(1-r_t) - \mathcal{Q}_t(a_t, s_t)$, where $\rho > 0$. We mention other possible models in the discussion.

Model fitting & comparison

Bayesian model comparison at the group level and model fitting procedures are described in detail in [58]. The key equations are provided in the Additional file 1. Briefly, model fitting was performed by Expectation-Maximisation to find group priors and individual (Laplace) approximate posterior distributions for the estimates for each parameter for each participant. For models 'Action' and 'Stimulus-action', this comprised parameters $\{\mathcal{Q}_0, \rho, \epsilon, \gamma\}$; for model 'Belief' it additionally included parameter ζ, and for model 'Punishment' also ρ^-. All parameters were represented as non-linearly transformed variables with support on the real line and normally distributed group priors.

More complex models will often fit the data better because they have more freedom. However, model complexity is better assessed by methods other than counting parameters [43,59]. Bayesian model comparison is a principled way of assessing model parsimony by computing the posterior probability of each model given the entire dataset for all participants. Because exact computation of these quantities is intractable, individual parameters were integrated out by sampling from the fitted priors, and a standard Bayesian information criterion (BIC)-like approximation was employed at the group level [43]. This procedure results in a measure we term iBIC (for 'integrated BIC') which captures how well each model explains the data given how complex it is [58]. The smaller this number, the greater the model parsimony. The difference between two such values, ΔiBIC, is an approximation of the models' relative log Bayes factor.

The same principles of model comparison also apply to the categorical question whether two groups differ in terms of their parameters. That is, when asking whether group A and B differ in terms of their reward sensitivity ρ or in terms of their learning rate ϵ, the correct approach is to compute the posterior likelihood of models *incorporating* these hypotheses about group differences. That is, we computed pairs of models for each dataset: in the first model of each pair $\mathcal{M}_\rho$, which allows group differences in ρ, participants share a common prior for all parameters except for ρ, for which the two groups have separate priors. In the second model $\mathcal{M}_\epsilon$, participants similarly share a prior for all parameters except for ϵ. Computing the Bayes factors (ΔiBIC) values for these two models relative to each other indicates whether group differences in one or the other parameter provide a more parsimonious account of the *entire* dataset while taking into account the relative flexibility each parameter accords the model and interactions between parameters. The Bayes factors of these models relative to the original model with no group separation, indicate whether the groups significantly differ in either characteristic.

Regression analyses

After model validation, we first assessed inter-correlations between specific questionnaire measures (AD, BDA, GDD, BDI\A, AA, GDA) and reward sensitivity or learning rate in the entire sample using one multiple linear regression analysis for ρ and one for ϵ (`regstats.m` in Matlab V7.14). However, this analysis neglected two aspects of the data: first that the questionnaire data are likely correlated; and secondly that parameters for different participants are estimated with different degrees of confidence. We therefore ran a weighted hierarchical multivariate regression. This is equivalent to a standard hierarchical multivariate regression, except that parameters were weighted by the precisions with which they were estimated (see [60] for details). Note that parameters are represented in the transformed space throughout to avoid issues with non-Gaussianity.

Results

Model validation

We built a set of models that embody key hypotheses about the course of learning in the different groups and fitted them to the data. The models parameterize the *type* of learning performed by participants, allowing us to assess whether they attach rewards to stimulus-action pairs; or just to actions, or to a mixture of the two. We also test the status of 'no reward': do participants treat this outcome as a punishment, reducing the probability of the associated action, or do they treat it as a non-informative null outcome, as intended?

Since we are interested in understanding the characteristics of groups of individuals, we need to ascertain at the group, rather than at the individual level, which model

does best [43,45,58]. We indeed found that, taking suitable account of complexity, a single model did capture all groups satisfactorily, allowing for a common and interpretable semantics for the parameters. These so-called random effects models capture inter-subject variation in a way that allows for differences in the extent to which individual participants' data constrain the parameters. When we perform correlation analyses (below), we weigh parameters according to the precision with which they were inferred.

The results are shown in Figure 2A. The most parsimonious account of the data was provided by model 'Belief'. This conclusion rests on the group-level Bayes factors iBIC (see Methods and Additional file 1 for further details). This compares the approximate posterior probability of each model given the data to that of model 'Belief'. Two key aspects of this quantity are that 1) it punishes overly complex models; 2) it assesses this at the group, rather than the individual level. The higher this number, the poorer the combination of model fit and model simplicity. Differences above 10 are typically considered to be strong evidence for one model over the other [43].

Both the standard Rescorla-Wagner model 'Stimulus-action' with separate stimulus-choice values $\mathcal{Q}(s, a)$, and model 'Punishment', which treats trials on which no reward was given as punishing, performed poorly with much worse iBIC values. This suggests that participants were not able to treat the two stimuli as entirely separate; and that they did not treat the absence of rewards as punishments with aversive properties. Model 'Action', which assumes that participants learn only action values, and thus do not separate between stimuli at all in the asymmetric component of choice ($\zeta = 0.5$) was an improvement over the basic model 'Stimulus-action' ($\zeta = 1$) but came a distant second to the model 'Belief' (ζ inferred for each session). Indeed, the belief parameter ζ was broadly distributed around 0.5 (mean 0.53, standard deviation 0.13). That is, on average, participants behave as model 'Action', neglecting the stimuli when learning and forming expectations. Individually, however, there was variability in how well they were able to discriminate the stimuli.

Figure 1F shows the probability of performing a correct choice for all participants in all datasets, with those not predicted better than chance by model 'Belief' circled in blue. This model correctly predicted significantly more choices than chance (binomial test, $p < .05$) for most experimental sessions (329 out of 392; black dots). Due to a smaller difference between long and short mouth, and hence a more difficult perceptual discrimination problem (Table 1), participants in the stress dataset– whether in the stress or the no stress condition–showed on average a lower probability of correct choice. They were consequently less well predicted by all models, but model

'Belief' still gave the best account of these data too (see Additional file 1: Section S2.1). Figure 2B shows that, as intended, the overall probability correct was captured by the instruction sensitivity parameter γ (Pearson correlation 0.93, $p < 10^{-20}$), which in turn frees the other parameters to capture trial-to-trial variation in the behaviour contingent on the reward outcomes.

Finally, Additional file 1: Section S2.2 also provides the results of a resampling analysis. This shows that if the model is run on the task, it spontaneously generates data that looks similar to the experimental data.

Regression analyses
Given that model 'Belief' captured the data satisfactorily, we proceeded to analyse the relationship between model parameters and self-report questionnaire measures. Our aims were primarily to identify correlations between measures of anhedonia and learning rates or reward sensitivities. A standard, unweighted, multiple linear regression analysis revealed a significant negative correlation between ρ and anhedonic depression (AD), the most specific measure of anhedonic symptoms available to us ($p = 0.004$; uncorrected for multiple comparisons). There was no correlation between ρ and other questionnaire measures, and no correlation between the questionairre measures and the learning rate ϵ (all $p > 0.1$). Participants in the 'Stress' dataset were tested twice. Repeating this analysis using only the session without stress yielded similar results ($p = 0.007$ for the correlation ρ vs AD, all other $p > 0.09$).

As expected, questionnaire scores were substantially correlated (Figure 3A), and the parameters of different experimental sessions were inferred with varying certainty. We therefore additionally orthogonalized the three anhedonic measures (BDA, GDD, AD) with respect to all the three other measures (BDI\A, AA, GDA) and fit a weighted generalised linear model (GLM). Figure 3B shows that anhedonic depression AD remained significantly and negatively related to the reward sensitivity ρ ($p = 0.005$, Bonferroni-corrected for 8 comparisons). No other correlation survived correction for multiple comparisons (all $p > 0.1$). In particular, no measure of anhedonia was associated with the learning rate ϵ. Figure 3C shows a scatter plot of AD scores vs. reward sensitivity with dot size proportional to the weight (inverse variance) in the weighted regression analysis.

Next, there was a negative linear correlation between ρ and ϵ of -0.41 ($p < 0.0001$; Figure 3D). To further question the selectivity of the correlation between AD and ρ, we orthogonalized ρ with respect to ϵ in addition to orthogonalising as above. This again yielded an (unweighted) significant correlation of AD with ρ ($p = 0.004$, multiple weighted linear regression, uncorrected for multiple comparisons), with no other correlations significant. The
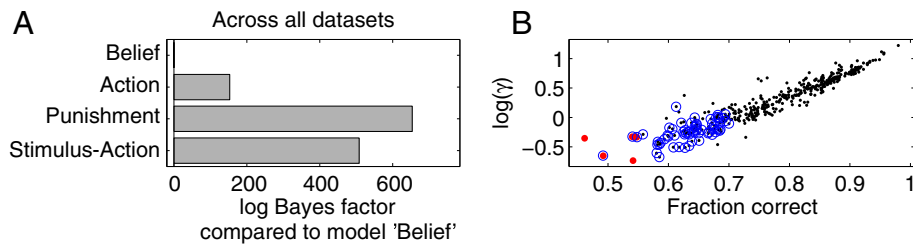
Figure 2 Model performance. A: Model comparison. Group-level log Bayes factors ΔiBIC for each model relative to model 'Belief' across all datasets. A difference ≥ 10 in this measure is strong evidence for the model with the lower score. **B**: The parameter γ in the model largely captures the probability with which participants made a correct choice. Note that, by design of the task, this explicitly captures the effect of symmetric instructions and perceptual difficulties, rather than the asymmetric effect of rewards.

reverse orthogonalization did not yield any significant correlations with ϵ.

