Perseverative behavior under uncertainty is enhanced by tryptophan depletion but diminished with subclinical obsessive–compulsive symptoms

Jonathan W. Kanen^{1,2}, Qiang Luo^{3,4,5}, Robyn Yellowlees⁶, Frédérique E. Arntz⁷, David M. Christmas^{8,9}, Barbara J. Sahakian^{2,8}, Annemieke M. Apergis-Schoute¹⁰, Rudolf N. Cardinal^{8,9}, Trevor W. Robbins^{1,2}

¹Department of Psychology, University of Cambridge, Cambridge, UK

²Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK

³Institute of Science and Technology for Brain-Inspired Intelligence, Ministry of Education Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Shanghai 200433, PR China

⁴State Key Laboratory of Medical Neurobiology and Ministry of Education Frontiers Center for Brain Science, Institutes of Brain Science and Human Phenome Institute, Fudan University, Shanghai 200433, PR China

⁵National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, 200040, PR China

⁶Section of Eating Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

⁷Department of Psychology, Leiden University, Leiden, The Netherlands

⁸Department of Psychiatry, University of Cambridge, Cambridge, UK

⁹Cambridgeshire and Peterborough NHS Foundation Trust, UK

¹⁰Department of Neuroscience, Psychology, and Behaviour, University of Leicester, Leicester, UK

Corresponding author: Jonathan W. Kanen, PhD

Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK.

Phone: +1 201 978 1612 (US)

Email: jonathan.kanen@gmail.com

Keywords: acute tryptophan depletion, serotonin, probabilistic reversal learning, reinforcement learning, compulsivity

Abstract

Serotonin is critically involved in flexible decision-making, yet the preponderance of evidence comes from non-human studies. Many insights about the serotonin's function have come from the technique of dietary acute tryptophan depletion (ATD). However, it is unclear whether ATD modulates choice behavior during probabilistic reversal learning (PRL), a widely used model of behavioral flexibility with significant translational and clinical value. To determine whether ATD affects latent choice tendencies, we applied computational models of reinforcement learning to PRL data from healthy human volunteers ($n = 62$; 29 females) who had received either ATD or placebo in a randomized, double-blind, placebo-controlled procedure. A secondary objective was to ascertain how model parameters related to clinically relevant self-report questionnaires. ATD did not affect the updating ("learning rates") or deployment ("reinforcement sensitivity") of value representations driving choice. However, ATD increased "stimulus stickiness", the tendency to choose a previously chosen visual stimulus again, regardless of reinforcement—a pattern seen in stimulant use disorder (SUD). Greater subclinical obsessive–compulsive symptoms were associated with lower stimulus stickiness, a pattern seen in clinically diagnosed obsessive–compulsive disorder (OCD). Low reinforcement sensitivity was instead related to intolerance of uncertainty and symptoms of depression and anxiety. Individuals with higher trait impulsivity showed lower reward and higher punishment learning rates, which parallels findings in SUD. Collectively, these results point to a role for serotonin in compulsive tendencies. They underscore the utility of computational modelling in illuminating the microstructure of behavior, which could point towards new markers of vulnerability to psychopathology.

Introduction

To understand the neurochemical mechanisms supporting flexible decision-making would have profound implications for improving the lives of individuals with a wide range of neuropsychiatric disorders. Serotonin has a well-established role in adapting behaviour flexibly to new circumstances; however, the preponderance of evidence comes from non-human animals (Bari et al. 2010; Lapiz-Bluhm et al. 2009; Matias et al. 2017; Rygula et al. 2015). Flexible decision-making is often modelled using probabilistic reversal learning (PRL) paradigms, in which subjects first learn, through trial and error, which of two stimuli more commonly leads to positive reinforcement; the contingencies then swap in a reversal phase. A core appeal of PRL is its translational value, amenable to study in rodents, monkeys, and humans both in health and disease (Robbins and Cardinal 2019).

Acute tryptophan depletion (ATD) is the technique most commonly used for studying serotonin in humans (Faulkner and Deakin, 2014) and has been in use for over 30 years (Bel and Artigas 1996; Biggio et al. 1974; Young, 2013). ATD capitalises on the principles of serotonin synthesis and transport across the blood-brain barrier (BBB), and involves dietary depletion of serotonin's biosynthetic precursor, tryptophan, in the presence of other amino acids (Hood et al. 2005). This decreases central serotonin synthesis (Crockett et al. 2012; Nishizawa et al. 1997). While many insights about the function of serotonin have been gleaned via ATD (Faulkner and Deakin 2014; Mendelsohn et al. 2009), evidence that ATD modulates choice behaviour during PRL remains notably absent.

Existing studies of ATD and PRL have focused on observable behavioural measures analysed using classical statistics. These have included “win–stay” behaviour (staying with the same choice after a win) and “lose–shift” behaviour (changing choices after negative feedback), which assay sensitivity to immediate feedback. Perseveration—inappropriately staying with the previously optimal choice following contingency reversal—has also been assessed. However, there has been a proliferation of ways in which perseveration has been measured, which hampers comparison across species, studies, and paradigms (den Ouden et al. 2013). No effects of ATD on conventional measures of choice during PRL have been reported to date (Evers et al. 2005; Finger et al. 2007; Murphy et al. 2002; Kanen et al. 2020). We recently attempted to overcome some of the shortcomings of prior studies

(Evers et al. 2005; Finger et al. 2007; Murphy et al. 2002) by nearly tripling the sample size, testing both sexes, using a between-subjects design to avoid practice effects, and calculating additional conventional measures—to no avail (Kanen et al. 2020). That we were unable to identify an effect of ATD on choice during PRL is puzzling, especially given that acute selective serotonin reuptake inhibitors (SSRIs) do have an effect on PRL in humans (Chamberlain et al. 2006; Skandali et al. 2018).

Here, we make another attempt to uncover the effects of ATD during PRL, by employing an entirely different methodology. We applied computational modelling of reinforcement learning processes, using Bayesian statistics. Computational modelling allows the examination of how learning evolves dynamically as experience accrues, using trial-by-trial raw data rather than relying on averaged data such as the probability of shifting (Daw 2011). This computational approach has shed light on the role of serotonin during PRL in marmosets (Rygula et al. 2015) and also revealed dissociable deficits during PRL in obsessive-compulsive disorder (OCD) and stimulant use disorder (SUD) (Kanen et al. 2019): a rigorous cross-species and clinical comparison is therefore possible and might represent a major advantage.

Guided by previous studies (Kanen et al. 2019; Rygula et al. 2015), the concepts we were most interested in examining were the reward learning rate, punishment learning rate, overall sensitivity of behaviour to reinforcement (via a softmax inverse temperature parameter), and stimulus stickiness. These model parameters enable the assessment of whether ATD affects the rate at which individuals update their valuation of choices following reinforcement; whether this differs for positive (reward learning rate) or negative feedback (punishment learning rate); whether ATD influences the extent to which actions are guided by reinforcement-driven valuation (e.g. versus random responding) (reinforcement sensitivity); and if ATD modulates the extent to which behaviour is driven by a basic tendency to choose recently chosen options, regardless of reinforcement (stimulus stickiness). Unlike conventional measures of perseveration, which could reflect a variety of underlying mechanisms (e.g. failure to update the internal “value” of stimuli), stimulus stickiness is conceptualized as a specific inclination towards stimulus-bound behaviour (e.g. Miller et al. 2019). Because focal serotonin depletion from the marmoset amygdala or orbitofrontal cortex (OFC), via the neurotoxin 5,7- dihydroxytryptamine (5,7-

DHT), led to alterations in stimulus stickiness as well as learning rates (Rygula et al. 2015), we predicted that ATD may affect the stimulus stickiness or learning rate parameters in humans.

