Effect of lysergic acid diethylamide (LSD) on reinforcement learning in humans

Jonathan W. Kanen[1,2], Qiang Luo[2,3,4,5], Mojtaba R. Kandroodi[6,7], Rudolf N. Cardinal[2,8,9], Trevor W. Robbins[1,2], Robin L. Carhart-Harris[10], Hanneke E.M. den Ouden[7]

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

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

[3]Institute 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

[4]State 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

[5]Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Zhangjiang Lab, Shanghai 201210, China

[6] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

[7] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

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

[9]Cambridgeshire and Peterborough NHS Foundation Trust, UK

[10]Centre for Psychedelic Research, Department of Brain Sciences, Imperial College London, United Kingdom

Abstract

The non-selective serotonin 2A (5-HT_{2A}) receptor agonist lysergic acid diethylamide (LSD) holds promise as a treatment for some psychiatric disorders. Psychedelic drugs such as LSD have been suggested to have therapeutic actions through their effects on learning. The behavioural effects of LSD in humans, however, remain largely unexplored. Here we examined how LSD affects probabilistic reversal learning in healthy humans. Conventional measures assessing sensitivity to immediate feedback ("win-stay" and "lose-shift" probabilities) were unaffected, whereas LSD increased the impact of the strength of initial learning on perseveration. Computational modelling revealed that the most pronounced effect of LSD was enhancement of the reward learning rate. The punishment learning rate was also elevated. Increased reinforcement learning rates suggest LSD induced a state of heightened plasticity. These results indicate a potential mechanism through which revision of maladaptive associations could occur.

Introduction

Research into lysergic acid diethylamide (LSD) as a potential therapeutic agent in psychiatry has been revitalised in recent years (Nutt and Carhart-Harris 2020; Vollenweider and Preller 2020). Theories on the putative beneficial effects of LSD on mental health centre on its effects on learning and plasticity (Carhart-Harris and Nutt 2017), yet few studies have examined its effect on human behaviour. LSD acts principally but not exclusively as an agonist at the serotonin (5-HT; 5-hydroxytryptamine) 2A [5-HT_{2A}] receptor (Marona-Lewicka et al. 2005, 2007; Nichols 2016). Indeed, blocking 5-HT_{2A} receptors inhibits the psychedelic effects of LSD (Nichols 2016). The 5-HT_{2A} receptor is involved in plasticity (Barre et al. 2016; Vaidya et al. 1997) and its modulation represents a putative neurobiological mechanism through which LSD could facilitate the revision of maladaptive associations (Carhart-Harris and Nutt 2017). Indeed, LSD and 5-HT_{2A} agonists have been shown to improve associative learning in non-human animals (Harvey 2003; Harvey et al. 1988; Romano et al. 2010; Schindler et al. 1986). However, studies of human learning and cognitive flexibility under the influence of psychedelic drugs using objective tests (rather than subjective experience) are limited in number (Pokorny et al. 2019). Here we tested whether LSD altered probabilistic reversal learning in healthy volunteers, and explored how LSD altered underlying learning mechanisms, using reinforcement learning models.

Serotonin is critically involved in adapting behaviour flexibly as environmental circumstances change (Barlow et al. 2015; Brigman et al. 2010; Clarke et al. 2004; Matias et al. 2017; Rygula et al. 2015), as well as processing aversive outcomes (Bari et al. 2010; Chamberlain et al. 2006; Cools et al. 2008; Crockett et al. 2009; Dayan and Huys 2009; Deakin 2013; den Ouden et al. 2013; Geurts et al. 2013). Both can be modelled in a laboratory setting using probabilistic reversal learning (PRL) paradigms. In these, individuals learn by trial and error the most adaptive action, in an “acquisition” stage, and this rule eventually changes in a “reversal” phase (Lawrence et al. 1999). Profound neurotoxin-induced depletion of serotonin from the marmoset orbitofrontal cortex (OFC) causes perseverative, stimulus-bound behaviour (Walker et al. 2009) – an

impaired ability to update action upon reversal (Clarke et al. 2004). At the same time, acute administration of selective serotonin reuptake inhibitors (SSRIs), which can paradoxically lower serotonin concentration (Nord et al. 2013), has resulted in an increased sensitivity to negative feedback (referred to as "lose-shift" behaviour) in healthy humans (Chamberlain et al. 2006; Skandali et al. 2018) and rats (Bari et al. 2010).

In addition to affecting the serotonin system, LSD has dopamine type 2 (D_2) receptor agonist properties (Marona-Lewicka et al. 2005, 2007; Nichols 2004). Dopamine is particularly well known to play a fundamental role in learning from feedback (Schultz 2019; Schultz et al. 1997) putatively mediating plasticity changes during associative learning (Shen et al. 2008; Yin and Knowlton 2006). Meanwhile, dopamine depletion of the marmoset caudate nucleus, like serotonergic OFC depletion, also induced perseveration (Clarke et al. 2011). Additionally, there is a body of evidence, across species, that D_2-modulating agents affect instrumental reversal learning (Boulougouris et al. 2009; Kanen et al. 2019; Lee et al. 2007).