At least part of the correlation between ρ and ϵ arises because the the two parameters can explain similar features of the data, i.e. alterations in one parameter can be compensated for by alterations in the other parameter (see Figure 1). To establish whether the association between AD and the reward sensitivity parameter was due to real features in the data, rather than due to inference issues, we asked whether the correlations with question-
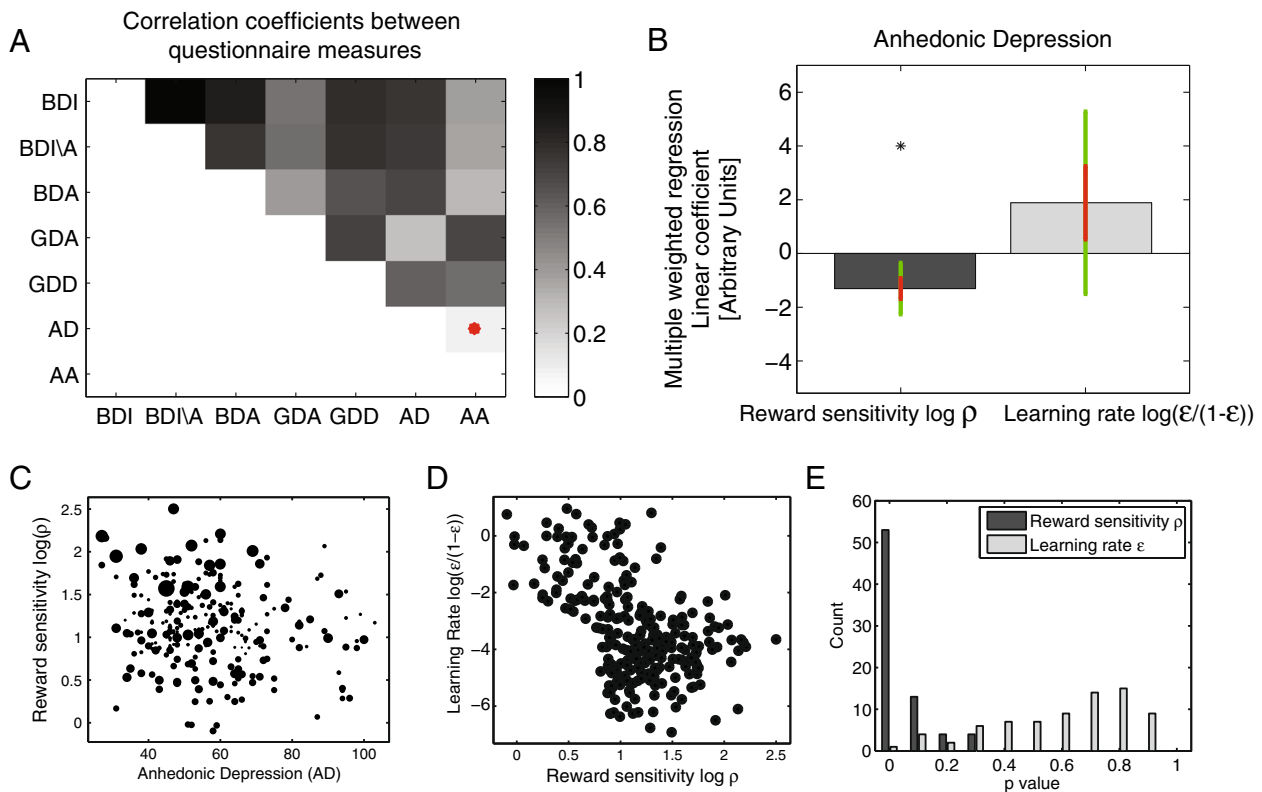
Figure 3 Correlates of anhedonia. A: Correlation coefficients for all pairwise correlations between questionnaire measures. All are highly significant ($p < .01$), except for the correlation between anhedonic depression and anxious anxiety, denoted by a red dot. **B**: Hierarchical weighted regression analysis across all datasets, involving all 255 participants with a full set of BDI, BDA and MASQ scores. The plots shows the linear coefficients between anhedonic depression (AD) score and the reward sensitivity and learning rate parameters ρ and ϵ. Each bars shows one linear coefficient; the red error bars indicate ± 1 standard error; and the green error bars indicate the 99.4% confidence interval (corresponding to a Bonferroni corrected level $p = .05/8$). AD is significantly and negatively correlated with the reward sensitivity ρ, but not significantly correlated with the learning rate ϵ. **C**: Scatter plot of anhedonic depression against reward sensitivity. Size of dots scale with weight (inference precision). **D**: Scatter plot of reward sensitivity vs. learning rate. **E**: Significance of correlations across parameter estimates from 70 surrogate datasets. There is a consistent and stably significant correlation between AD and reward sensitivity ρ, but not between AD and learning rate ϵ.

naire measures remained stable and identifiable in the surrogate data. For each of the 70 surrogate datasets of 392 experimental sessions, we repeated the standard, unweighted multiple linear regression. Figure 3E shows the distribution of p values for the correlation of AD with ρ and ϵ. While the median p value for the correlation between ρ and AD was 0.04, that for ϵ and AD was 0.71.

Finally, all correlation analyses using the reward sensitivity and learning parameters inferred from the second-best model 'Action' yielded the same results, showing that the results are not dependent on a particular model formulation.

Categorical comparisons

We next examined how learning rate and reward sensitivity were affected by the factors explored in each of the individual datasets. For each dataset, we compared two models: one which assumes that the two experimental groups differed in terms of ρ, the other in terms of ϵ.

Figure 4 shows the Bayes factors for models $\mathcal{M}_\rho$ compared with $\mathcal{M}_\epsilon$ for each of the individual datasets. Given the correlation results, we additionally performed a comparison between all participants with the 20% highest and lowest AD scores. Bayes factors above 20 (or below 1/20) are very strong evidence in favour of a hypothesis [43], while likelihood ratios of 3 to 10 (or the inverse) are weak evidence. We found strong evidence that MDD and AD were better accounted for by a change in reward sensitivity ρ, rather than learning rate ϵ. The opposite was true for pramipexole, which mainly acted by reducing the learning rate ϵ.

However, the Bayes factors comparing models $\mathcal{M}_\rho$ and $\mathcal{M}_\epsilon$ to the basic 'Belief' model without split parameters

were all ≤ 3. This means that there was no strong evidence that either of these parameters categorically separates any two groups. Stress appeared most likely to have a genuinely shared effect on both parameters. Participants with a history of depression, or currently euthymic bipolar participants showed weak evidence of reductions in reward sensitivity. Note that all these ratios are by necessity numerically more modest than those in Additional file 1: Figure S1 because they are inferred from far less data.

Discussion

Our results suggest that anhedonia (as measured by AD) and MDD affect appetitive learning more by reducing the primary sensitivity to rewards ρ than by affecting the learning rates. Non-significant trends for such an effect were observed for participants with a past history of depression, and amongst euthymic bipolar disorder patients. By contrast, a dopaminergic manipulation was found to preferentially affect the speed of learning, ϵ. Acute stress had no preferential effect on reward sensitivity or learning rate. These two parameters appear to be state, rather than trait measures: Neither a past history of depression, nor one of BPD, had a significant effect on ρ or ϵ.

Anhedonia

Two self-report measures of anhedonia were used in this paper: the anhedonic depression subscore of the MASQ questionnaire, and the anhedonic subscore of the BDI. The former was clearly related to the reward sensitivity ρ. This subscore quantifies participants' verbally expressed inability to experience pleasure, and so might be expected to capture something akin to consummatory 'liking' [40].

Figure 4 Comparing models incorporating hypotheses about group differences in terms of reward sensitivity ρ or learning rate ϵ. Each bar of the large panel shows the Bayes factor comparing $\mathcal{M}_\rho$ to $\mathcal{M}_\epsilon$. Green bars indicate very strong evidence for model $\mathcal{M}_\rho$ (Bayes factor ≥ 20), yellow bars weak evidence (Bayes factor $3 - 10$) and cyan bars insufficient evidence (Bayes factor < 3). MDD and high scores of anhedonic depression (AD) result in a reduction in ρ. Pramipexole instead reduces the learning rate. Stress has no differential effect on reward sensitivity compared with learning rate. Participants with a past history of depression, or euthymic individuals with bipolar disorder, showed weak evidence for a reduction in ρ. The insets at the top show the inferred priors over the relevant parameters of the winning model for each group.

Note, however, that various studies (e.g. [18,61,62]) have failed to find a direct correlate between anhedonia measures such as those employed here and participants' ratings of how pleasant a sucrose solution is—a widely-used animal model of anhedonia [63].