The recent modelling effort in OCD and SUD (Kanen et al. 2019), which revealed novel PRL deficits beyond those previously detected via conventional means (Ersche et al. 2011), motivated another set of empirical questions tested here. Perhaps counterintuitively, individuals with OCD demonstrated diminished perseverative tendencies during PRL, reflected by lower stimulus stickiness (Kanen et al. 2019). That is, people with OCD showed increased shifting of their choice away from their previous choice, regardless of the outcome of their actions. In the present study, we tested whether lower stimulus stickiness in healthy volunteers was related to higher levels of obsessive-compulsive symptoms (Foa et al. 2002).

We also explored the specificity of the hypothesized inverse relationship between subclinical OCD symptoms and stimulus stickiness, by examining whether low stimulus stickiness might alternatively be related to self-reported intolerance of uncertainty (Carleton et al. 2007), symptoms of anxiety (Spielberger et al. 1983), or depression (Beck et al. 1996). We hypothesized that low reinforcement sensitivity (inverse temperature) might instead be related to elevated symptoms of depression in healthy volunteers, which would be reminiscent of the overall blunting of affect seen in depression (Huys et al. 2012; Mukherjee et al. 2020).

Moreover, individuals with SUD, a characteristically impulsive population (Ersche et al. 2010), have shown blunted reward learning rates yet elevated learning rates for negative feedback during PRL (Kanen et al. 2019). We therefore tested whether self-reported impulsivity (Patton et al. 1995) in healthy individuals was likewise correlated with a lower reward learning rate and an elevated punishment learning rate.

This study addresses the following questions. Does ATD modulate latent reinforcement learning mechanisms, in particular stimulus stickiness or reinforcement learning rates? Is stimulus stickiness related to obsessive-compulsive symptoms or instead intolerance of uncertainty, anxiety, or depressive symptoms in a non-clinical sample? How does reinforcement sensitivity relate to these self-report measures, in particular depressive

symptoms? Is trait impulsivity correlated with learning rate parameters? Are any such relationships influenced by ATD? We investigated these questions in a large sample of healthy volunteers ($n = 62$), in a randomised double-blind placebo-controlled between-subjects design: the primary objective was to better determine whether ATD modulates choice behaviour in PRL.

Methods

Participants

Participants (33 males, 17 placebo, 16 ATD; 29 females, 15 placebo, 14 ATD) were screened to be psychiatrically healthy using the Mini International Neuropsychiatric Interview (MINI; Sheehan et al. 1998). They were medically healthy, free of regular medication besides contraception, and had not taken psychiatric or neurological medication in the past. Participants reported having no first-degree relatives with a diagnosed mental illness at the time of screening. Further participant characteristics and exclusion criteria can be found in Kanen et al. (2020, 2021). Participants gave informed consent and were paid.

General procedure

The study was approved by the Cambridge Central Research Ethics Committee (16/EE/0101) and was carried out at the National Institute for Health Research/Wellcome Trust Clinical Research Facility at Addenbrooke's Hospital in Cambridge, England. Volunteers arrived in the morning having fasted for at least 9 hours beforehand, completed a 16-item visual analogue scale (VAS) to rate baseline mood and other feelings including alertness, gave a baseline blood sample, and ingested either a tryptophan-containing placebo mixture, or a tryptophan depletion drink, administered in a randomised and double-blinded fashion. Approximately 4.5 hours after ingesting the drink (Carpenter et al. 1998), participants completed the VAS again, gave another blood sample, and completed the PRL task. Specific quantities of amino acids used can be found in Kanen et al. (2020).

Probabilistic reversal learning task

The task is shown in Figure 1 and contained 40 trials during acquisition and 40 following reversal for a total of 80 trials (Chamberlain et al., 2006; Murphy et al., 2002; Skandali et al., 2018). During acquisition, one option yielded positive feedback on 80% of trials, the other option on 20% of trials. These contingencies reversed for the latter 40 trials. Probabilistic reinforcement provides an element of uncertainty not present in tasks that employ deterministic reinforcement.

Classical statistics

Homogeneity of variance for *t* tests was verified with Levene's test, and degrees of freedom were adjusted when this assumption was violated. Multiple comparisons correction was conducted using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). The false discovery rate was set *a priori* (McDonald, 2014) at $q = .15$ (Genovese et al., 2002; Skandali et al., 2018).

Computational modelling of behaviour

Overview

These methods are based on Kanen et al. (2019). Four reinforcement learning (RL) models were fitted to the behavioural data, which incorporated parameters that have been studied previously using a hierarchical Bayesian method (Kanen et al. 2019). The priors used for each parameter are shown in Table 1. Trials were sequenced across all 80 trials of the PRL task, and on each trial the computational model was supplied with the participant's identification number and drug condition, whether the trial resulted in positive or negative feedback, and which stimuli were presented.

Models

Model 1 incorporated three parameters and was used to test the hypothesis that ATD would affect the speed at which the action values guiding behaviour are updated following positive versus negative feedback. Separate learning rates for reward (positive feedback)

α^{rew} and punishment (negative feedback, nonreward) α^{pun} were implemented. Positive reinforcement led to an increase in the value V_i of the stimulus i that was chosen, at a speed governed by the *reward rate* α^{rew} , via $V_{i,t+1} \leftarrow V_{i,t} + \alpha^{rew}(R_t - V_{i,t})$. R_t represents the outcome on trial t (defined as 1 on trials where positive feedback occurred), and $(R_t - V_{i,t})$ the prediction error. On trials where negative feedback occurred $R_t = 0$, which led to a decrease in value of V_i at a speed governed by the *punishment rate* α^{pun} , according to $V_{i,t+1} \leftarrow V_{i,t} + \alpha^{pun}(R_t - V_{i,t})$. Stimulus value was incorporated into the final quantity controlling choice according to $Q^{reinf}_t = \tau^{reinf} V_t$. The additional parameter τ^{reinf} , termed *reinforcement sensitivity*, governs the degree to which behaviour is driven by reinforcement history. The quantities Q associated with the two available choices, for a given trial, were then input to a standard softmax choice function to compute the probability of each choice:

$$P(\text{action}_a) = \text{softmax}_\beta^a(Q_1 \dots Q_n) = \frac{e^{\beta Q_a}}{\sum_{k=1}^n e^{\beta Q_k}}$$

for $n = 2$ choice options. The probability values for each trial emerging from the softmax function (the probability of choosing stimulus 1) were fitted to the subject's actual choices (did the subject choose stimulus 1?). Softmax inverse temperature was set to $\beta = 1$, and as a result the reinforcement sensitivity parameter (τ^{reinf}) directly represented the weight given to the exponents in the softmax function.

Model 2 was as model 1 but incorporated a “stimulus stickiness” parameter τ^{stim} , which measures the tendency to repeat a response to a specific perceptual stimulus, irrespective of the action's outcome. This four-parameter model served to test whether accounting for stimulus-response learning, in addition to learning about action-outcome associations, would best characterise behaviour under ATD. The stimulus stickiness effect was modelled as $Q^{stim}_t = \tau^{stim} s_{t-1}$, where s_{t-1} was 1 for a stimulus that was chosen on the previous trial and was otherwise 0. The final quantity controlling choice incorporated this additional parameter as $Q_t = Q^{reinf}_t + Q^{stim}_t$. Quantities Q , corresponding to the two choice options on a given trial, were then fed into the softmax function as above.

Model 3 incorporated three parameters and served to test whether a single learning rate α^{reinf} , rather than separate learning rates for positive and negative feedback, optimally

characterised behaviour under ATD. Reinforcement led to a change in the value V_i of the stimulus i that was chosen, at a speed controlled by the *reinforcement rate* α^{reinf} , via $V_{i,t+1} \leftarrow V_{i,t} + \alpha^{reinf}(R_t - V_{i,t})$. R_t represents the outcome on trial t (1 for positive feedback, 0 for negative feedback). Model 3 also included the stimulus stickiness parameter. The final quantity controlling choice was determined by $Q_t = Q^{reinf}_t + Q^{stim}_t$.