The aim of the current study was to examine the effects of LSD on learning in humans, to inform the psychological mechanisms by which LSD could have salubrious effects on mental health. To do so, we tested the acute effects of LSD on PRL, in a placebo-controlled study of healthy human volunteers. We predicted LSD modulates either sensitivity to negative feedback or the impact of learned values on subsequent perseverative behaviour (den Ouden et al. 2013). Measuring "staying" (repeating a choice) or "shifting" (choosing another stimulus) after wins or losses assesses sensitivity to immediate reinforcement, but does not account for the integration of feedback history across multiple experiences to influence behaviour (Daw 2011). We applied computational models of reinforcement learning to test the hypothesis that LSD alters the rate at which value is updated following reward or punishment. Through modelling we additionally investigated whether LSD affects the degree to which behaviour is stimulus-driven ("stimulus sticky"), independent of an action's outcome.

Results

Learning and perseveration

First, we verified that LSD did not impair participants' overall ability to perform the task. Behavioural performance is depicted in Figure 1 and 2. To examine whether LSD affected the number of times each stimulus was chosen, repeated-measures analysis of variance (ANOVA) was conducted with drug (LSD, placebo), phase (acquisition, reversal), and stimulus type (75%, 50%, or 25% rewarded) as within-subjects factors. This revealed a main effect of stimulus ($F_{1,23}$ = 30.66, p = 3 x 10^{-6}, η_p^2 = .63), a stimulus × phase interaction (F = 28.62, p = 2 x 10^{-6}, η_p^2 = .61), and no interaction of LSD with stimulus or phase (F < 1.5, p > .24, η_p^2 < .08, for terms involving LSD). The number of correct responses did not differ between placebo and LSD during the acquisition (paired-sample t test, t_{18} = 0.84, p = .4, d = .19) or reversal phases (t_{18} = 0.23, p = .8, d = .05).

We then examined the relationship between initial learning and perseveration, following den Ouden et al. (2013) (Figure 2B). LSD enhanced the relationship between the number of correct responses during the acquisition phase and the number of perseverative errors made during the subsequent reversal stage (acquisition correct responses [LSD minus placebo] versus reversal perseverative errors [LSD minus placebo]: linear regression coefficient β = .56, p = 0.002). Confirming this, making fewer errors during the acquisition phase predicted more perseverative errors when on LSD (β = 0.44, p = 0.003) but not when under placebo (β = 0.04, p = .8). Perseverative errors, a subset of all reversal errors, alone did not differ between conditions (t_{18} = 0.03, p = .98, d = .01).

Feedback sensitivity

We next assessed whether LSD influenced individuals' responses on trials immediately after positive versus negative feedback – whether participants stayed with the same choice after a win or a loss (win-stay/lose-stay; Figure 1D, 2D). Repeated-measures ANOVA with drug (LSD, placebo) and valence (win, loss) as within-subjects factors revealed a main effect of valence – participants "stayed" more after wins than losses

($F_{1,18}$ = 37.76, p = 8.0 × 10^{-6}, η_p^2 = 0.68) – and no main effect of LSD ($F_{1,18}$ = 0.20, p = .66, η_p^2 = .01). There was also no interaction of valence × LSD ($F_{1,18}$ = 0.63, p = .44, η_p^2 = .03).

Choice of reinforcement learning model

The core modelling results are displayed in Figure 2E. We fitted and compared three reinforcement learning models. Convergence was good with all three models having $\hat{R}$ < 1.2. Behaviour was best characterised by a reinforcement learning model with four parameters (Table 2). The four parameters in the winning model were: 1) reward learning rate, which reflects the degree to which the chosen stimulus value is increased following a positive outcome (reward prediction error); 2) punishment learning rate, degree to which the chosen stimulus value is decreased following a negative outcome (punishment prediction error); 3) reinforcement sensitivity (comparable to inverse temperature), which is the degree to which the values learned through reinforcement contribute to final choice; and 4) "stimulus stickiness", which indexes the tendency to get "stuck" to a stimulus and choose it because it was chosen on the previous trial, irrespective of outcome. The last two parameters resemble the explore/exploit trade-off: low values of stickiness or reinforcement sensitivity index two different types of exploratory behaviour.

Reward and punishment learning rates

The reward learning rate was significantly elevated on LSD (mean 0.87) compared to placebo (mean 0.28) (with the posterior 99.9% highest posterior density interval [HDI] of the difference between these means excluding zero; 0 ∉ 99.9% HDI; Figure 2E). There was also an increased punishment learning rate under LSD (mean 0.48) relative to placebo (mean 0.39) (drug difference, 0 ∉ 99% HDI). Importantly, LSD increased the reward learning rate to a greater extent than the punishment learning rate ($[\alpha^{rew,LSD} - \alpha^{rew,placebo}] - [\alpha^{pun,LSD} - \alpha^{pun,placebo}] > 0$; drug difference, 0 ∉ 99% HDI).