By contrast, in our paradigm, ρ falls squarely within the behavioural definition of 'wanting' whereby past experiences are collated into values that influence choice. One possibility is thus that our results reflect the way that 'liking' is coupled into 'wanting', with the reward having less impact when it occurred (ρ was lower), rather than the same amount of pleasure not being translated efficiently into anticipation (which would be the consequence of lowered ϵ). Such a reduction in reward sensitivity parallels reductions in emotional reactions seen in MDD in response to appetitive, but also aversive, stimuli [64].

At a neurobiological level, 'liking', has been linked to μ-opioid signalling [40,65-67] in NAcc shell. Recent PET studies have reported prefrontal changes in μ-opioid receptors in MDD [68], and μ-opioid receptors have been found to affect the response to antidepressant medications [69]. We also note that substance abuse disorders, including of opiates, are prominently co-morbid with depressive disorders [70,71], although it is not clear whether abuse causes depression or *vice versa*. It would be interesting to examine the effect of opiates on tasks like those examined here. If the primary reward sensitivity relates to "liking" and consummatory pleasure, then one would expect a preferential impact on ρ, unlike dopamine.

Dopamine in depression

The temporal difference learning model of phasic dopamine signalling posits that DA reports the prediction error δ. As mentioned in the introduction, a multiplicative change in this signal would be equivalent to a change in the learning rate ϵ. Although an alteration in the reward sensitivity ρ necessarily also leads to an alteration in the prediction error δ, the consequences of an abnormality that arises upstream of the prediction error are subtly different from an abnormality affecting the already computed prediction error signal. It is the latter that is most closely aligned with a primary alteration in dopaminergic function.

In the current study, we report tentative findings suggesting that pramipexole reduced the learning rate rather than affecting the reward sensitivity, which might correspond to a direct reduction in the signal reported by phasic DA. Although pramipexole is a non-ergot D_2/D_3 *agonist* (and is clinically used as such in Parkinson's disease, restless leg syndrome, and occasionally in treatment-resistant MDD), it has previously been found to have behavioural effects of a DA *antagonist* at the low doses (0.5 mg) used in our dataset [72-74]. Similarly, low doses of the D_2 agonist cabergoline have been found to

specifically reduce reward go learning [75]. With both drugs, it has been postulated that this may be due to activation of inhibitory [76] presynaptic D_2 receptors, which have a higher affinity for DA [77] and *reduce* phasic DA release ([78-82]; see also [56]) and DA cell firing [83]. Thus, our findings echo those of a recent study showing that dopamine agonists can reinstate prediction errors by restoring the component of the prediction error related to expected value, rather than that of reward [84]. However, we emphasize that the conclusions pertaining to dopamine rest on the contribution of only one study and that there are important technical caveats to be borne in mind (see below).

As indicated in the introduction, DA has multiple and profound involvements with depression. These range from the fact that DA manipulations affect mood [85] and that many antidepressants have pro-DA effects (including buproprion, sertraline, nomifensine, tranylcypromine amongst others; [86]), to a role in resilience [87,88], and suicidality in MDD [89,90]. Critically, depression is common in other disorders that also involve DA dysfunction, such as Parkinson's disease and schizophrenia [91,92]. However, the findings reported here speak to the specific aspects of phasic DA in learning, rather than to all the manifold ways in which DA may be involved in anhedonia and depression. Tonic DA levels, which are partially independent of phasic bursts of activity [93,94] have a profound impact on energy levels and vigour [95,96] making this feature of DA a likely candidate for the severe psychomotor retardation seen in melancholic depression [32], and to the distinction between pleasure and motivational aspects of anhedonia [18,97]. In the context of our task, this may relate to reaction times (albeit noting that these can also be affected by phasic DA signals; [98]), which were not analysed. Rather, here, we focused on the role of DA in instrumental learning of action values. Note, though, that the current design cannot completely disentangle instrumental effects from those due to Pavlovian approach [58,99]. Such Pavlovian influences involve dopamine and the nucleus accumbens (NAcc; [100-105]), and may indeed constitute one of the ways in which DA supports antidepressant function [106,107].

Overall, our findings are consistent with the fact that DA by itself is not a major target for psychopharmacological treatment of anhedonia [92], and, modulo the issues mentioned above, that DA appears to mediate 'wanting', more than 'liking' [40,108]. There has been a number of recent functional imaging investigations into reward-related decision making in MDD (e.g. [11,109-114]). Our findings have key implications for the interpretations of these studies, as it is important to separate contributions to the correlates of prediction errors that arise from changes in reward sensitivity from changes in learning rate. Because a) 'wanting' (reinforcement) and 'liking'

(hedonic) aspects of tasks will often overlap; and b) both changes in ρ and ϵ would affect correlations with a regressor based on prediction errors δ, our findings do not invalidate these imaging results, but call for studies and analyses that further separate them.

Analysis methods

Our conclusions were derived from a detailed, model-based meta-analysis combining behavioural data from six datasets in 392 sessions. Several features and limitations of this analysis deserve comment.

First, the interpretability of the parameters is maximised by using rigorous model comparisons. The Bayesian approach we used prevents overfitting by *integrating out* individual subject parameters [43,45,46,58]. Rather than just comparing how careful parameter tuning can allow a model to fit the data extremely well, we extensively sampled the model's entire parameter space, and ask how well, on average, the *class* of model, independent of its particular parameters, can fit the given data (for an insightful explanation, see chapter 28 of [59]). Next, our model comparison is done at the group level, rather than at the individual level, again through sampling and Bayesian model comparison. This ensures that conclusions about the parameters do apply to the group.

Second, our approach takes individual model fits into account. The regression analyses are true random effects analyses, weighting parameters by how strongly they are constrained by each participant's own choice data, and by how well the model fits that particular participant. This, in combination with the explicit modelling of stimulus uncertainty (beliefs) and instruction weights, ensures that any non-specific performance variability does not unduly affect our parameters of interest. Furthermore, the weighted regression ensures that each participant influences the conclusions proportionally to how well they are fit by the model.

Third, it is standard practice to constrain the parameters when fitting models, for instance to avoid extreme outlier inference. We use two types of constraints. The parameter transformations generate hard constraints that force parameters to remain inside feasible regions. The empirical Bayesian inference of the group priors additionally yields the most appropriate soft constraints [59,115].

Fourth, learning rate and reward sensitivity were correlated in all models tested. To alleviate this, we enforced independence at the group level. One standard approach would have been to compare models in which one dimension is constrained by forcing the parameters for all participants along that dimension to be equal. However, this a) is an unrealistic constraint; b) fails to address the fact that parameters may not give the model equal flexibility; and c) renders parameters hard to interpret as variability from one is squeezed into all the other parameters

(which further aggravates point b). To circumvent these issues, we contrasted the parsimony of models that explicitly allowed participants to fall into distinct groups. Doing this at the group level addressed the questions at the group level, which is where we sought to draw conclusions. We also performed the regression analysis in a number of ways to assess the selectivity of the relationship with ρ and its stability in the inference procedure.

Fifth, it is important to note that our failure to discover correlations between AD and learning rate, or between the effects of pramipexol and reward sensitivity might be due to limits of power. This is particularly true for the categorical comparisons on which arguments about the differential effect of dopamine and anhedonia or depression rest mainly. It will be critical to replicate these findings in larger samples, potentially requiring paradigms explicitly adapted to separating the two factors.

Sixth, we find no evidence that either learning rate of reward sensitivity clearly separates any of the groups when comparing the basic model to models $\mathcal{M}_\rho$ or $\mathcal{M}_\epsilon$ that allow for the two groups to have different means in one of the other parameter. However, our central motivation was whether anhedonia associated more with unusual ρ or unusual ϵ, because of the different psychiatric and psychological interpretations of these possibilities. The most direct test of this is to compare $\mathcal{M}_\rho$ to $\mathcal{M}_\epsilon$ rather than comparing each to a mid-point in the shape of the basic model. Furthermore, there are a number of caveats to this finding, which form the basis for reporting the more lenient comparison between models $\mathcal{M}_\rho$ and $\mathcal{M}_\epsilon$. First, the model comparison technique may be too conservative for this particular analysis. For instance, at the top level we very stringently punish according to the total number of observations. Although we have found this to overestimate the rate at which the group-level variance should drop with observations, in our hands this penalty has proven to yield the correct responses most reliably when tested on a variety of surrogate datasets (unpublished data in preparation). This is likely to be particularly important given that the power in the analyses in Figure 4 is much lower than that in the preceding analyses using the entire dataset. These considerations do not bear on comparisons between models $\mathcal{M}_\rho$ and $\mathcal{M}_\epsilon$. Furthermore, the panels showing the parameters in Figure 4 appear to show robust group differences. However, a separation in parameter space does not guarantee that the increase in fit outweighs the increase in model complexity.

Of course, even though our best fitting model did an excellent job predicting the data of a plurality of participants, there could be a model that we did not try that would do even better. This is particularly true of the participants in the Stress dataset, who were fit the worst. There is a conventional ANOVA-like procedure in these

circumstances that involves assessing the extent to which responses are potentially predictable [44,116] from running the same task multiple times. Unfortunately, it is not clear how to execute this in cases involving learning which is contingent on the participants' behaviour.