Model 4 took a different approach, and had three parameters: ϕ (phi), ρ (rho), and β (beta). Derived from the experienced-weighted attraction model (EWA) of Camerer and Ho (1999), here it was implemented as in den Ouden et al. (2013) where the EWA model best described behaviour on a nearly identical task. A key difference to the other reinforcement learning models tested in this study is that here the learning rate can decline over time, governed by a decay factor ρ (rho). The EWA model weighs the value of new information against current expectations or beliefs, accumulated from previous experience.

Learning from reinforcement is modulated by an “experience weight”, $n_{c,t}$, which is a measure of how often the subject has chosen a stimulus (i.e. experienced the action), and is updated every time the stimulus is chosen (where c is choice and t is trial) according to the experience decay factor ρ (range $0 < \rho < 1$) and can increase without bounds (den Ouden et al. 2013):

$$n_{c,t} \leftarrow n_{c,t-1} \rho + 1$$

The value of a choice is updated according to the outcome, λ , and the decay factor for previous payoffs, ϕ (range $0 < \phi < 1$) (den Ouden et al. 2013)

$$v_{c,t} \leftarrow (v_{c,t-1} \phi n_{c,t-1} + \lambda_{t-1}) / n_{c,t}$$

The payoff decay factor ϕ (phi) is related to a Rescorla–Wagner-style (Rescorla and Wagner, 1972) learning rate α (as in Models 1–3), by $\alpha = 1 - \phi$. A high value of ϕ means that stimuli keep a high fraction of their previous value and thus learning from reinforcement is slow. When ρ is high, then “well-known” actions (with high n) are updated relatively little by reinforcement, by virtue of the terms involving n , whilst reinforcement has a proportionately larger effect on novel actions (with low n). For comparison to Models 1–3, when $\rho = 0$, the experience weight n , is 1, which reduces to a learning rate α controlling

the influence of learning from prediction error. Choice in the EWA model is also governed by a softmax process, only here the softmax inverse temperature β was also a parameter able to vary, in contrast to Models 1–3.

Model fitting and comparison

Models were fitted using Hamiltonian Markov chain Monte Carlo sampling via Stan 2.17.2 (Carpenter et al. 2017). Convergence was checked with the potential scale reduction factor measure $\hat{R}$ (Brooks and Gelman, 1998; Gelman et al. 2012), which approaches 1 for perfect convergence; values below 1.2 are typically used as a guideline for convergence and a cut-off of <1.1 is a stringent criterion for convergence (Brooks and Gelman 1998). The use of multiple simulation runs with measurement of convergence is an important check for simulation reliability (Wilson and Collins 2019) and is an intrinsic part of Stan. Parameter recovery from simulated data for this modelling approach has been confirmed by Kanen et al. (2019).

Models were compared using a bridge sampling estimate of the marginal likelihood (Gronau et al. 2017a) via the “bridgesampling” R package (Gronau et al. 2017b). This procedure directly estimates the marginal likelihood, and thus the posterior probability of each model, given the data, prior model probabilities, and the assumption that the models represent the entire family of those to be considered. All models were assumed to have equal prior probability.

In addition to the estimated parameters, group comparisons (e.g. between ATD and placebo for a given parameter) were also sampled directly to give a posterior probability distribution for each quantity of interest. Posterior distributions were interpreted using the 95% highest posterior density interval (HDI), the Bayesian “credible interval”. Identical priors were used for all groups and conditions.

Results

Blood results and mood

Robust depletion of tryptophan was achieved ($t_{(49)} = -17.726$, $p = 4.857 \times 10^{-23}$; blood samples unavailable for 3 participants), and mood was unaffected ($t_{(55)} = -1.341$, $p = .186$; data unavailable for 5 participants).

Choice of reinforcement learning model

The core modelling results are summarised in Figure 2. Computational modelling accounts for how behaviour is influenced by an integration of previous choices and feedback history from multiple experiences. Behaviour in the present PRL task was best characterised by a simple reinforcement learning model, as determined by a bridge sampling estimate of the marginal likelihood (Table 2). Four reinforcement learning models were fitted and compared. Convergence was nearly perfect with all four models having $\hat{R} < 1.001$. The winning model (Model 2) included four parameters: 1) reward learning rate, the speed at which an action's value is updated following positive feedback; 2) punishment learning rate, the speed at which an action's value decreases following negative feedback; 3) reinforcement sensitivity, which is the degree to which overall outcome of behaviour contributes to choice (how heavily stimulus value learned through reinforcement is weighted); and 4) "stimulus stickiness" which indexes the tendency to get "stuck" to a cue: was the chosen stimulus selected on the previous trial, irrespective of outcome? It should be noted that Model 2 (which had separate learning rates for reward and punishment) and Model 3 (which had a single learning rate) were nearly equiprobable. Comparison via parameter estimation is in general better than via model comparison (Kruschke 2011): Model 2 both wins via model comparison and allows for a direct comparison of reward and punishment learning rates. In other words, a very close model comparison has gone the way that does not lead to potential loss of information.

Effect of ATD on model parameters

Modelling results are summarised in Figure 2. The key result was that stimulus stickiness was elevated under ATD compared to placebo (with the posterior 95% highest posterior

density interval [HDI] of the difference between these means excluding zero; $0 \notin 95\%$ HDI). There were no differences between placebo and ATD for the reward learning rate, punishment learning rate, or reinforcement sensitivity parameters, which all had a 95% HDI containing zero ($0 \in 95\%$ HDI).

Relationship between obsessive–compulsive symptoms and stimulus stickiness

Results are summarised in Figure 3a. Clinically diagnosed OCD has been associated with diminished stimulus stickiness (Kanen et al. 2019). Here we sought to determine whether an analogous pattern of low stimulus stickiness applies in healthy volunteers with greater subclinical OCD symptoms. Indeed, stimulus stickiness was significantly correlated with obsessive–compulsive (OC) symptoms ($r_{(62)} = -.393$, $p = .002$), such that individuals with more OC symptoms demonstrated lower stimulus stickiness, consistent with clinical data on OCD (Kanen et al. 2019).

The relationship between OC symptoms and stickiness was not modulated by serotonergic status. Analysis of covariance (ANCOVA) was performed to test an interaction effect of serotonin (placebo versus depletion) and OC symptoms (OCI-R scores) (between-subjects factors) on stimulus stickiness, controlling for main effects. A serotonin $\times$ OC interaction was not present, indicating that the relationship between total OCI-R score and stickiness was not modulated by serotonergic status ($F_{(1,61)} = 2.239$, $p = .14$, $\eta_p^2 = .037$). To characterise the relationship between OCI-R score and stickiness at baseline, we also conducted an exploratory analysis on the placebo group alone: lower stimulus stickiness remained correlated with elevated OCI-R scores ($r_{(32)} = -.529$, $p = .002$).

Specificity of low stimulus stickiness to obsessive–compulsive symptoms

Results are summarised in Figure 3b–3d. To examine the specificity of the hypothesized relationship between low stimulus stickiness and greater OC symptoms, we conducted exploratory correlations to determine whether reduced stickiness in healthy volunteers is also related to elevations in trait anxiety, depressive symptoms, or intolerance of uncertainty. Stimulus stickiness was not correlated with intolerance of uncertainty ($r_{(62)} = -.032$, $p = .804$), trait anxiety ($r_{(62)} = -.142$, $p = .270$), or depressive symptoms ($r_{(62)} = -.054$, $p = .674$), collapsing across serotonergic status. Repeating the above ANCOVA approach

separately for intolerance of uncertainty, trait anxiety, and depressive symptoms confirmed there were no trait $\times$ serotonin interaction effects on stimulus stickiness ($F < 1.4$, $p > .05$, $\eta_p^2 < .03$ for all interaction terms).