Stimulus stickiness and reinforcement sensitivity

Stimulus stickiness was lowered by LSD (mean 0.23) relative to placebo (mean 0.43) (drug difference, $0 \notin 90\%$ HDI; Figure 2E), whilst reinforcement sensitivity was not modulated by LSD (LSD mean 4.70, placebo mean 5.57; no drug difference, $0 \in 95\%$ HDI). This is in line with the absence of an effect of LSD on the tendency to 'stay' following reward or punishment (see conventional analysis above).

Discussion

There has been a recent surge of interest in potential therapeutic effects of psychedelics, particularly LSD. Theorising on the mechanisms of such effects centres on their role in enhancing learning and plasticity. In the current study we tested these postulated effects of LSD in flexible learning in humans and find that LSD increased learning rates as well as the impact of previously learnt values on subsequent perseverative behaviour. Specifically, LSD increased the speed at which value representations were updated following prediction error (the mismatch between expectations and experience). Whilst LSD enhanced the impact of both positive and negative feedback, it augmented learning from reward significantly more than it augmented learning from punishment.

The observation that LSD enhanced learning rates may be particularly important for understanding the mechanisms through which LSD might be therapeutically useful. Psychedelic drugs have been hypothesised to destabilise pre-existing beliefs (i.e. relax prior beliefs or "priors"), making them amenable to revision (Carhart-Harris and Friston 2019). The notion of relaxed priors is directly compatible with increased reinforcement learning rates: in our study, LSD rendered subjects more sensitive to prediction errors, which naturally implies downweighting of prior beliefs (Carhart-Harris and Friston 2019). That LSD affected a fundamental belief-updating process is notable given that psychedelics are under investigation trans-diagnostically for diverse clinical phenomena including depression (Carhart-Harris et al. 2016, 2018; Ross et al. 2016), anxiety (Griffiths et al. 2016; Grob et al. 2011), alcohol (Bogenschutz et al. 2015) and nicotine abuse (Johnson et al. 2014), OCD (Moreno et al. 2006), and eating disorders (Lafrance et

al. 2017); a unifying feature of these conditions is maladaptive associations in need of revision.

Given the broad effect of LSD on a range of neurotransmitter systems (Nichols 2004, 2016), it is not possible to determine the specific neurochemical mechanism underlying the observed LSD effects on learning. Nonetheless, obvious possibilities involve the serotonin and dopamine system, in particular 5-HT_{2A} and D_2 receptors (Marona-Lewicka et al. 2005, 2007; Nichols 2004, 2016). Specifically, the psychological plasticity purportedly promoted by psychedelics is believed to be mediated through action at 5-HT_{2A} receptors (Carhart-Harris and Nutt 2017) via downstream enhancement of NMDA (N-methyl-D-aspartate) glutamate receptor transmission (Barre et al. 2016) and brain-derived neurotrophic factor (BDNF) expression (Vaidya et al. 1997). The hypothesis that the present results are driven by serotonergic effects of LSD is supported by two recent studies in mice. Optogenetically stimulating dorsal raphé serotonin neurons enhanced reinforcement learning rates (Iigaya et al. 2018), whilst activation of these neurons tracked prediction errors during reversal learning (Matias et al. 2017).

In addition to affecting the serotonin system, however, LSD also acts at dopamine receptors (Nichols 2004, 2016). Dopamine has long been known to play a crucial role in belief updating following reward (Schultz et al. 1997), and more recent evidence shows that dopaminergic manipulations may alter learning rates (Schultz 2019; Swart et al. 2017). A dopaminergic effect would be in line with our previous study where genetic variation in the dopamine, but not serotonin transporter polymorphism, was associated with the same enhanced relationship between acquisition and perseveration as reported here under LSD (den Ouden et al. 2013).

Serotonin–dopamine interactions represent another candidate mechanism that could underlie the present findings. For example, stimulation of 5-HT_{2A} receptors in the prefrontal cortex of the rat, enhanced ventral tegmental area (VTA) dopaminergic activity (Bortolozzi et al. 2005). Indeed, the initial action of LSD at 5-HT_{2A} receptors has been proposed to sensitise dopamine neuron firing, which subsequently potentiates the direct

dopaminergic effects of LSD (Nichols 2016). LSD action at D_2 receptors, consequently, appears to be especially pronounced at a later time following LSD administration (Marona-Lewicka et al. 2005, 2007), which is relevant given the relatively long delay between LSD administration and performance of the current task (see Methods). However, arguing against a late dopaminergic effect is a previous study in rodents where the effects of LSD on reversal learning were consistent across four different time lags between drug administration and behavioural testing (King et al. 1974).

LSD impaired flexibility such that under LSD, better initial learning led to more perseverative responding. This result was in line with a recent study showing that LSD induced spatial working memory deficits and higher-order cognitive inflexibility in a set-shifting paradigm (Pokorny et al. 2019). Importantly, these effects were blocked by co-administration of the 5-HT_{2A} antagonist ketanserin (Pokorny et al. 2019), showing that the LSD-induced impairments were mediated by 5-HT_{2A} agonism, consistent with a 5-HT_{2A} mechanism underlying the present results. These data collectively suggest that LSD "stamps in" new learning following drug administration, which may subsequently be harder to update.