Alternative models

One advantage of the simplicity of the task is that it is likely insensitive to several aspects of reinforcement learning that are under current investigation, such as goal-directed *versus* habitual decision-making [117,118]. However, one interesting direction would be to consider a more Bayesian treatment of the learning process, according to which the effective learning rate ϵ should be a function of the amount of experience that the participant has had, modulated by the participant's belief that the contingencies are constant [119-121].

This elaborate account raises an important question about the factor ϵ in the models that we did build. Here, we considered the effect on learning of manipulating the magnitude of δ_t; but noted that this is exactly confounded in the magnitude of ϵ. Neurobiologically, though, there could be an alternative realization of ϵ, notably via cholinergic influences [122-124]. It would be interesting to examine the effect of cholinergic manipulations on the basic task.

Equally, in conventional reinforcement learning models, it is common to employ a variant of Equation 2 in which the terms $\mathcal{W}$ are multiplied by another arbitrary constant which is usually written β and called an inverse temperature. The larger β, the more deterministic the choice between a_t and $\bar{a}_t$, all else being equal. Thus β is often used as a surrogate for controlling exploration, as more stochastic choices (lower β) are more exploratory. However, in our model, β can substitute *exactly* for ρ (in a very similar way that ϵ can substitute for the magnitude of δ). Thus reward insensitivity could masquerade as over-exploration or (particularly in depression) as under-exploitation. These are not differentiable in this task. Nevertheless, the neurobiological realization of the exploration constant β has been suggested as being rather different – notably involving noradrenergic neuromodulation [125] – opening up another line of experimental investigation.

Finally, two relevant findings may be important for future model development. First, positive affective responses to positive events in daily life have recently been found to be *stronger* in depressed than in non-depressed individuals [126]. Second, the impact of negative events has been found to be determined not by the strength of the immediate emotional reaction, but by the attributions made about it [127]. Thus, it may be that the effects seen here can be modified by higher-order processes in ways

that are critical for the development of psychopathology, and it may be useful to extend the present methods to such interactions [60].

Conclusions

This paper presented a model-based meta-analysis of behavioural data spanning several related manipulations and adds to a growing literature of behavioural correlates of depression [64,128]. We concluded that anhedonia in depressive states was mediated by a change in reward sensitivity, which has different behavioural consequences from either stress or DA manipulations. Our analysis allowed us to draw these conclusions while taking into account as much as possible the variability between the datasets and participants; and the variability in other, unrelated aspects of task performance. We believe that similar analyses could be readily applied to other datasets. We hope that our findings will encourage the re-analysis of the dissociation between reward sensitivity and dopaminergic processes in depressive states.

Additional file

Additional file 1: Supplementary methods and results [43,63-65].

Author details
[1]Gatsby Computational Neuroscience Unit, UCL, London, UK. [2]Wellcome Trust Centre for Neuroimaging, UCL, London, UK. [3]Translational Neuromodeling Unit, University of Zurich and ETH Zurich, Switzerland. [4]Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital of Psychiatry, Zurich, Switzerland. [5]Department of Psychiatry, Harvard Medical School, MA, USA. [6]Department of Psychology and Neuroscience, Duke University, NC, USA.

References

1. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*: American Psychiatric Association Press; 1994.
2. World Health Organization: *International Classification of Diseases*: World Health Organization Press; 1990.
3. Myin-Germeys I, Peeters F, Havermans R, Nicolson NA, DeVries MW, Delespaul P, Os JV: **Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study.** *Acta Psychiatr Scand* 2003, **107**(2):124–131.
4. Costello CG: **Depression: Loss of reinforcers or loss of reinforcer effectiveness?.** *Behav Ther* 1972, **3**:240–247.
5. Akiskal HS, McKinney WT: **Depressive disorders: toward a unified hypothesis.** *Science* 1973, **182**(107):20–29.
6. Blaney PH: **Contemporary theories of depression: critique and comparison.** *J Abnorm Psychol* 1977, **86**(3):203–223.
7. Nelson RE, Craighead WE: **Selective recall of positive and negative feedback, self-control behaviors, and depression.** *J Abnorm Psychol* 1977, **86**(4):379–388.
8. Henriques JB, Glowacki JM, Davidson RJ: **Reward fails to alter response bias in depression.** *J Abnorm Psychol* 1994, **103**(3):460–466.
9. Henriques JB, Davidson RJ: **Decreased Responsiveness to reward in depression.** *Cogn Emotion* 2000, **14**(5):711–724.
10. Pizzagalli DA, Jahn AL, O'Shea JP: **Toward an objective characterization of an anhedonic phenotype: a signal-detection approach.** *Biol Psychiatry* 2005, **57**(4):319–327. [http://dx.doi.org/10.1016/j.biopsych.2004.11.026]
11. Steele JD, Kumar P, Ebmeier KP: **Blunted response to feedback information in depressive illness.** *Brain* 2007, **130**(Pt 9):2367–2374. [http://dx.doi.org/10.1093/brain/awm150]
12. Huys QJM: **Reinforcers and control. Towards a computational ætiology of depression.** *PhD thesis*. Gatsby Computational Neuroscience Unit, UCL, University of London 2007, [http://www.gatsby.ucl.ac.uk/ qhuys/pub.html]
13. Huys QJM, Vogelstein J, Dayan P, Bottou L: **Psychiatry: Insights into depression through normative decision-making models.** In *Advances in Neural Information Processing Systems 21*. Edited by Koller D, Schuurmans D, Bengio Y: MIT Press; 2009:729–736.
14. Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW: **Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia.** *Psychol Med* 2010, **40**(3):433–440 doi:10.1017/S0033291709990468.
15. Chase HW, Michael A, Bullmore ET, Sahakian BJ, Robbins TW: **Paradoxical enhancement of choice reaction time performance in patients with major depression.** *J Psychopharmacol* 2010, **24**(4):471–479. [http://dx.doi.org/10.1177/0269881109104883]
16. Beck A, Steer R, Brown G: *Manual for the Beck Depression Inventory-II.* San Antonio: Psychological Corporation; 1996.
17. Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, McCormick RA: **Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales.** *J Abnorm Psychol* 1995, **104**:3–144.
18. Treadway MT, Zald DH: **Reconsidering anhedonia in depression: lessons from translational neuroscience.** *Neurosci Biobehav Rev* 2011, **35**(3):537–555. [http://dx.doi.org/10.1016/j.neubiorev.2010.06.006]
19. Papp M, Klimek V, Willner P: **Parallel changes in dopamine D_2 receptor binding in limbic forebrain associated with chronic mild stress-induced anhedonia and its reversal by imipramine.** *Psychopharmacology* 1994, **115**(4):441–446.
20. Ichikawa J, Meltzer HY: **Effect of antidepressants on striatal and accumbens extracellular dopamine levels.** *Eur J Pharmacol* 1995, **281**(3):255–261.
21. D'Aquila PS, Collu M, Gessa GL, Serra G: **The role of dopamine in the mechanism of action of antidepressant drugs.** *Eur J Pharmacol* 2000, **405**(1-3):365–373.
22. Barr AM, Markou A, Phillips AG: **A 'crash' course on psychostimulant withdrawal as a model of depression.** *Trends Pharmacol Sci* 2002, **23**(10):475–482.
23. Montague PR, Dayan P, Sejnowski TJ: **A framework for mesencephalic dopamine systems based on predictive Hebbian learning.** *J Neurosci* 1996, **16**(5):1936–1947.
24. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ: **Temporal difference models and reward-related learning in the human brain.** *Neuron* 2003, **38**(2):329–337.
25. Bayer HM, Glimcher PW: **Midbrain dopamine neurons encode a quantitative reward prediction error signal.** *Neuron* 2005, **47**:129–141. [http://dx.doi.org/10.1016/j.neuron.2005.05.020]
26. Waelti P, Dickinson A, Schultz W: **Dopamine responses comply with basic assumptions of formal learning theory.** *Nature* 2001, **412**(6842):43–48. [http://dx.doi.org/10.1038/35083500]
27. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD: **BOLD responses reflecting dopaminergic signals in the human ventral tegmental area.** *Science* 2008, **319**(5867):1264–1267. [http://dx.doi.org/10.1126/science.1150605]
28. Kapur S, Mann JJ: **Role of the dopaminergic system in depression.** *Biol Psychiatry* 1992, **32**:1–17. [http://dx.doi.org/10.1016/0006-3223(92)90137-O]
29. Nutt DJ: **The role of dopamine and norepinephrine in depression and antidepressant treatment.** *J Clin Psychiatry* 2006, **67**(Suppl 6):3–8.
30. Nestler EJ, Carlezon WA: **The mesolimbic dopamine reward circuit in depression.** *Biol Psychiatry* 2006, **59**(12):1151–1159. [http://dx.doi.org/10.1016/j.biopsych.2005.09.018]
31. Dunlop BW, Nemeroff CB: **The role of dopamine in the pathophysiology of depression.** *Arch Gen Psychiatry* 2007, **64**(3):327–337. [http://dx.doi.org/10.1001/archpsyc.64.3.327]
32. Parker G: **Defining melancholia: the primacy of psychomotor disturbance.** *Acta Psychiatr Scand* 2007, **115 (Suppl 433)**(2):21–30.
33. Mathew SJ, Manji HK, Charney DS: **Novel drugs and therapeutic targets for severe mood disorders.** *Neuropsychopharmacology* 2008, **33**(9):2080–2092. [http://dx.doi.org/10.1038/sj.npp.1301652]
34. Schlaepfer TE, Cohen MX, Frick C, Kosel M, Brodesser D, Axmacher N, Joe AY, Kreft M, Lenartz D, Sturm V: **Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression.** *Neuropsychopharmacology* 2008, **33**(2):368–377. [http://dx.doi.org/10.1038/sj.npp.1301408]
35. Bewernick BH, Hurlemann R, Matusch A, Kayser S, Grubert C, Hadrysiewicz B, Axmacher N, Lemke M, Cooper-Mahkorn D, Cohen MX, Brockmann H, Lenartz D, Sturm V, Schlaepfer TE: **Nucleus accumbens deep brain stimulation decreases ratings of depression and anxiety in treatment-resistant depression.** *Biol Psychiatry* 2010, **67**(2):110–116. [http://dx.doi.org/10.1016/j.biopsych.2009.09.013]
36. Rescorla R, Wagner A: **A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.** In *Classical Conditioning II: Current Research and Theory*. Edited by Black AH, Prokasy WF: Appleton-Century-Crofts; 1972:64–99.
37. Widrow B, Hoff M: **Adaptive switching circuits.** In *WESCON Convention Report, Volume 4*: Institute of, Radio Engineers; 1960:96–104.
38. Sutton R, Barto A, et al: **Toward a modern theory of adaptive networks Expectation and prediction.** *Psychol Rev* 1981, **88**(2):135–170.
39. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. Cambridge: MIT Press; 1998. [http://www.cs.ualberta.ca/~sutton/book/the-book.html]
40. Berridge KC, Robinson TE: **What is the role of dopamine in reward hedonic impact, reward learning, or incentive salience?.** *Brain Res Rev* 1998, **28**(3):209–269.
41. Smith A, Li M, Becker S, Kapur S: **A model of antipsychotic action in conditioned avoidance: a computational approach.** *Neuropsychopharm* 2004, **29**(6):1040–1049.
42. Daw N: **Decision Making, Affect, and Learning: Attention and Performance XXIII.** Edited by Delgado MR, Phelps EA, Robbins TW: OUP; 2009.
43. Kass R, Raftery A: **Bayes factors.** *J Am Stat Assoc* 1995, **90**(430):773–795.
44. Gelman A, Carlin J, Stern H, Rubin D: *Bayesian Data Analysis*: Chapman and Hall/CRC Press; 2004.
45. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ: **Bayesian model selection for group studies.** *Neuroimage* 2009, **46**(4):1004–1017. [http://dx.doi.org/10.1016/j.neuroimage.2009.03.025]
46. Huys QJM, Moutoussis M, Williams J: **Are computational models of any use to psychiatry?.** *Neural Netw* 2011, **24**(6):544–551. [http://dx.doi.org/10.1016/j.neunet.2011.03.001]