Relationship between obsessive-compulsive symptom categories and stimulus stickiness

Because OCD and OC symptoms have numerous manifestations, which can be assessed using subscales of tools like the OCI-R, we followed up the significant correlation between stickiness and total OCI-R score by exploring categories of symptoms, collapsed across serotonergic status. Lower stimulus stickiness was significantly correlated with higher scores of checking ($r_{(62)} = -.392$, $p = .002$), ordering ($r_{(62)} = -.261$, $p = .04$), neutralising ($r_{(62)} = -.380$, $p = .002$), washing ($r_{(62)} = -.304$, $p = .016$), and obsessing ($r_{(62)} = -.352$, $p = .005$), but not hoarding ($r_{(62)} = -.129$, $p = .319$). All five significant correlations survived correction for six (subscale) comparisons via the Benjamini-Hochberg method (see Methods). These correlations, furthermore, held up when controlling for BDI, STAI, and IUS (all $p < .05$). Partial correlation analyses, moreover, showed that checking was not the dominant correlation: controlling for washing, for instance, rendered the checking correlation not significant.

These relationships were impervious to experimental changes in serotonin, as determined via ANCOVA: interaction effects, between serotonin status and each subscale, on stimulus stickiness were absent for all subscales ($F < 3$, $p > .05$, $\eta_p^2 < .05$ for all interaction terms). This converges with evidence from individuals with OCD, who show aberrantly low stickiness despite serotonergic medication (Kanen et al. 2019).

Relationship between trait measures and reinforcement sensitivity

Results are shown in Figure 4. Collapsing across serotonergic status, diminished reinforcement sensitivity was correlated with depressive symptoms ($r_{(62)} = -.295$, $p = .02$) assessed with the Beck Depression Inventory-II (BDI-II; Beck et al. 1996), as hypothesized. The relationship between reinforcement sensitivity and depressive symptoms did not differ between placebo and ATD assessed via ANCOVA ($F = .508$, $p = .479$, $\eta_p^2 = .009$). When collapsing across serotonergic status, reinforcement sensitivity,

meanwhile was not correlated with intolerance of uncertainty ($r_{(62)} = -.236, p = .064$), trait anxiety ($r_{(62)} = -.201, p = .117$), or OC symptoms ($r_{(62)} = -.104, p = .42$). The significant correlation between elevated depressive symptoms and diminished reinforcement sensitivity survived the Benjamini-Hochberg procedure for four comparisons.

We then explored the relationship between individual subject reinforcement sensitivity parameter estimates and self-report measures in the placebo group only, to characterise these relationships at baseline, when unaffected by ATD, which may have been obscured by including all participants in the same analysis. As was the case for the entire sample, diminished reinforcement sensitivity remained correlated with elevated BDI scores when isolating the placebo group ($r_{(32)} = -.421, p = .016$). On placebo, lower reinforcement sensitivity was additionally correlated with greater intolerance of uncertainty ($r_{(32)} = -.433, p = .013$) and higher trait anxiety ($r_{(32)} = -.366, p = .04$), but not with OCI-R scores ($r_{(32)} = -.246, p = .174$). These three significant correlations involving reinforcement sensitivity in the placebo group survived the Benjamini-Hochberg procedure for four comparisons.

Relationship between learning rates and trait impulsivity

Results are displayed in Figure 5. Next, we tested the relationship between learning rates and trait impulsivity, assessed using the Barratt Impulsiveness Scale (BIS; Patton et al. 1995). This analysis was prompted by Kanen et al. (2019): individuals with stimulant use disorder (SUD), who are characteristically impulsive (Ersche et al. 2010), demonstrated diminished reward and elevated punishment learning rates during PRL. Indeed, analysis of healthy volunteers under placebo revealed that individuals higher in trait impulsivity showed a lower reward learning rate ($r_{(32)} = -.363, p = .041$) and a higher punishment learning rate ($r_{(32)} = .375, p = .034$). Correlations, within the depletion, were not observed between trait impulsivity and the reward learning rate ($r_{(30)} = -.086, p = .652$) or the punishment learning rate ($r_{(30)} = .267, p = .153$). Critically, the two significant correlations in the placebo group survived the Benjamini-Hochberg procedure for four comparisons. There was no serotonin $\times$ impulsivity interaction effect on either the reward or punishment learning rate, as assessed by two ANCOVAs ($F < .7, p > .05, \eta_p^2 < .015$, for both interaction terms).

Discussion

In light of the previous difficulty in detecting effects of ATD on choice behaviour during PRL (Evers et al. 2005; Finger et al. 2007; Murphy et al. 2002; Kanen et al. 2020), we reasoned that the microstructure of behaviour as revealed through computational modelling, might be more sensitive to ATD. Indeed, the present study uncovered effects of ATD on choice tendencies during PRL for the first time, enabled by computational methods. Beyond the effect of ATD, the modelling approach also enabled new relationships to be uncovered between the parameters governing behaviour and both self-reported traits and subclinical symptoms. Specifically, this study has shown that 1) ATD increased stimulus stickiness; 2) diminished stimulus stickiness was related to elevated obsessive-compulsive symptoms and not to depressive symptoms, trait anxiety, or intolerance of uncertainty; 3) diminished stimulus stickiness was correlated with all OC subscales (checking, washing, etc.) except for hoarding; 4) diminished reinforcement sensitivity was related to elevated depressive symptoms, trait anxiety, and intolerance of uncertainty, but not OC symptoms; and 5) higher trait impulsivity was related to diminished reward and enhanced punishment learning rates.

That stimulus stickiness was elevated under ATD may be particularly relevant for understanding stimulant use disorder, where enhanced stimulus stickiness was recently found during PRL (Kanen et al. 2019). There is robust evidence that serotonin in the OFC is critical for reversal learning (Barlow et al. 2015; Clarke et al. 2004; Lapiz-Bluhm et al. 2009). Indeed, it has been shown that individuals with SUD have decreased serotonin concentration in the OFC assessed *post mortem* (Wilson et al., 1996), as well as diminished levels of the OFC serotonin transporter (Kish et al., 2009). Furthermore, rats prone to compulsive drug-taking have been shown to have diminished forebrain serotonin (assessed via the 5-HT:5-HIAA ratio) in ventral and dorsal PFC, ventral and dorsal striatum, and amygdala (Pelloux et al. 2012). Cox et al. (2011), moreover, studied non-dependent cocaine users and found that during cocaine administration, ATD enhanced drug craving, which was accompanied by elevated striatal dopamine responses.

Our result, that ATD enhanced stimulus stickiness, is strengthened when aligned with a series of studies using other behavioural paradigms showing that ATD increases compulsive tendencies in healthy volunteers (Seymour et al. 2012; Worbe et al. 2015; 2016). One study found that ATD elevated a comparable measure of stickiness on a four-

choice probabilistic task (Seymour et al. 2012). Other studies have shown a shift away from goal-directed behaviour towards habitual responding for rewards under ATD (Worbe et al. 2015; 2016), a pattern of responding seen in SUD (Ersche et al. 2016; Voon et al. 2015a) but also relevant for understanding alcohol use disorder (Sjoerds et al. 2013; Chen et al. 2021), nicotine dependence (Luijten et al. 2020), binge eating disorder (Voon et al. 2015a), Tourette syndrome (Delorme et al. 2016), and OCD (Gillan et al. 2011; Voon et al. 2015a; but see Voon et al. 2015b, Worbe et al. 2016 for how effects in OCD and ATD differ when responding to avoid punishment).

Enhanced stickiness under ATD is also consistent with results in marmosets after serotonin depletion using 5,7-DHT (Rygula et al. 2015). Following administration of 5,7-DHT to either the OFC or amygdala, marmosets showed heightened location stickiness (a tendency to repeat choices in the same location as before, regardless of outcome; Rygula et al. 2015). Additionally, marmosets that received 5,7-DHT in the OFC tended to repeat choices to recently chosen stimuli across a longer timescale (were marginally slower than controls to “update” the effects of stimulus stickiness following their choices), whereas 5,7-DHT in the amygdala produced a more ephemeral tendency to repeat choices (Rygula et al. 2015).