LSD's effects to induce cognitive inflexibility are ostensibly at odds with the observation that LSD enhanced plasticity (through enhanced learning rates). However, these results can be reconciled by considering the timing of drug administration with respect to initial learning and tests of cognitive flexibility. In both the present experiment and the previous set-shifting study (Pokorny et al. 2019), all phases of learning (acquisition and reversal) were conducted after LSD administration. In contrast, when acquisition learning was conducted prior to LSD administration, LSD resulted in improved reversal learning (using a reversal paradigm in rats; King et al. 1974). Likewise, when acquisition learning was conducted prior to administration of a 5-HT_{2A} antagonist, reversal learning was impaired (Boulougouris et al. 2008). Collectively, these findings suggest that whether a prior belief is down- or up-weighted under LSD may depend on whether the prior is formed before or during drug administration, respectively. This observation is of great

relevance for a putative therapeutic setting, where maladaptive beliefs will have been formed before treatment.

In summary, LSD enhanced the rate at which humans updated their beliefs based on feedback. Learning rate was most enhanced by LSD when receiving reward, and to a lesser extent following punishment. This study represents one of the few applications of objective measures to investigate fundamental cognitive processes in humans under LSD. These findings have implications for understanding the mechanisms through which LSD might be therapeutically useful for revising deleterious associations.

Methods

Subjects and drug administration

Nineteen healthy volunteers, over the age of 21, attended two sessions at least two weeks apart where they received either intravenous LSD (75µg in 10 mL saline) or placebo (10mL saline), in a single-blind within-subjects balanced-order design. All participants provided written informed consent after briefing on the study and screening. Participants had no personal history of diagnosed psychiatric disorder, or immediate family history of a psychotic disorder. Other inclusion criteria were normal electrocardiogram (ECG), routine blood tests, negative urine test for pregnancy and recent recreational drug use, negative breathalyser test for recent alcohol use, alcohol use limited to less than 40 units per week, and absence of a significant medical condition. Participants had previous experience with a classic psychedelic drug (e.g. LSD, mescaline, psilocybin/magic mushrooms, or DMT/ayahuasca) without an adverse reaction, and had not used these within six weeks of the study. Screening was conducted at the Imperial College London Clinical Research Facility (ICRF) at the Hammersmith Hospital campus, and the study was carried out at the Cardiff University Brain Research Imaging Centre (CUBRIC). Participants were blinded to the condition but the experimenters were not. A cannula was inserted and secured in the antecubital fossa and injection was performed over the course of two minutes. Participants reported noticing subjective effects of LSD 5 to 15 minutes after dosing. The PRL task was administered approximately five hours after injection. Once the subjective drug effects subsided, a psychiatrist assessed suitability for

discharge. This experiment was part of a larger study, the data from which are published elsewhere (e.g. Carhart-Harris et al. 2016). Additional information, including subjective ratings, can be found in Carhart-Harris et al. (2016).

Probabilistic reversal learning task

A schematic of the task is shown in Figure 1A. On every trial, participants could choose from three visual stimuli, presented at three of four randomised locations on a computer screen. In the first half of the task (40 trials), choosing one of the stimuli resulted in positive feedback in the form of a green smiling face on 75% of trials. A second stimulus resulted in positive feedback 50% of the time, whilst the third stimulus yielded positive feedback on only 25% of trials. Negative feedback was provided in the form of a red frowning face. The first stimulus that was selected, was defined as the initially rewarded stimulus; the choice on trial 1 always resulted in reward. The second stimulus that was selected was defined as the mostly punished stimulus, and by definition the 3rd stimulus was then the "neutral" stimulus. After 40 trials, the most and least optimal stimuli reversed, such that the stimulus that initially was correct 75% of the time was then only correct 25% of the time, and likewise the 25% correct stimulus then resulted in positive feedback on 75% of trials. This is a novel version (Kandroodi et al. 2020) of a widely used PRL task (Lawrence et al. 1999; den Ouden et al. 2013): novel due to the addition of a 50% "neutral" stimulus in order to distinguish learning to select the mostly rewarding stimulus from learning to avoid the mostly punishing stimulus.

Conventional analysis of behaviour

We examined whether LSD impaired participants' overall ability to perform the task by analysing the number of responses made to each stimulus during the acquisition and reversal phases. We measured feedback sensitivity by determining whether participants stayed with the same choice following positive or negative feedback (win-stay or lose-stay). The win-stay probability was defined as the number of times an individual repeated a choice after a win, divided by the number of trials on which positive feedback occurred (opportunities to stay after a win). Lose-stay probability was calculated in the same manner: number of times a choice was repeated following a loss, divided by the total

losses experienced. Note that in previous studies with a choice between only two stimuli (or responses), this metric is usually referred to as "win-stay / lose-shift", which also captures the tendency to repeat (rather than switch) responses following a win, and the tendency to switch (rather than repeat) choices following a loss. Random choice would result in 50% win-stay and 50% lose-shift; however, in the current paradigm with 3 stimuli, this base rate is 33% (win-)stay and 67% (lose-)shift. We therefore encode both variables with respect to the stay (rather than shift) rate, but they are still conceptually identical to earlier studies. Perseveration was defined according to den Ouden et al. (2013) and was assessed based on responses in the reversal phase. A perseverative error 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. The first trial in the reversal phase (trial 41 of 80) was excluded from the perseveration analysis, however, as at that point behaviour cannot yet be shaped by the new feedback structure. Note again that this metric is not entirely identical to the previous studies cited employing two stimuli, as the base-rate choice for each stimulus is now 1/3, so the "chance" level of perseverative errors is lower. Null hypothesis significance tests used $\alpha = 0.05$.