47. Bogdan R, Santesso DL, Fagerness J, Perlis RH, Pizzagalli DA: **Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning.** *J Neurosci* 1324, **31**(37):6–13254. [http://dx.doi.org/10.1523/JNEUROSCI.2661-11.2011]

48. Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, Fava M: **Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task.** *J Psychiatr Res* 2008, **43**:76–87. [http://dx.doi.org/10.1016/j.jpsychires.2008.03.001]

49. Dutra S, Brooks N, Lempert K, Guardado A, Goetz E, Pizzagalli D: **Reward responsiveness in a remitted depressed sample: Effects of gender and trait negative affect.** In *23rd Annual Meeting of the Society for Research in Psychopathology*. Minneapolis; 2009.

50. Pizzagalli DA, Goetz E, Ostacher M, Iosifescu DV, Perlis RH: **Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task.** *Biol Psychiatry* 2008, **64**(2):162–168. [http://dx.doi.org/10.1016/j.biopsych.2007.12.001]

51. Young RC, Biggs JT, Ziegler VE, Meyer DA: **A rating scale for mania reliability, validity and sensitivity.** *Br J Psychiatry* 1978, **133:** 429–435.

52. Hamilton M: **A rating scale for depression.** *J Neurol Neurosurg Psychiatry* 1960, **23**:56–62.

53. Pizzagalli DA, Evins AE, Schetter EC, Frank MJ, Pajtas PE, Santesso DL, Culhane M: **Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness.** *Psychopharmacol (Berl)* 2008, **196**(2):221–232. [http://dx.doi.org/10.1007/s00213-007-0957-y]

54. Bogdan R, Pizzagalli DA: **Acute stress reduces reward responsiveness: implications for depression.** *Biol Psychiatry* 2006, **60**(10):1147–1154. [http://dx.doi.org/10.1016/j.biopsych.2006.03.037]

55. Tripp G, Alsop B: **Sensitivity to reward frequency in boys with attention deficit hyperactivity disorder.** *J Clin Child Psychol* 1999, **28**(3):366–375. [http://dx.doi.org/10.1207/S15374424jccp280309]

56. Santesso DL, Evins AE, Frank MJ, Schetter EC, Bogdan R, Pizzagalli DA: **Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function.** *Hum Brain Mapp* 2009, **30**(7):1963–1976. [http://dx.doi.org/10.1002/hbm.20642]

57. Green DM, Swets JA, et al: *Signal Detection Theory and Psychophysics, Volume 1974*. New York: Wiley; 1966.

58. Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P: **Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding.** *PLoS Comput Biol* 2011, **7**(4):e1002028. [http://dx.doi.org/10.1371/journal.pcbi.1002028]

59. MacKay DJ: *Information Theory, Inference and Learning Algorithms.* Cambridge: CUP; 2003.

60. Huys QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP: **Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees.** *PLoS Comput Biol* 2012, **8**(3):e1002410. [http://dx.doi.org/10.1371/journal.pcbi.1002410]

61. Berlin I, Givry-Steiner L, Lecrubier Y, Puech AJ: **Measures of anhedonia and hedonic responses to sucrose in depressive and schizophrenic patients in comparison with healthy subjects.** *Eur Psychiatry* 1998, **13**(6):303–309. [http://dx.doi.org/10.1016/S0924-9338(98)80048-5]

62. Dichter GS, Smoski MJ, Kampov-Polevoy AB, Gallop R, Garbutt JC: **Unipolar depression does not moderate responses to the Sweet Taste Test.** *Depress Anxiety* 2010, **27**(9):859–863. [http://dx.doi.org/10.1002/da.20690]

63. Willner P, Towell A, Sampson D, Sophokleous S, Muscat R: **Reduction of sucrose preference by chronic unpredictable mild stress, and its restoration by a tricyclic antidepressant.** *Psychopharmacology* 1987, **93**(3):358–364. [http://dx.doi.org/10.1007/BF00187257]

64. Bylsma LM, Morris BH, Rottenberg J: **A meta-analysis of emotional reactivity in major depressive disorder.** *Clin Psychol Rev* 2008, **28**(4):676–691. [http://dx.doi.org/10.1016/j.cpr.2007.10.001]

65. Smith KS, Berridge KC: **The ventral pallidum and hedonic reward neurochemical maps of sucrose "liking" and food intake.** *J Neurosci* 2005, **25**(38):8637–8649.

66. Pecina S, Berridge KC: **Hedonic hot spot in nucleus accumbens shell where do mu-opioids cause increased hedonic impact of sweetness?.** *J Neurosci* 1177, **25**(50):7–11786.

67. Berridge KC: **'Liking' and 'wanting' food rewards: brain substrates and roles in eating disorders.** *Physiol Behav* 2009, **97**(5):537–550. [http://dx.doi.org/10.1016/j.physbeh.2009.02.044]

68. Kennedy SE, Koeppe RA, Young EA, Zubieta JK: **Dysregulation of endogenous opioid emotion regulation circuitry in major depression in women.** *Arch Gen Psychiatry* 2006, **63**(11):1199–1208. [http://dx.doi.org/10.1001/archpsyc.63.11.1199]

69. Garriock HA, Tanowitz M, Kraft JB, Dang VC, Peters EJ, Jenkins GD, Reinalda MS, McGrath PJ, von Zastrow M, Slager SL, Hamilton SP: **Association of mu-opioid receptor variants and response to citalopram treatment in major depressive disorder.** *Am J Psychiatry* 2010, **167**(5):565–573. [http://dx.doi.org/10.1176/appi.ajp.2009.08081167]

70. Grant BF: **Comorbidity between DSM-IV drug use disorders and major depression: results of a national survey of adults.** *J Subst Abuse* 1995, **7**(4):481–497.