The modelling approach allowed for further comparisons to clinical conditions, beyond the effects of ATD. Healthy volunteers with higher OC symptom scores showed lower stimulus stickiness, which replicates previous results in clinically diagnosed OCD (Hauser et al. 2017; Kanen et al. 2019) and highlights a possible marker of vulnerability. Having established the relationship between diminished stimulus stickiness and OC symptoms in the present healthy sample, we were able to address a series of follow-up questions. Previous studies had conjectured that the tendency of patients with OCD not to repeat previous choices during PRL could be a manifestation of checking behaviour—checking the previously unchosen option (Hauser et al. 2017; Kanen et al. 2019). Here we showed that it was not only the “checking” subscale of the OCI-R that was correlated with stimulus stickiness, but that all other subscales (neutralising, obsessing, ordering, washing) save for “hoarding” were correlated. The suggestion here is that the phenomenon of low stimulus stickiness is correlated with OCD symptoms globally, except perhaps hoarding. This could be relevant for the recent reclassification of hoarding disorder as a distinct condition from OCD, albeit in the same DSM5 section: obsessive-compulsive and related disorders (APA

2013). That hoarding was not related to diminished stimulus stickiness in this study raises the possibility that this parameter may be an important novel objective measure for distinguishing other related conditions such as body dysmorphic disorder, excoriation (skin-picking) disorder, and trichotillomania (hair-pulling disorder; APA 2013).

The significant correlations involving OC symptoms were impervious to changes in serotonin, which is consistent with Kanen et al. (2019), in which individuals with OCD showed diminished stimulus stickiness despite the fact that most were taking SSRIs. Whilst we also clarified that the stickiness parameter was not related to intolerance of uncertainty, anxiety, or depressive symptoms (regardless of serotonergic status), which we had charted as candidate phenomena underlying this behavioural pattern, the question remains: What does reduced stimulus stickiness mean for understanding OCD?

One intriguing account posits that the brain in OCD has difficulty estimating the probability of transitioning between states of the world (Fradkin et al. 2020a, 2020b). In the context of PRL, one state would be the acquisition phase, another state would be the reversal phase. In OCD, this would map onto statistical uncertainty (variance) about the likelihood of transitioning from dirty to clean hands after washing, for example. The optimal response to uncertainty about the current state would be to continue gathering information (Fradkin et al. 2020a). In the context of PRL, this would manifest as overly exploratory responses, consistent with low stimulus stickiness as seen here in individuals with high subclinical OCD symptoms, and in clinical OCD (Kanen et al. 2019). Importantly, the OFC, a key area of dysfunction in OCD (Chamberlain et al. 2008; Remijnse et al 2006), is implicated in representing states (Schuck et al. 2016; Wilson et al. 2014). In extension of this framework, it should be noted that SUD, where high stimulus stickiness has been observed (Kanen et al. 2019), may be characterised by over-encoding of state-specific rules and information (Mueller et al. 2020).

Subsequent analyses primarily focused on the placebo group in order to assess baseline relationships between model parameters and self-report measures. There was a striking dissociation in how reinforcement sensitivity as opposed to stimulus stickiness related to

self-report measures: Blunted reinforcement sensitivity was related to depressive symptoms, as well as trait anxiety and intolerance of uncertainty, but not to OC symptoms.

The learning rates of impulsive healthy volunteers, furthermore, aligned with what has been reported in SUD (lower reward and higher punishment learning rates; Kanen et al. 2019), providing additional replications between the non-clinical sample studied here, and clinical populations. It also raises the possibility that alterations in learning rates may precede and potentially predispose to problematic use of stimulants like cocaine and methamphetamine, rather than being a consequence of drug toxicity. That impulsivity is a behavioural endophenotype, or vulnerability trait for stimulant abuse, found also in first-degree unaffected relatives (Ersche et al. 2010), suggests that learning rates assessed using computational modelling may represent an additional endophenotype warranting further study. It is advantageous that learning rates are objective measures that can be assessed across paradigms, populations, and species.

The present study has, for the first time, identified a latent mechanism supporting behaviour that was modulated by ATD during PRL. That computational methods unearthed effects in the face of null results on conventional measures (Evers et al. 2005; Finger et al. 2007; Kanen et al. 2020; Murphy et al. 2002) highlights the sensitivity and promise of applying reinforcement learning models. ATD enhanced a basic perseverative tendency, which, remarkably, aligned with behaviour in SUD (Kanen et al. 2019). Considering a series of self-report measures in a hypothesis-driven approach yielded several important clinically relevant insights. Behavioural patterns in clinical OCD and SUD were replicated using analogous methods in healthy volunteers high in OC symptoms and trait impulsivity, respectively. Applying reinforcement learning models to PRL may hold promise for refining psychiatric classification (Cuthbert and Insel 2013) and identifying vulnerability factors to an array of psychopathologies (Marx et al. 2020), including disorders of compulsivity.

References

- APA (2013) Diagnostic and Statistical Manual of Mental Disorders, 5th edn. American Psychiatric Association, Washington, DC
- Bari A, Theobald DE, Caprioli D, et al (2010) Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats. *Neuropsychopharmacology* 35:1290–1301. doi: 10.1038/npp.2009.233
- Barlow RL, Alsö J, Jupp B, et al (2015) Markers of Serotonergic Function in the Orbitofrontal Cortex and Dorsal Raphe Nucleus Predict Individual Variation in Spatial-Discrimination Serial Reversal Learning. *Neuropsychopharmacology* 40:1619–1630. doi: 10.1038/npp.2014.335
- Beck AT, Steer RA, Ball R, Ranieri WF (1996) Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *J Pers Assess* 67:588–597. doi: 10.1207/s15327752jpa6703_13
- Beck SG, Pan YZ, Akanwa AC, Kirby LG (2004) Median and Dorsal Raphe Neurons Are Not Electrophysiologically Identical. *J Neurophysiol* 91:994–1005. doi: 10.1152/jn.00744.2003
- Bel N, Artigas F (1996) Reduction of serotonergic function in rat brain by tryptophan depletion: Effects in control and fluvoxamine-treated rats. *J Neurochem* 67:669–676. doi: 10.1046/j.1471-4159.1996.67020669.x
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* 57:289–300
- Biggio G, Fadda F, Fanni P, et al (1974) Rapid depletion of serum tryptophan, brain tryptophan, serotonin and 5-hydroxyindoleacetic acid by a tryptophan-free diet. *Life Sci* 14:1321–1329. doi: 10.1016/0024-3205(74)90440-8
- Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7:434–455
- Carleton RN, Norton MAPJ, Asmundson GJG (2007) Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *J Anxiety Disord* 21:105–117. doi: 10.1016/j.janxdis.2006.03.014
- Carpenter B, Gelman A, Hoffman MD, et al (2017) Stan: A Probabilistic Programming Language. *J Stat Softw* 76:1–32. doi: 10.18637/jss.v076.i01
- Carpenter LL, Anderson GM, Pelton GH, et al (1998) Tryptophan depletion during continuous CSF sampling in healthy human subjects. *Neuropsychopharmacology* 19:26–35. doi: 10.1016/S0893-133X(97)00198-X