Computational modelling of behaviour

Model fitting, comparison, and interpretation

These methods are based on our previous work (Kanen et al. 2019). We fitted three reinforcement learning (RL) models to the behavioural data using a hierarchical Bayesian method, via Hamiltonian Markov chain Monte Carlo sampling implemented in Stan 2.17.2 (Carpenter et al. 2017). Convergence was checked according to $\hat{R}$, the potential scale reduction factor measure (Gelman et al. 2012; Brooks and Gelman 1998), which approaches 1 for perfect convergence. Values below 1.2 are typically used as a guideline for determining model convergence (Brooks and Gelman 1988). We assumed the three models had the same prior probability (0.33). Models were compared via a bridge sampling estimate of the marginal likelihood (Gronau et al. 2017a), using the "bridgesampling" package in R (Gronau et al. 2017b). Bridge sampling directly estimates the marginal likelihood, and therefore the posterior probability of each model given the data (and prior model probabilities), as well as the assumption that the models represent

the entire group of those to be considered. Posterior distributions were interpreted using the 95% highest posterior density interval (HDI), which is the Bayesian "credible interval". Parameter recovery for this modelling approach has been confirmed in a previous study (Kanen et al. 2019).

The Bayesian hierarchy consisted of "drug condition" at the highest level, and "subject" at the level below. For each parameter, each drug condition (e.g. LSD) had its own mean (with a prior that was the same across conditions, i.e. with priors that were unbiased with respect to LSD versus placebo). This was then merged with the intersubject variability (assumed to be normally distributed; mean 0 by definition, standard deviation determined by a further prior). The priors used for each parameter are shown in Table 1. For instance, the learning rate for a given subject under LSD was taken as: the group mean LSD value for learning rate, plus the subject-specific component of learning rate. The learning rate for a given subject under placebo was taken as: the group mean placebo value for learning rate, plus the subject-specific component of learning rate for the same subject. This method accounts for the within-subjects structure of the study design. This was done similarly (and separately) for all other model parameters.

To determine the change (LSD – placebo) in parameters, we calculated [group mean LSD learning rate] – [group mean placebo learning rate] for each of the ~8,000 simulation runs and tested them against zero via the HDI. This approach also removes distributional assumptions and provides an automatic multiple comparisons correction (Gelman and Tuerlinckx 2000; Gelman et al. 2012; Kruschke 2011).

Models

The parameters contained in each model are summarised in Tables 1 and 2. With Model 1, we tested the hypothesis that positive versus negative feedback guides behaviour differentially, and that LSD affects this. We augmented a basic RL model (Rescorla & Wagner 1972) with separate learning rates for reward α^{rew} and punishment α^{pun}. Positive feedback led to an increase in the value V_i of the stimulus i that was chosen, at a speed governed by the *reward learning 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 learning 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 three available choices, for a given trial, were then input to a standard softmax choice function to compute the probability of each choice:

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

for $n = 3$ 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 again augmented a simple RL model, but now also described the tendency to repeat a response, irrespective of the outcome that followed it (in other words, the tendency to "stay" regardless of outcome). With Model 2 we tested the hypothesis that LSD affects this basic perseverative tendency. This was implemented using a "stimulus stickiness" parameter τ^{stim}. The stimulus stickiness effect was modelled as $Q^{stim}_t = \tau^{stim}s_{t-1}$, where s_{t-1} was 1 for the stimulus that was chosen on the previous trial and was 0 for the other two stimuli. In this model we used only a single learning rate α^{reinf}. Positive reinforcement led to an increase in the value V_i of the stimulus i that was chosen, at a speed controlled by the *learning rate* α^{reinf}, via $V_{i,t+1} \leftarrow V_{i,t} + \alpha^{reinf}(R_t - V_{i,t})$. The final quantity controlling choice incorporated the additional stickiness parameter as $Q_t = Q^{reinf}_t + Q^{stim}_t$. Quantities Q, corresponding to the three choice options on a given trial, were then fed into the softmax function as above. It should be noted that if τ^{stim} is not in the model (or is zero), then τ^{reinf} is mathematically identical to the notion of softmax inverse

temperature typically implemented as β. The notation τ^{reinf} is used, however, because it contributes to Q^{reinf}_t but not to Q^{stim}_t. A standard implementation of β, by contrast, would govern the effects of both Q^{reinf}_t and Q^{stim}_t by weighting the sum of the two (Q_t).

Model 3 was the full model that incorporated separate reward and punishment learning rates as well as the stimulus stickiness parameter. With Model 3, we tested the hypothesis that LSD affects both how positive versus negative feedback guides behaviour differentially, and how LSD affects a basic perseverative tendency. Again, the final quantity controlling choice was determined by $Q_t = Q^{reinf}_t + Q^{stim}_t$.