71. Swendsen J, Conway KP, Degenhardt L, Glantz M, Jin R, Merikangas KR, Sampson N, Kessler RC: **Mental disorders as risk factors for substance use, abuse and dependence: results from the 10-year follow-up of the National Comorbidity Survey.** *Addiction* 2010, **105**(6):1117–1128. [http://dx.doi.org/10.1111/j.1360-0443.2010.02902.x]

72. Samuels ER, Hou RH, Langley RW, Szabadi E, Bradshaw CM: **Comparison of pramipexole and amisulpride on alertness, autonomic and endocrine functions in healthy volunteers.** *Psychopharmacology (Berl)* 2006, **187**(4):498–510. [http://dx.doi.org/10.1007/s00213-006-0443-y]

73. Samuels ER, Hou RH, Langley RW, Szabadi E, Bradshaw CM: **Comparison of pramipexole and modafinil on arousal, autonomic, and endocrine functions in healthy volunteers.** *J Psychopharmacol* 2006, **20**(6):756–770. [http://dx.doi.org/10.1177/0269881106060770]

74. Samuels ER, Hou RH, Langley RW, Szabadi E, Bradshaw CM: **Comparison of pramipexole with and without domperidone co-administration on alertness, autonomic, and endocrine functions in healthy volunteers.** *Br J Clin Pharmacol* 2007, **64**(5):591–602. [http://dx.doi.org/10.1111/j.1365-2125.2007.02938.x]

75. Frank MJ, O'Reilly RC: **A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol.** *Behav Neurosci* 2006, **120**(3):497–517. [http://dx.doi.org/10.1037/0735-7044.120.3.497]

76. Grace AA: **Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia.** *Neuroscience* 1991, **41**:1–24.

77. Cooper J, Bloom F, Roth R: *The Biochemical Basis of Neuropharmacology.* 8th edition. USA: Oxford University Press; 2003.

78. Baudry M, Martres MP, Schwartz JC: **In vivo binding of 3H-pimozide in mouse striatum: effects of dopamine agonists and antagonists.** *Life Sci* 1977, **21**(8):1163–1170.

79. Fuller RW, Clemens JA, Hynes MD: **Degree of selectivity of pergolide as an agonist at presynaptic versus postsynaptic dopamine receptors implications for prevention or treatment of tardive dyskinesia.** *J Clin Psychopharmacol* 1982, **2**(6):371–375.

80. Tissari AH, Rossetti ZL, Meloni M, Frau MI, Gessa GL: **Autoreceptors mediate the inhibition of dopamine synthesis by bromocriptine and lisuride in rats.** *Eur J Pharmacol* 1983, **91**(4):463–468.

81. Sumners C, de Vries JB, Horn AS: **Behavioural and neurochemical studies on apomorphine-induced hypomotility in mice.** *Neuropharmacology* 1981, **20**(12A):1203–1208.

82. Schmitz Y, Benoit-Marand M, Gonon F, Sulzer D: **Presynaptic regulation of dopaminergic neurotransmission.** *J Neurochem* 2003, **87**(2):273–289.

83. Piercey MF, Hoffmann WE, Smith MW, Hyslop DK: **Inhibition of dopamine neuron firing by pramipexole, a dopamine D3 receptor-preferring agonist comparison to other dopamine receptor agonists.** *Eur J Pharmacol* 1996, **312**:35–44.

84. Chowdhury R, Guitart-Masip M, Lambert, Dayan P, Huys QJM, Düzel E, Dolan RJ: **Dopamine restores reward prediction errors in older age.** *Nat Neurosci* 2013. in press.

85. Tremblay LK, Naranjo CA, Cardenas L, Herrmann N, Busto UE: **Probing brain reward system function in major depressive disorder: altered response to dextroamphetamine.** *Arch Gen Psych* 2002, **59**(5): 209–215.

86. Berton O, Nestler EJ: **New approaches to antidepressant drug discovery: beyond monoamines.** *Nat Rev Neurosci* 2006, **7**(2):137–151. [http://dx.doi.org/10.1038/nrn1846]

87. Krishnan V, Han MH, Graham DL, Berton O, Renthal W, Russo SJ, Laplant Q, Graham A, Lutter M, Lagace DC, Ghose S, Reister R, Tannous P, Green TA, Neve RL, Chakravarty S, Kumar A, Eisch AJ, Self DW, Lee FS, Tamminga CA, Cooper DC, Gershenfeld HK, Nestler EJ: **Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions.** *Cell* 2007, **131**(2):391–404. [http://dx.doi.org/10.1016/j.cell.2007.09.018]

88. Vialou V, Robison AJ, Laplant QC, Covington HE, Dietz DM, Ohnishi YN, Mouzon E, Rush AJ, Watts EL, Wallace DL, Iñiguez SD, Ohnishi YH, Steiner MA, Warren BL, Krishnan V, Bolaÿos CA, Neve RL, Ghose S, Berton O, Tamminga CA, Nestler EJ: **DeltaFosB in brain reward circuits mediates resilience to stress and antidepressant responses.** *Nat Neurosci* 2010, **13**(6):745–752. [http://dx.doi.org/10.1038/nn.2551]

89. Pitchot W, Hansenne M, Ansseau M: **Role of dopamine in non-depressed patients with a history of suicide attempts.** *Eur Psychiatry* 2001, **16**(7):424–427.

90. Pitchot W, Reggers J, Pinto E, Hansenne M, Fuchs S, Pirard S, Ansseau M: **Reduced dopaminergic activity in depressed suicides.** *Psychoneuroendocrinology* 2001, **26**(3):331–335.

91. Tandberg E, Larsen JP, Aarsland D, Cummings JL: **The occurrence of depression in Parkinson's disease. A community-based study.** *Arch Neurol* 1996, **53**(2):175–179.

92. Gelder M, Harrison P, Cowen P: *Shorter Oxford Textbook of Psychiatry*. Oxford: Oxford University Press; 2006.

93. Floresco SB, West AR, Ash B, Moore H, Grace AA: **Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission.** *Nat Neurosci* 2003, **6**(9):968–973. [http://dx.doi.org/10.1038/nn1103]

94. Goto Y, Grace AA: **Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior.** *Nat Neurosci* 2005, **8**(6):805–812. [http://dx.doi.org/10.1038/nn1471]

95. Salamone JD, Correa M, Farrar AM, Nunes EJ, Pardo M: **Dopamine behavioral economics, and effort.** *Front Behav Neurosci* 2009, **3**(13). [http://dx.doi.org/10.3389/neuro.08.013.2009]

96. Niv Y, Daw ND, Joel D, Dayan P: **Tonic dopamine: opportunity costs and the control of response vigor.** *Psychopharmacol(Berl)* 2007, **191**(3):507–520. [http://dx.doi.org/10.1007/s00213-006-0502-4]

97. Gard D, Gard M, Kring A, John O: **Anticipatory and consummatory components of the experience of pleasure: a scale development study.** *J Res Pers* 2006, **40**(6):1086–1102.

98. Satoh T, Nakai S, Sato T, Kimura M: **Correlated coding of motivation and outcome of decision by dopamine neurons.** *J Neurosci* 2003, **23**(30):9913–9923.

99. Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJM, Dayan P, Dolan RJ, Duzel E: **Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain.** *J Neurosci* 2011, **31**(21):7867–7875. [http://dx.doi.org/10.1523/JNEUROSCI.6376-10.2011]

100. Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ: **Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine.** *J Neurosci* 1999, **19**(6):2401–2411.

101. Dickinson A, Smith J, Mirenowicz J: **Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists.** *Behav Neurosci* 2000, **114**(3):468–483.

102. Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, Rudarakanchana N, Halkerston KM, Robbins TW, Everitt BJ: **Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function.** *Behav Brain Res* 2002, **137**(1-2):149–163.

103. Di Ciano P, Cardinal RN, Cowell RA, Little SJ, Everitt BJ: **Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior.** *J Neurosci* 2001, **21**(23):9471–9477.

104. Murschall A, Hauber W: **Inactivation of the ventral tegmental area abolished the general excitatory influence of Pavlovian cues on instrumental performance.** *Learn Mem* 2006, **13**(2):123–126. [http://dx.doi.org/10.1101/lm.127106]

105. Lex B, Hauber W: **The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning.** *Neurobiol Learn Mem* 2010, **93**(2):283–290. [http://dx.doi.org/10.1016/j.nlm.2009.11.002]

106. Sasaki-Adams DM, Kelley AE: **Serotonin-Dopamine interactions in the control of conditioned reinforcement and motor behaviour.** *Neuropsychopharm* 2001, **25**(3):440–452.

107. Iiguez SD, Warren BL, Bolaos-Guzmn CA: **Short- and long-term functional consequences of fluoxetine exposure during adolescence in male rats.** *Biol Psychiatry* 2010, **67**(11):1057–1066. [http://dx.doi.org/10.1016/j.biopsych.2009.12.033]

108. Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PEM, Akil H: **A selective role for dopamine in stimulus-reward learning.** *Nature* 2011, **469**(7328):53–57. [http://dx.doi.org/10.1038/nature09588]

109. Elliott R, Sahakian BJ, Michael A, Paykel ES, Dolan RJ: **Abnormal neural response to feedback on planning and guessing tasks in patients with unipolar depression.** *Psychol Med* 1998, **28**(3):559–71.

110. Steele JD, Meyer M, Ebmeier KP: **Neural predictive error signal correlates with depressive illness severity in a game paradigm.** *Neuroimage* 2004, **23**:269–2680.