- Chamberlain SR, Menzies L, Hampshire A, et al (2008) Orbitofrontal Dysfunction in Patients with Obsessive-Compulsive Disorder and Their Unaffected Relatives. *Science* (80-) 321:421–422. doi: 10.1126/science.1154433
- Chamberlain SR, Müller U, Blackwell AD, et al (2006) Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* (80-) 311:861–3. doi: 10.1126/science.1121218
- Chen H, Mojtahedzadeh N, Belanger MJ, et al (2021) Model-based and model-free control predicts alcohol consumption developmental trajectory in young adults – a three-year prospective study. *Biol Psychiatry* 3223:1–13. doi: 10.1016/j.biopsych.2021.01.009
- Clarke HF, Dalley JW, Crofts HS, et al (2004) Cognitive Inflexibility After Prefrontal Serotonin Depletion. *Science* (80-) 304:878–880. doi: <https://doi.org/10.1126/science.1094987>
- Cox SML, Benkelfat C, Dagher A, et al (2011) Effects of lowered serotonin transmission on cocaine-induced striatal dopamine response: PET [11C]raclopride study in humans. *Br J Psychiatry* 199:391–397. doi: 10.1192/bjp.bp.110.084178
- Crockett MJ, Clark L, Roiser JP, et al (2012) Converging evidence for central 5-HT effects in acute tryptophan depletion. *Mol Psychiatry* 17:121–123. doi: 10.1038/mp.2011.106
- Cuthbert BN, Insel TR (2013) Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 11:126. doi: 10.1186/1741-7015-11-126
- Daw ND (2011) Trial-by-trial data analysis using computational models. *Decis Making, Affect Learn Atten Perform XXIII* 1–26. doi: 10.1093/acprof:oso/9780199600434.003.0001
- den Ouden H, Daw ND, Fernandez G, et al (2013) Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron* 80:1090–1100. doi: 10.1016/j.neuron.2013.08.030
- Ersche KD, Gillan CM, Jones PS, et al (2016) Carrots and sticks fail to change behavior in cocaine addiction. *Science* (80-) 352:1468–1471. doi: <https://doi.org/10.1126/science.aaf3700>
- Ersche KD, Turton AJ, Pradhan S, et al (2010) Drug addiction endophenotypes: Impulsive versus sensation-seeking personality traits. *Biol Psychiatry* 68:770–773. doi: 10.1016/j.biopsych.2010.06.015
- Ersche KD, Roiser JP, Abbott S, et al (2011) Response perseveration in stimulant dependence is associated with striatal dysfunction and can be ameliorated by a D2/3 receptor agonist. *Biol Psychiatry* 70:754–762. doi: 10.1016/j.biopsych.2011.06.033

- Evers EA, Cools R, Clark L, et al (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology* 30:1138–1147. doi: 10.1038/sj.npp.1300663
- Faulkner P, Deakin JFW (2014) The role of serotonin in reward, punishment and behavioural inhibition in humans: Insights from studies with acute tryptophan depletion. *Neurosci Biobehav Rev* 46P3:365–378. doi: 10.1016/j.neubiorev.2014.07.024
- Finger EC, Marsh AA, Buzas B, et al (2007) The impact of tryptophan depletion and 5-HTTLPR genotype on passive avoidance and response reversal instrumental learning tasks. *Neuropsychopharmacology* 32:206–215. doi: 10.1038/sj.npp.1301182
- Fischer AG, Jocham G, Ullsperger M (2015) Dual serotonergic signals: A key to understanding paradoxical effects? *Trends Cogn Sci* 19:21–26. doi: 10.1016/j.tics.2014.11.004
- Foa EB, Huppert JD, Leiberg S, et al (2002) The Obsessive-Compulsive inventory: Development and Validation of a Short Version. *Psychol Assess* 14:485–496. doi: 10.1037/1040-3590.14.4.485
- Fradkin I, Adams RA, Parr T, et al (2020) Searching for an Anchor in an Unpredictable World: A Computational Model of Obsessive Compulsive Disorder. *Psychol Rev*. doi: 10.1037/rev0000188
- Fradkin I, Ludwig C, Eldar E, Huppert JD (2020) Doubting what you already know: Uncertainty regarding state transitions is associated with obsessive compulsive symptoms. *PLoS Comput Biol* 16:1–26. doi: 10.1371/journal.pcbi.1007634
- Gelman A, Hill J, Yajima M (2012) Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Eff* 5:189–211. doi: <https://doi.org/10.1080/19345747.2011.618213>
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878. doi: 10.1006/nimg.2001.1037
- Gillan CM, Papmeyer M, Morein-Zamir S, et al (2011) Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry* 168:718–726. doi: 10.1176/appi.ajp.2011.10071062
- Gronau QF, Sarafoglou A, Matzke D, et al (2017) A tutorial on bridge sampling. *J Math Psychol* 81:80–97. doi: 10.1016/j.jmp.2017.09.005
- Gronau QF, Singmann H, Wagenmakers E-J (2017) bridgesampling: An R Package for Estimating Normalizing Constants. *ArXiv* 1710.08162:

- Hauser TU, Iannaccone R, Dolan RJ, et al (2017) Increased fronto-striatal reward prediction errors moderate decision making in obsessive-compulsive disorder. *Psychol Med* 47:1246–1258. doi: 10.1017/S0033291716003305
- Hood SD, Bell CJ, Nutt DJ (2005) Acute tryptophan depletion. Part I: Rationale and methodology. *Aust N Z J Psychiatry* 39:558–564. doi: 10.1111/j.1440-1614.2005.01627.x
- Huys QJM, Eshel N, O’Nions E, et al (2012) Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol* 8:. doi: 10.1371/journal.pcbi.1002410
- Kanen JW, Arntz FE, Yellowlees R, et al (2021) Serotonin depletion amplifies distinct human social emotions as a function of individual differences in personality. *Transl Psychiatry* 11:. doi: 10.1101/789669
- Kanen JW, Arntz FE, Yellowlees R, et al (2020) Probabilistic reversal learning under acute tryptophan depletion in healthy humans: a conventional analysis. *J Psychopharmacol* 34:580–583. doi: 10.1177/0269881120907991
- Kanen JW, Ersche KD, Fineberg NA, et al (2019) Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. *Psychopharmacology (Berl)* 236:2337–2358. doi: 10.1007/s00213-019-05325-w
- Kish SJ, Fitzmaurice PS, Boileau I, et al (2009) Brain serotonin transporter in human methamphetamine users. *Psychopharmacology (Berl)* 202:649–661. doi: 10.1007/s00213-008-1346-x
- Kruschke JK (2011) Bayesian assessment of null values via parameter estimation and model comparison. *Perspect Psychol Sci* 6:299–312. doi: 10.1177/1745691611406925
- Lapiz-Bluhm MDS, Soto-Piña AE, Hensler JG, Morilak DA (2009) Chronic intermittent cold stress and serotonin depletion induce deficits of reversal learning in an attentional set-shifting test in rats. *Psychopharmacology (Berl)* 202:329–341. doi: 10.1007/s00213-008-1224-6
- Luijten M, Gillan CM, de Wit S, et al (2020) Goal-Directed and Habitual Control in Smokers. *Nicotine Tob Res* 22:188–195. doi: 10.1093/ntr/ntz001
- Marx W, McGuinness AJ, Rocks T, et al (2020) The kynurenine pathway in major depressive disorder, bipolar disorder, and schizophrenia: a meta-analysis of 101 studies. *Mol Psychiatry*. doi: 10.1038/s41380-020-00951-9