Figures

Figure 1. A) Schematic of the probabilistic reversal learning task. Subjects chose one of three stimuli. The timeline of a trial is depicted: stimuli appear, a choice is made, the outcome is shown, a fixation cross is presented during the intertrial interval, stimuli appear for the next trial (etc.) (RT, reaction time). One stimulus delivered positive feedback (green smiling face) with a 75% probability, one with 50%, and one with 25%. The probabilistic alternative was negative feedback (red sad face). Midway through the task, the contingencies for the best and worst stimuli swapped. s = seconds. **B)** Trial-by-trial average probability of choosing each stimulus, averaged over subjects and sessions, collapsed across LSD and placebo sessions. A sliding 5-trial window was used for smoothing. The vertical dotted line indicates the reversal of contingencies. Shading indicates 1 standard error of the mean (SE). **C)** Distributions depicting the average per-subject probability (scattered dots) of choosing each stimulus during the acquisition (shown in dark blue) and reversal (light blue) phases, collapsed across LSD and placebo sessions. Mean value for each distribution is illustrated with a single dot at the base of each distribution, and the mean values for the probability of choosing different stimuli in each phase are connected by a line. One SE is shown by black error bars around the mean value. Horizontal dotted line indicates chance-level stay-behaviour (33%). **D)** Distributions depicting the average per-subject probability (scattered dots) of repeating a choice (staying) after receiving positive or negative feedback during the acquisition (dark blue) and reversal (light blue) phases, collapsed across LSD and placebo sessions. Horizontal dotted line indicates chance-level stay-behaviour (33%).

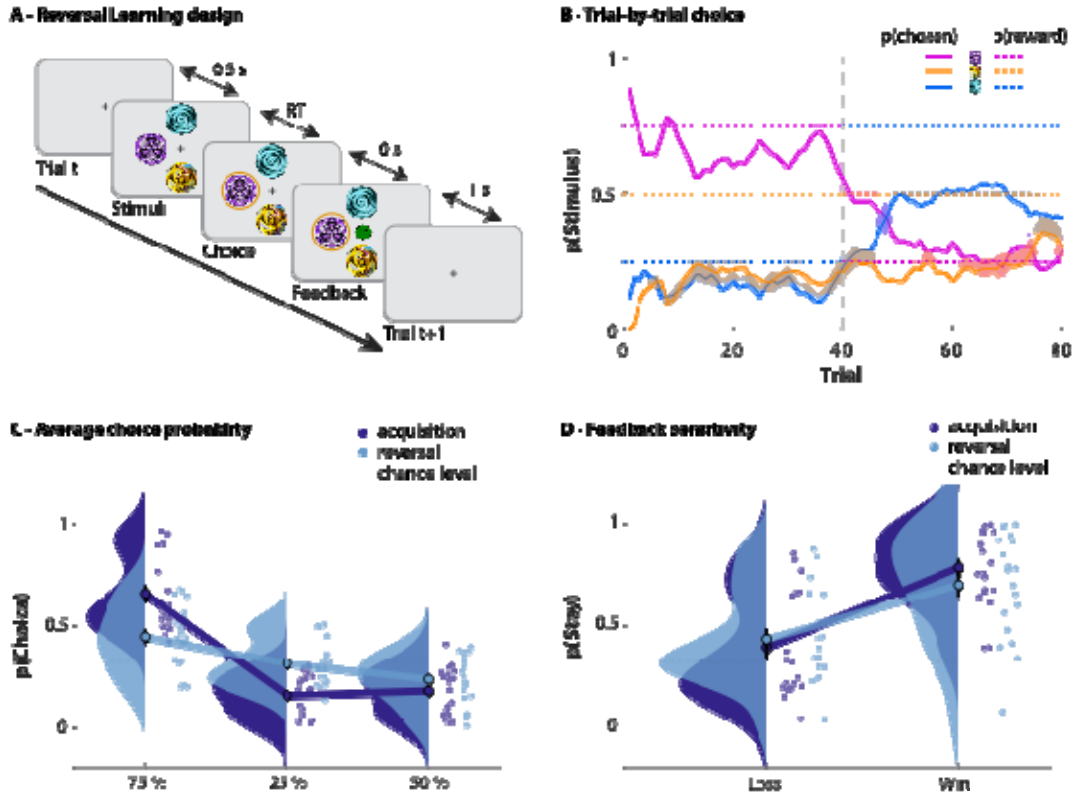
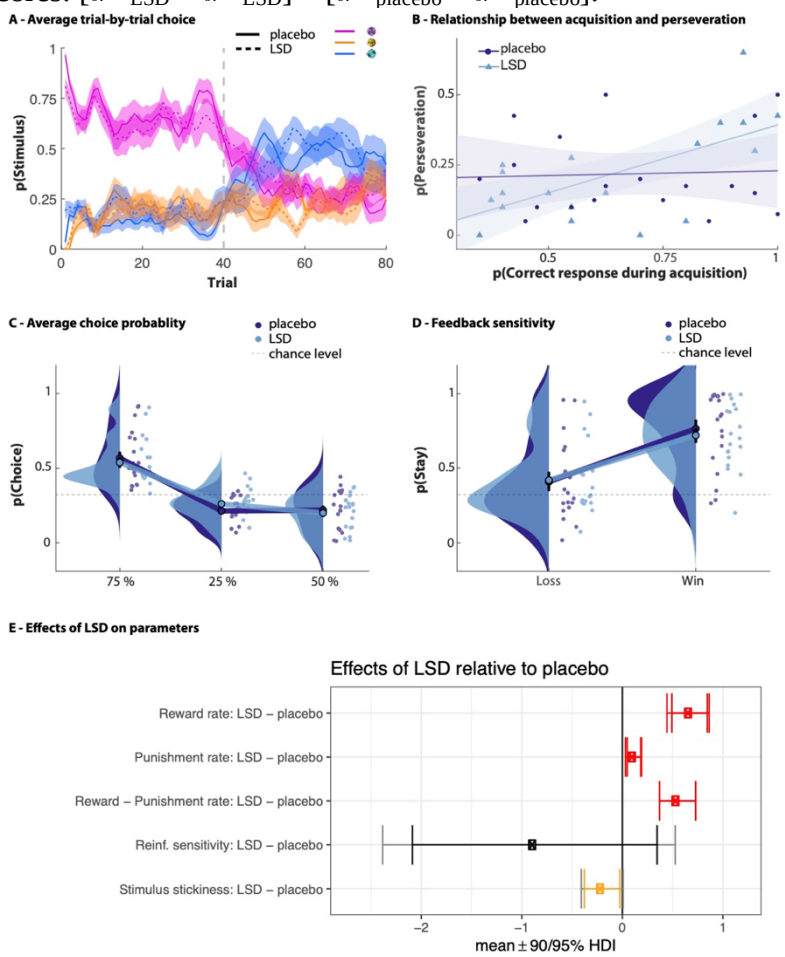