111. Knutson B, Bhanji JP, Cooney RE, Atlas LY, Gotlib IH: **Neural responses to monetary incentives in major depression.** *Biol Psychiatry* 2008, **63**(7):686–692. [http://dx.doi.org/10.1016/j.biopsych.2007.07.023]

112. Kumar P, Waiter G, Ahearn T, Milders M, Reid I, Steele JD: **Abnormal temporal difference reward-learning signals in major depression.** *Brain* 2008, **131**(Pt 8):2084–2093. [http://dx.doi.org/10.1093/brain/awn136]

113. Pizzagalli DA, Holmes AJ, Dillon DG, Goetz EL, Birk JL, Bogdan R, Dougherty DD, Iosifescu DV, Rauch SL, Fava M: **Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder.** *Am J Psychiatry* 2009, **166**(6):702–710. [http://dx.doi.org/10.1176/appi.ajp.2008.08081201]

114. Smoski MJ, Felder J, Bizzell J, Green SR, Ernst M, Lynch TR, Dichter GS: **fMRI of alterations in reward selection, anticipation, and feedback in major depressive disorder.** *J Affect Disord* 2009, **118**(1-3):69–78. [http://dx.doi.org/10.1016/j.jad.2009.01.034]

115. Dempster A, Laird N, Rubin D: **Maximum likelihood from incomplete data via the EM algorithm.** *J Royal Stat Soc Series B (Methodological)* 1977, **39**:1–38.

116. Sahani M, Linden JF: **How linear are auditory cortical responses?** In *Advances in Neural Information Processing Systems, Volume 15*. Edited by Becker S, Thrun S, Obermayer K. Cambridge: MIT Press; 2003:109–116.

117. Dickinson A, Balleine B: **The role of learning in the operation of motivational systems.** In *Stevens' Handbook of Experimental Psychology, Volume 3*. Edited by Gallistel R. New York: Wiley; 2002:497–534.

118. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.** *Nat Neurosci* 2005, **8**(12):1704–1711. [http://dx.doi.org/10.1038/nn1560]

119. Sutton R: **Integrated architectures for learning, planning, and reacting based on approximating dynamic programming.** In *Proceedings of the Seventh International Conference on Machine Learning, Volume 216*; 1990:224.

120. Dayan P, Kakade S, Montague PR: **Learning and selective attention.** *Nat Neurosci* 2000, **3 Suppl:**1218–1223. [http://dx.doi.org/10.1038/81504]

121. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS: **Learning the value of information in an uncertain world.** *Nat Neurosci* 2007, **10**(9):1214–1221. [http://dx.doi.org/10.1038/nn1954]

122. Baxter MG, Chiba AA: **Cognitive functions of the basal forebrain.** *Curr Opin Neurobiol* 1999, **9**(2):178–183.

123. Yu AJ, Dayan P: **Uncertainty, neuromodulation, and attention.** *Neuron* 2005, **46**(4):681–692. [http://dx.doi.org/10.1016/j.neuron.2005.04.026]

124. Hasselmo: **Neuromodulation: acetylcholine and memory consolidation.** *Trends Cogn Sci* 1999, **3**(9):351–359.
125. Aston-Jones G, Cohen JD: **Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance.** *J Comp Neurol* 2005, **493**:99–110. [http://dx.doi.org/10.1002/cne.20723]
126. Bylsma LM, Taylor-Clift A, Rottenberg J: **Emotional reactivity to daily events in major and minor depression.** *J Abnorm Psychol* 2011, **120**:155–167. [http://dx.doi.org/10.1037/a0021662]
127. Haeffel GJ, Abramson LY, Brazy PC, Shah JY, Teachman BA, Nosek BA: **Explicit and implicit cognition: a preliminary test of a dual-process theory of cognitive vulnerability to depression.** *Behav Res Ther* 2007, **45**(6):1155–1167. [http://dx.doi.org/10.1016/j.brat.2006.09.003]
128. Eshel N, Roiser JP: **Reward and punishment processing in depression.** *Biol Psychiatry* 2010, **68**(2):118–124. [http://dx.doi.org/10.1016/j.biopsych.2010.01.027]

1 Supplementary Methods

1.1 Fitting procedures

Our Bayesian model fitting and comparison approach has been described in detail previously [58,60], but we repeat it here very briefly for completeness.

Each model contains, for each participant, a vector of parameters $\mathbf{h}$. The maximum a posteriori estimate of each parameter $\mathbf{h}_i = \operatorname{argmax}_{\mathbf{h}} p(\mathbf{A}_i|, \mathbf{h}) p(\mathbf{h}|\boldsymbol{\theta})$ is inferred for each participant i based on that participants' actions $\mathbf{A}_i$. The parameters of the prior distribution $\boldsymbol{\theta}$ are inferred via approximate Expectation-Maximisation [59] and set to the maximum likelihood given all the data by *all* the N experimental sessions:

$$\hat{\boldsymbol{\theta}}^{ML} = \operatorname*{argmax}_{\boldsymbol{\theta}} p(\mathcal{A}|\boldsymbol{\theta})$$

Here we let $\mathcal{A} = \{\mathbf{A}_i\}_{i=1}^{N}$ be all the actions in all the experimental session. The approximation concerns the use of a Laplacian approximation for the E-step, letting:

$$p(\mathbf{h}|\mathbf{A}_i) \approx \mathcal{N}(\mathbf{h}_i^{(k)}, \Sigma_i^{(k)})$$

$$\mathbf{h}_i^{(k)} = \operatorname*{argmax}_{\mathbf{h}} p(\mathbf{A}_i|\mathbf{h}) p(\mathbf{h}|\boldsymbol{\theta}^{(k-1)})$$

$\mathcal{N}(\cdot)$ denotes a Gaussian distribution with variance $\Sigma_i^{(k)}$ and mean $\mathbf{h}_i^{(k)}$. In the M-step, the group level mean parameters $\mathbf{m}$ and (factorized) variance $\mathbf{v}^2$ are inferred simply as:

$$\mathbf{m}^{(k)} = \frac{1}{N} \sum_i \mathbf{h}_i^{(k)}$$

$$(\mathbf{v}^{(k)})^2 = \frac{1}{N} \sum_i \left[(\mathbf{h}_i^{(k)})^2 + \Sigma_i^{(k)} \right] - (\mathbf{m}^{(k)})^2$$

1.2 Model comparison

Models are compared based on the likelihood of the data given the model $\log p(\mathcal{A}|\mathcal{M})$ because we have no *a priori* preference for any particular model. As previuosly, we approximate this quantity in two steps. At the group level [43], we utilize the standard Bayesian Information Criterion (BIC) approach:

$$\log p(\mathcal{A}|\mathcal{M}) = \int d\boldsymbol{\theta} \, p(\mathcal{A}|\boldsymbol{\theta}) \, p(\boldsymbol{\theta}|\mathcal{M})$$

$$\approx -\frac{1}{2} \text{iBIC} = \log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}) - \frac{1}{2}|\mathcal{M}| \log(|\mathcal{A}|)$$

Here, $|\mathcal{A}|$ is the total number of choices summed over all experimental sessions while $|\mathcal{M}|$ is the number of prior parameters fitted (note that there are one group-level mean and one group-level variance for each parameter).

Notably, though, $\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML})$ is the sum of *integrals* over the individual parameters (hence the prescript "i" to BIC). We evaluate this by sampling from the empirical prior and averaging over the samples:

$$
\begin{aligned}
\log p(\mathcal{A}|\hat{\boldsymbol{\theta}}^{ML}) &\approx \sum_i \log \frac{1}{K} \sum_{k=1}^{K} p(\mathbf{A}_i|\mathbf{h}^k) \\
\mathbf{h}^k &\sim p(\mathbf{h}|\boldsymbol{\theta}^{ML})
\end{aligned}
$$

1.3 Belief model specification

The model 'Belief' applies the belief variable ζ only to the learned $\mathcal{Q}$ values, and not to the instructions $\mathcal{I}$. Prima facie this is surprising, as if a subject is able to correctly identify the stimulus to follow instructions, then they should be able to do so for the learned values, too. Here we briefly show that the models are, in fact, equivalent.

In model 'Belief', choices are determined by the term in the softmax:

$$
\mathcal{W}_t(a_t, s_t) - \mathcal{W}_t(\bar{a}_t, s_t) = \gamma(\mathcal{I}(a_t, s_t) - \mathcal{I}(\bar{a}_t, s_t)) + f(\mathcal{Q})
$$

where $f(\mathcal{Q})$ is a term that only involves $\mathcal{Q}$ and so is not important for the effect of ζ on $\mathcal{I}$.

An alternative model, where ζ applies to both the learned component $\mathcal{Q}$ and the instructions $\mathcal{I}$, would read:

$$
\begin{aligned}
\mathcal{W}_t(a_t, s_t) =& \gamma(\zeta\mathcal{I}(a_t, s_t) + (1-\zeta)\mathcal{I}(a_t, \bar{s}_t)) + \\
& \zeta\,\mathcal{Q}_t(a_t, s_t) + (1-\zeta)\,\mathcal{Q}_t(a_t, \bar{s}_t)
\end{aligned}
$$

and so the term in the choice-determining softmax would be

$$
\begin{aligned}
\mathcal{W}_t(a_t, s_t) - \mathcal{W}_t(\bar{a}_t, s_t) =& \gamma(\zeta\mathcal{I}(a_t, s_t) + (1-\zeta)\mathcal{I}(a_t, \bar{s}_t)) - \\
& \gamma(\zeta\mathcal{I}(\bar{a}_t, s_t) + (1-\zeta)\mathcal{I}(\bar{a}_t, \bar{s}_t)) + f(\mathcal{Q})
\end{aligned}
$$

however, $\mathcal{I}(a_t, s_t) = \mathcal{I}(\bar{a}_t, \bar{s}_t)$, and similarly for the off-diagonal terms of $\mathcal{I}$, so we have

$$
\mathcal{W}_t(a_t, s_t) - \mathcal{W}_t(\bar{a}_t, s_t) = \gamma((2\zeta - 1))\mathcal{I}(a_t, s_t) - (2\zeta - 1)\mathcal{I}(\bar{a}_t, s_t)f(\mathcal{Q})
$$

which is this equivalent to redefining γ as $\gamma(2\zeta - 1)$.