- Matias S, Lottem E, Dugué GP, Mainen ZF (2017) Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife* 6:1–24. doi: 10.7554/eLife.20552
- McDonald JH (2014) *Handbook of Biological Statistics*, 3rd edn. Sparky House Publishing, Baltimore, Maryland
- Mendelsohn D, Riedel WJ, Sambeth A (2009) Effects of acute tryptophan depletion on memory, attention and executive functions: A systematic review. *Neurosci Biobehav Rev* 33:926–952. doi: 10.1016/j.neubiorev.2009.03.006
- Miller KJ, Shenhav A, Ludvig EA (2019) Habits without values. *Psychol Rev* 126:292–311. doi: 10.1037/rev0000120
- Mueller LE, Sharpe MJ, Stalnaker TA, et al (2020) Prior cocaine use alters the normal evolution of information coding in striatal ensembles during value-guided decision making. *J Neurosci* JN-RM-1755-20. doi: 10.1523/jneurosci.1755-20.2020
- Mukherjee D, Filipowicz ALS, Vo K, et al (2020) Reward and Punishment Reversal-Learning in Major Depressive Disorder. *J Abnorm Psychol* 129:810–823. doi: 10.1037/abn0000641
- Murphy FC, Smith KA, Cowen PJ, et al (2002) The effects of tryptophan depletion on cognitive and affective processing in healthy volunteers. *Psychopharmacology (Berl)* 163:42–53. doi: 10.1007/s00213-002-1128-9
- Nishizawa S, Benkelfat C, Young SN, et al (1997) Differences between males and females in rates of serotonin synthesis in human brain. *Proc Natl Acad Sci* 94:5308–5313. doi: 10.1073/pnas.94.10.5308
- Patton JH, Stanford MS, Barratt E. (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 51:768–774
- Pelloux Y, Dilleen R, Economidou D, et al (2012) Reduced Forebrain Serotonin Transmission is Causally Involved in the Development of Compulsive Cocaine Seeking in Rats. *Neuropsychopharmacology* 37:2505–2514. doi: 10.1038/npp.2012.111
- Remijnse PL, Nielen MMA, van Balkom AJLM, et al (2006) Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Arch Gen Psychiatry* 63:1225–1236. doi: 10.1001/archpsyc.63.11.1225
- Richardson-Jones JW, Craige CP, Guiard BP, et al (2010) 5-HT_{1A} Autoreceptor Levels Determine Vulnerability to Stress and Response to Antidepressants. *Neuron* 65:40–52. doi: 10.1016/j.neuron.2009.12.003

- Robbins TW, Cardinal RN (2019) Computational psychopharmacology: a translational and pragmatic approach. *Psychopharmacology (Berl)* 236:2295–2305. doi: 10.1007/s00213-019-05302-3
- Rygula R, Clarke HF, Cardinal RN, et al (2015) Role of Central Serotonin in Anticipation of Rewarding and Punishing Outcomes: Effects of Selective Amygdala or Orbitofrontal 5-HT Depletion. *Cereb Cortex* 25:3064–3076. doi: 10.1093/cercor/bhu102
- Schuck NW, Cai MB, Wilson RC, Niv Y (2016) Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* 91:1402–1412. doi: 10.1016/j.neuron.2016.08.019
- Seymour B, Daw ND, Roiser JP, et al (2012) Serotonin Selectively Modulates Reward Value in Human Decision-Making. *J Neurosci* 32:5833–5842. doi: 10.1523/JNEUROSCI.0053-12.2012
- Sheehan D V., Lecrubier Y, Sheehan KH, et al (1998) The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 59:22–23
- Sjoerds Z, de Wit S, van den Brink W, et al (2013) Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl Psychiatry* 3:e337. doi: 10.1038/tp.2013.107
- Skandali N, Rowe JB, Voon V, et al (2018) Dissociable effects of acute SSRI (escitalopram) on executive, learning and emotional functions in healthy humans. *Neuropsychopharmacology* 43:2645–2651. doi: 10.1038/s41386-018-0229-z
- Soiza-Reilly M, Gaspar P (2020) From B1 to B9: a guide through the hindbrain serotonin neurons with additional views from multidimensional characterization. In: Muller C, Cunningham KA (eds) *Handbook of the Behavioral Neurobiology of Serotonin*, 2nd edn. Academic Press, pp 23–40
- Spielberger C, Gorsuch R, Lushene R (1983) STAI manual for the State-Trait Anxiety Inventory. Consult Psychol Press 1–78
- Trulson ME (1985) Dietary tryptophan does not alter the function of brain serotonin neurons. *Life Sci* 37:1067–1072
- Voon V, Baek K, Enander J, et al (2015) Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Transl Psychiatry* 5:e670. doi: 10.1038/tp.2015.165
- Voon V, Derbyshire K, Rück C, et al (2015) Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry* 1–8. doi: 10.1038/mp.2014.44

- Wilson J, Kalasinsky K, Levey A, et al (1996) Striatal dopamine nerve terminal markers in human, chronic methamphetamine users. *Nat Med* 2:699–703
- Wilson RC, Collins AGE (2019) Ten simple rules for the computational modeling of behavioral data. *PsyArXiv* January 7,:1–35
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81:267–279. doi: 10.1016/j.neuron.2013.11.005
- Worbe Y, Palminteri S, Savulich G, et al (2016) Valence-dependent influence of serotonin depletion on model-based choice strategy. *Mol Psychiatry* 21:624–629. doi: 10.1038/mp.2015.46
- Worbe Y, Savulich G, De Wit S, et al (2015) Tryptophan depletion promotes habitual over goal-directed control of appetitive responding in humans. *Int J Neuropsychopharmacol* 18:1–9. doi: 10.1093/ijnp/pyv013
- Worbe Y, Savulich G, Voon V, et al (2014) Serotonin depletion induces “waiting impulsivity” on the human four-choice serial reaction time task: cross-species translational significance. *Neuropsychopharmacology* 39:1519–26. doi: 10.1038/npp.2013.351
- Young SN (2013) Acute tryptophan depletion in humans: a review of theoretical, practical and ethical aspects. *J Psychiatry Neurosci* 38:294–305. doi: 10.1503/jpn.120209

Figures and Tables

Figure 1. Task schematic. Two visual stimuli were presented simultaneously in two of four randomised locations on each trial. Participants needed to choose one stimulus by pressing on the touchscreen. Feedback not shown. The task was self-paced.

Figure 2. Effects of ATD relative to placebo. Stimulus stickiness was enhanced following ATD. Red signifies a difference between the parameter per-condition mean according to the Bayesian “credible interval”, $0 \notin 95\% \text{ HDI}$.

Figure 3. (a) Elevated obsessive–compulsive symptoms were correlated with lower stimulus stickiness; (b) intolerance of uncertainty, (c) trait anxiety, and (d) depressive symptoms were not correlated with stimulus stickiness. These data are collapsed across both the placebo and depletion groups. Grey background indicates a significant correlation; white background denotes lack of significance. The relationship between obsessive–compulsive symptoms in healthy volunteers and stimulus stickiness is consistent with the diminished stimulus stickiness observed by Kanen et al. (2019) in clinical OCD.

Figure 4. Under placebo, diminished reinforcement sensitivity was not related to **(a)** obsessive-compulsive symptoms, but was significantly correlated with **(b)** greater intolerance of uncertainty, **(c)** higher trait anxiety, and **(d)** elevated depressive symptoms.

Figure 5. Under placebo, higher trait impulsivity was related to **(a)** diminished reward learning rate and **(b)** elevated punishment learning rate. This pattern is consistent with learning rates during PRL in stimulant use disorder (SUD), shown by Kanen et al. (2019).