Figure 2. A) Trial-by-trial average probability of choosing each stimulus, averaged over subjects, separated by drug session. A sliding 5-trial window was used for smoothing. The vertical dotted line indicates the reversal of contingencies. Shading indicates 1 standard error of the mean (SE). **B)** Better initial learning was predictive of more perseveration on LSD and not on placebo. Shading indicates 1 SE. **C)** Distributions depicting the average per-subject probability (scattered dots) of choosing each stimulus while under placebo (shown in dark blue) and LSD (light blue). Mean value for each distribution is illustrated with a single dot at the base of each distribution, and the mean values for the probability of choosing different stimuli in each condition are connected by a line. One SE is shown by black error bars around the mean value. Horizontal dotted line indicates chance-level stay-behaviour (33%). **D)** Conventional analyses of feedback sensitivity were unaffected by LSD. Distributions depicting the average per-subject probability (scattered dots) of repeating a choice (staying) after receiving positive or negative feedback under placebo (dark blue) and LSD (light blue). Horizontal dotted line indicates chance-level stay-behaviour (33%). **E)** Effects of LSD relative to placebo on model parameters. Contrasts with the posterior 95% (or greater) highest posterior density interval [HDI] of the difference between means excluding zero ($0 \notin$ 95% HDI) are shown in red. Orange signifies $0 \notin$ 90% HDI. The third row represents a difference of differences scores: $[\alpha^{rew}{}_{\mathrm{LSD}} - \alpha^{pun}{}_{\mathrm{LSD}}] - [\alpha^{rew}{}_{\mathrm{placebo}} - \alpha^{pun}{}_{\mathrm{placebo}}]$.



Tables

Table 1. Prior distributions for model parameters

Model parameters	**Models using each parameter**	Range	**Prior**	**Reference**
reward learning rate, α^{rew}	1, 3	[0, 1]	Beta(1.2, 1.2)	den Ouden et al. (2013)
punishment learning rate, α^{pun}	1, 3	[0, 1]	Beta(1.2, 1.2)	den Ouden et al. (2013)
combined reward/punishment learning rate, α^{reinf}	2	[0, 1]	Beta(1.2, 1.2)	den Ouden et al. (2013)
reinforcement sensitivity, τ^{reinf}	1, 2, 3	[0, +∞]	Gamma(α=4.82, β=0.88)	Gershman (2016)
stimulus stickiness, τ^{stim}	2, 3	[−∞, +∞]	Normal(0, 1)	Christakou et al. (2013)
Intersubject variability in parameters				
Intersubject standard deviations for α^{rew}, α^{pun}, α^{reinf}, τ^{loc}	As above	[0, +∞]	Half-normal: Normal(0, 0.05) constrained to ≥0	Kanen et al. (2019)
Intersubject standard deviations for τ^{reinf}	As above	[0, +∞]	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)**
2	Model 1	α^{rew}, α^{pun}, τ^{reinf}	-2401.49	-33.28
3	Model 2	α^{reinf}, τ^{reinf}, τ^{stim}	-2428.52	-60.32
1	Model 3	α^{rew}, α^{pun}, τ^{reinf}, τ^{stim}	-2368.21	0