Another alternative, would be to argue that uncertainty affects not only the selection of the appropriate $\mathcal{Q}$ value, but also the learning. To test this legitimate idea, we fitted an additional model in which, on each

iteration, the $\mathcal{Q}$ value of the chosen action a_t for both the presented stimulus s_t and the other stimulus $\bar{s}_t$ is updated to an extent proportional to the belief:

$$
\begin{aligned}
\mathcal{Q}(a_t, s_t) &= \mathcal{Q}(a_t, s_t) + \zeta\epsilon(\rho r_t - \mathcal{Q}(a_t, s_t)) \\
\mathcal{Q}(a_t, \bar{s}_t) &= \mathcal{Q}(a_t, \bar{s}_t) + (1 - \zeta)\epsilon(\rho r_t - \mathcal{Q}(a_t, \bar{s}_t))
\end{aligned}
$$

This model performed very similarly to the model 'Action' and performed worse than the model 'Belief' (log Bayes factor compared to model 'Belief' 138.04). We therefore did not analyse it further.

2 Supplementary Results
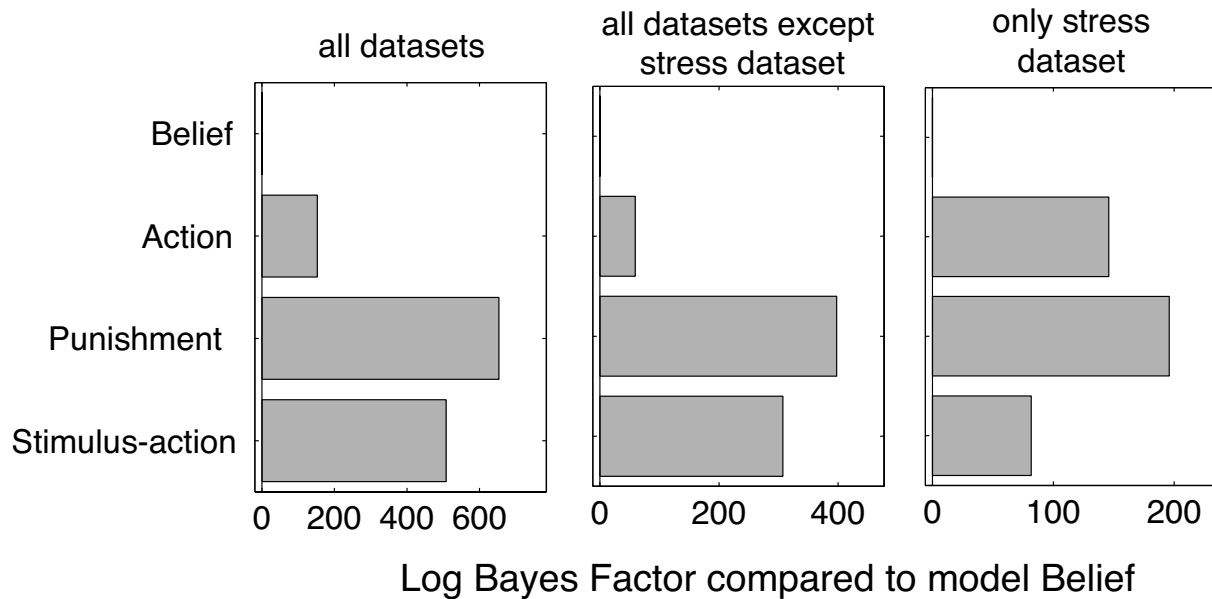### 2.1 Model comparison for Stress dataset

Figure S1: Model comparison log Bayes factors. Left panel is the same as in Figure 2A for comparison; the middle panel for all but the Stress dataset; and the right panel for the Stress dataset alone. Throughout, there is clear evidence for model 'Belief'.

Because they faced a more difficult perceptual task (a smaller difference between the 'long' and the 'short' mouth), participants in the stress dataset, showed on average a lower probability of correct choice, and were consequently less well predicted by all models (c.f. Figure 2). We therefore performed additional, separate model fits and comparisons for the Stress dataset, and the remaining five datasets; and for all datasets separately. Figure S1 shows the results for the Stress dataset and all other datasets combined. The same 'Belief' model provided the (jointly) best fit in all datasets. In addition, the belief values did not differ between the Stress dataset and the other datasets (data not shown). Combined with the finding that participants who responded correctly less often had lower γ values (Figure 2C), this shows that these participants specifically assigned less weight to the instructions, but not that they differed in their ability to identify, or learn about, the stimuli. Thus, the accuracy difference between the Stress dataset and the other datasets was well captured by the parameter γ of the model and does not require separate treatment of the Stress dataset.
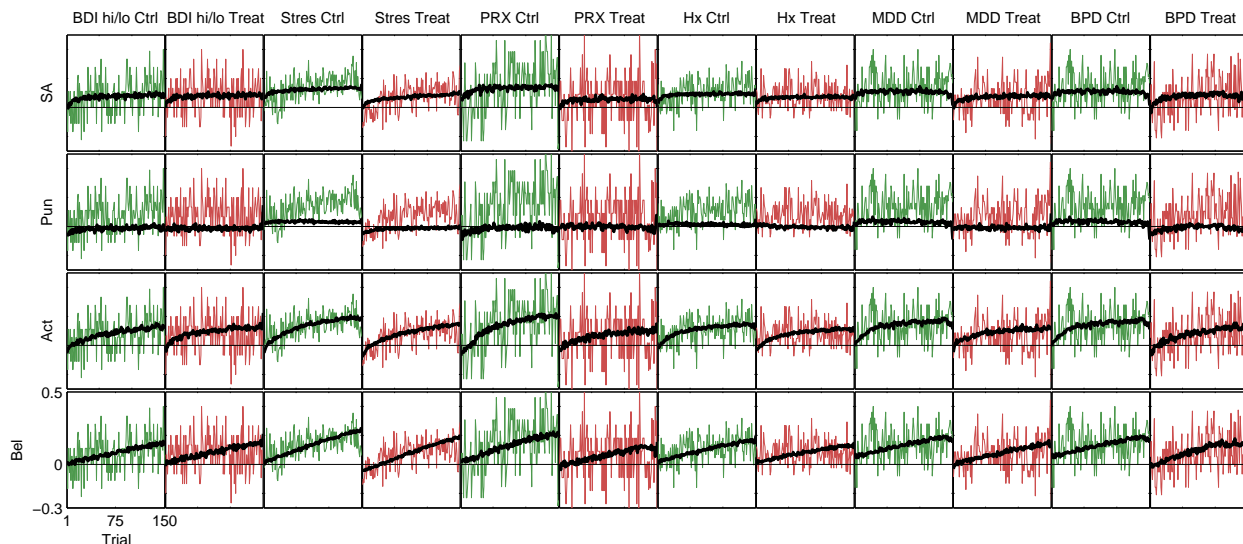
2.2 Surrogate data comparison

Figure S2: Evolution of choice bias for all datasets (left to right) and all models (top to bottom). Each dataset is split into two panels, showing the 'control' and 'treatment' subgroups in green and red, respectively. Each curve shows the difference in the log probability of a correct response to the rich minus the lean stimulus, averaged across participants. The black curves show the same measure averaged across the 70 surrogate datasets. Note that these models were fit to all six datasets jointly, yet all datasets are fitted reasonably well.

As additional test of model fit, we generated data from the model. Sampling is a good guide as to whether the model captures the data; but does not measure parsimony. We generated 70 surrogate datasets from each models for direct comparison with the experimental results on the evolution of the choice bias log p(correct response—rich stimulus) - log p(correct — lean stimulus). We obtained the *maximum a posteriori (MAP)* parameters $\mathbf{h}^k$ for each participant, and then used this together with the original sequence of stimuli for that participant to generate a novel sequence of 300 choices, although we used a truly random reward sequence rather than the pseudorandom sequence used in the experiment to generate this data. Because datasets were pooled for parameter inference, we expect the resampled data to regress towards the mean of all datasets (this effect should be stronger in smaller datasets). However, we should expect roughly to replicate the general pattern observed in the raw data.

Figure S2 shows the evolving bias towards the rich stimulus for all models and all datasets. The evolution of the bias was reasonably well captured across all datasets. It is notable that the model 'Action' seems to perform similarly to the model 'Belief' despite being simpler. We therefore repeated the regression analyses with the parameters derived from this model.