Table 1. Prior distributions for model parameters

	Models using each parameter	Prior	Reference, if applicable
Model parameters			
reward learning rate, α^{rew}	1, 2	Beta(1.2, 1.2)	den Ouden et al. (2013)
punishment learning rate, α^{pun}	1, 2	Beta(1.2, 1.2)	den Ouden et al. (2013)
combined reward/punishment learning rate, α^{reinf}	3	Beta(1.2, 1.2)	den Ouden et al. (2013)
reinforcement sensitivity, τ^{reinf}	1, 2, 3	Gamma($\alpha=4.82$, $\beta=0.88$)	Gershman (2016)
stimulus stickness, τ^{stim}	2, 3	Normal(0, 1)	Christakou et al. (2013)
experience decay factor, ρ	4	Beta(1.2, 1.2)	den Ouden et al. (2013)
decay factor for previous payoffs, ϕ	4	Beta(1.2, 1.2)	den Ouden et al. (2013)
softmax inverse temperature, β	4 [note that $\beta = 1$ in all other models]	Gamma($\alpha=4.82$, $\beta=0.88$)	Gershman (2016)
Intersubject variability in parameters			
Intersubject standard deviations for α^{rew} , α^{pun} , α^{reinf} , ρ , ϕ	As above	Half-normal; Normal(0, 0.05) constrained to ≥ 0	Kanen et al. (2019)
Intersubject standard deviations for τ^{reinf} , β	As above	Half-normal; Normal(0, 1) constrained to ≥ 0	Kanen et al. (2019)

rew reward, *pun* punishment, *reinf* reinforcement, *stim* stimulus

Table 2. Model comparison

Rank	Name	Parameters	log marginal likelihood	log posterior P(model)
4	Model 1	α^{rew} , α^{pun} , τ^{reinf}	-3438.016	-2181.1517
1	Model 2	α^{rew} , α^{pun} , τ^{reinf} , τ^{stim}	-1257.556	-0.6911
2	Model 3	α^{reinf} , τ^{reinf} , τ^{stim}	-1257.560	-0.6952
3	Model 4	ρ , ϕ , β	-3437.999	-2181.1341

rew reward, *pun* punishment, *reinf* reinforcement, *stim* stimulus

Supplementary Material

Relationship between model parameters and conventional measures

In order to understand how the model parameters relate to the conventional measures of PRL previously reported (Kanen et al. 2020), exploratory correlations were conducted and are summarised in Supplementary Table 1. Correlations were tested between individual subject parameter values from the winning model and the following six conventional measures of behaviour: the number of correct responses in acquisition, correct responses in reversal, win-stay (as a proportion of veracious positive feedback received during acquisition and reversal combined), lose-shift (as a proportion of spurious negative feedback received during acquisition and reversal combined), “immediate perseveration”, and “phase perseveration”. Phase perseveration occurred when two or more (now incorrect) responses were made to the previously correct stimulus, and these errors could occur at any point in the reversal phase (e.g. “right WRONG WRONG WRONG WRONG right WRONG WRONG” would be six phase perseverative errors) (den Ouden et al. 2013). Immediate perseveration, by contrast, refers to perseverative errors contiguous with the reversal (Murphy et al. 2002); however, the first trial in the reversal phase (trial 41 of 80) was excluded from the perseveration analysis because at that point behaviour cannot yet be shaped by the new feedback structure (e.g. an incorrect response on trials 42, 43, and 44 would be three immediate perseverative errors). These tests were performed in the placebo and depletion groups separately. The results were subsequently subjected to the Benjamini–Hochberg procedure. Stimulus stickiness was the only parameter correlated with several conventional measures of behaviour after accounting for 48 comparisons. Significant correlations were followed up by testing whether the relationships were significantly different between the placebo and depletion groups.

Stimulus stickiness was highly correlated with correct responses during both the acquisition phase when on placebo ($r_{(32)} = .853$, $p = 5.78 \times 10^{-10}$) and following ATD ($r_{(30)} = .745$, $p = 2.0 \times 10^{-6}$). The relationship between stimulus stickiness and acquisition performance was significantly stronger under placebo compared to depletion: ANCOVA with serotonin status and stimulus stickiness as a between-subjects interaction term, controlling for main effects, revealed serotonin and stimulus stickiness interacted to

modulate acquisition behaviour ($F_{(1,58)} = 7.225$, $p = .009$, $\eta_p^2 = .111$). Stimulus stickiness was also highly correlated with correct responses in the reversal phase following both placebo ($r_{(32)} = .802$, $p = 3.45 \times 10^{-8}$) and ATD ($r_{(30)} = .761$, $p = 1 \times 10^{-6}$); however, this relationship was not modulated by serotonin status as determined using the same ANCOVA method ($F_{(1,58)} = .028$, $p = .868$, $\eta_p^2 = 4.8 \times 10^{-4}$). Stimulus stickiness was highly correlated with win-stay behaviour both on placebo ($r_{(32)} = .91$, $p = 2.44 \times 10^{-18}$) and under ATD ($r_{(30)} = .946$, $p = 3.42 \times 10^{-15}$), and the correlation was significantly stronger following depletion ($F_{(1,58)} = 9.944$, $p = .003$, $\eta_p^2 = .146$). Stimulus stickiness was negatively correlated with lose-shift behaviour both on placebo ($r_{(32)} = -.940$, $p = 1.50 \times 10^{-15}$) and following ATD ($r_{(30)} = -.943$, $p = 6.43 \times 10^{-15}$). The relationship between stimulus stickiness and lose-shift behaviour was not significantly different between placebo and ATD: ANCOVA showed no stimulus stickiness $\times$ serotonin interaction: $F_{(1,58)} = .014$, $p = .905$, $\eta_p^2 = 2.5 \times 10^{-4}$. Stimulus stickiness was negatively correlated with phase perseveration under both placebo ($r_{(32)} = -.554$, $p = .001$) and ATD ($r_{(30)} = -.508$, $p = .004$), and these relationships were not significantly different from one another ($F_{(1,58)} = .085$, $p = .771$, $\eta_p^2 = .001$).

Relationship between model parameters and extent of depletion

Individual subject values of each of the four parameters from the winning model were tested for correlation with the extent of depletion achieved. Using data from all subjects, none of the four parameters were correlated with the change in the tryptophan to large neutral amino acid (TRP:LNA) ratio: reward rate ($r_{(59)} = -.241$, $p = .065$), punishment rate ($r_{(59)} = .151$, $p = .253$), reinforcement sensitivity ($r_{(59)} = -.023$, $p = .861$), and stimulus stickiness ($r_{(59)} = -.057$, $p = .666$). When examining the ATD group alone, there also was no correlation between the extent of depletion and stimulus stickiness ($r_{(28)} = -.040$, $p = .839$).

Supplementary Table 1. Correlations between stimulus stickiness and conventional measures

		Stimulus Stickiness, τ^{stim}
Correct responses in acquisition	Placebo	↑↑
	ATD	↑
Correct responses during reversal	Placebo	↑
	ATD	↑
Perseveration (phase)	Placebo	↓
	ATD	↓
Lose-Shift	Placebo	↓
	ATD	↓
Win-Stay	Placebo	↑
	ATD	↑↑

Correlations between conventional behavioural measures and the stimulus stickiness parameter. ↑ denotes a significant correlation. ↑↑ denotes the correlation is significantly stronger compared to the other condition. All correlations displayed survived the Benjamini- Hochberg procedure for multiple comparisons (see Results). The conventional measures shown here are correct responses in the acquisition phase, reversal phase; degree of phase (not immediate) perseveration (as used in den Ouden et al. 2013); proportion of lose-shift; proportion of win-stay. *stim* stimulus.

Funding and Disclosure

T.W.R. discloses consultancy with Cambridge Cognition, Arcadia, Greenfield Bioventures, Heptares, Shionogi and Takeda. Royalties with Cambridge Cognition. Research grants with Shionogi and GlaxoSmithKline and editorial honoraria with Elsevier and Springer-Nature. R.N.C. consults for Campden Instruments and receives royalties from Cambridge Enterprise, Routledge, and Cambridge University Press. B.J.S discloses consultancy with Cambridge Cognition and Greenfield BioVentures, and receives royalties for CANTAB from Cambridge Cognition. All other authors declare no conflicts of interest. J.W.K. was supported by a Gates Cambridge Scholarship and an Angharad Dodds John Bursary in Mental Health and Neuropsychiatry, and T.W.R. by a Wellcome Trust Senior Investigator Grant 104631/Z/14/Z. R.N.C.'s research is funded by the UK Medical Research Council (MC_PC_17213). Q.L. was supported by the National Key Research and Development Program of China (grant 2018YFC0910503), the National Natural Science Foundation of China (grant 81873909), the Natural Science Foundation of Shanghai (grant 20ZR1404900), the Shanghai Municipal Science and Technology Major Project (grant 2018SHZDZX01), and the Zhangjiang Lab. B.J.S. receives funding from the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (Mental Health Theme); the views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.