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

References

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, Alsiö J, Jupp B, et al (2015) Markers of Serotonergic Function in the Orbitofrontal Cortex and Dorsal Raphé Nucleus Predict Individual Variation in Spatial-Discrimination Serial Reversal Learning. Neuropsychopharmacology 40:1619–1630. doi: 10.1038/npp.2014.335

Barre A, Berthoux C, De Bundel D, et al (2016) Presynaptic serotonin 2A receptors modulate thalamocortical plasticity and associative learning. Proc Natl Acad Sci U S A 113:E1382–E1391. doi: 10.1073/pnas.1525586113

Bogenschutz MP, Forcehimes AA, Pommy JA, et al (2015) Psilocybin-assisted treatment for alcohol dependence: A proof-of-concept study. J Psychopharmacol 29:289–299. doi: 10.1177/0269881114565144

Bortolozzi A, Díaz-Mataix L, Scorza MC, et al (2005) The activation of 5-HT2A receptors in prefrontal cortex enhances dopaminergic activity. J Neurochem 95:1597–1607. doi: 10.1111/j.1471-4159.2005.03485.x

Boulougouris V, Castañé A, Robbins TW (2009) Dopamine D2/D3 receptor agonist quinpirole impairs spatial reversal learning in rats: Investigation of D3 receptor involvement in persistent behavior. Psychopharmacology (Berl) 202:611–620. doi: 10.1007/s00213-008-1341-2

Boulougouris V, Glennon JC, Robbins TW (2008) Dissociable effects of selective 5-HT2A and 5-HT2C receptor antagonists on serial spatial reversal learning in rats. Neuropsychopharmacology 33:2007–2019. doi: 10.1038/sj.npp.1301584

Brigman JL, Mathur P, Harvey-White J, et al (2010) Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. Cereb Cortex 20:1955–1963. doi: 10.1093/cercor/bhp266

Brolsma SCA, Vrijsen JN, Vassena E, et al (2020) Challenging the negative learning bias hypothesis of depression: reversal learning in a naturalistic psychiatric sample. Psychol Med 1–11

Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455

Carhart-Harris RL, Bolstridge M, Day CMJ, et al (2018) Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. Psychopharmacology (Berl) 235:399–408

Carhart-Harris RL, Friston KJ (2019) REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics. Pharmacol Rev 71:316–344. doi: 10.1124/pr.118.017160

Carhart-Harris RL, Nutt DJ (2017) Serotonin and brain function: A tale of two receptors. J Psychopharmacol 31:1091–1120. doi: 10.1177/0269881117725915

Carhart-Harris RL, Bolstridge M, Rucker J, et al (2016) Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. The Lancet Psychiatry 3:619–627. doi: 10.1016/S2215-0366(16)30065-7

Carhart-Harris RL, Muthukumaraswamy S, Roseman L, et al (2016) Neural correlates of the LSD experience revealed by multimodal neuroimaging. Proc Natl Acad Sci U S A 113:4853–4858. doi: 10.1073/pnas.1518377113

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

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

Christakou A, Gershman S, Niv Y, et al (2013) Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood. J Cogn Neurosci 25:1807–1823. doi: 10.1162/jocn_a_00447

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

Clarke HF, Hill GJ, Robbins TW, Roberts AC (2011) Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. J Neurosci 31:4290–4297. doi: 10.1523/JNEUROSCI.5066-10.2011

Cools R, Roberts AC, Robbins TW (2008) Serotoninergic regulation of emotional and behavioural control processes. Trends Cogn Sci 12:31–40. doi: 10.1016/j.tics.2007.10.011

Crockett MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. J Neurosci 29:11993–11999. doi: 10.1523/JNEUROSCI.2513-09.2009

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

Dayan P, Huys QJM (2009) Serotonin in Affective Control. Annu Rev Neurosci 32:95–126. doi: 10.1146/annurev.neuro.051508.135607

Deakin JFW (2013) The origins of “5-HT and mechanisms of defence” by Deakin and Graeff: a personal perspective. J Psychopharmacol 27:1084–9. doi: 10.1177/0269881113503508

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

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

Gelman A, Tuerlinckx F (2000) Type S error rates for classical and Bayesian single and multiple comparison procedures 1 Introduction. Comput Stat 15:373–90

Gershman SJ (2016) Empirical priors for reinforcement learning models. J Math Psychol 71:1–6. doi: 10.1016/j.jmp.2016.01.006

Geurts DEM, Huys QJM, den Ouden HEM, Cools R (2013) Serotonin and Aversive Pavlovian Control of Instrumental Behavior in Humans. J Neurosci 33:18932–18939. doi: 10.1523/JNEUROSCI.2749-13.2013

Goldberg SB, Pace BT, Nicholas CR, et al (2020) The experimental effects of psilocybin on symptoms of anxiety and depression: A meta-analysis. Psychiatry Res 284:. doi: 10.1016/j.psychres.2020.112749

Griffiths RR, Johnson MW, Carducci MA, et al (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. J Psychopharmacol 30:1181–1197. doi: 10.1177/0269881116675513

Grob CS, Danforth AL, Chopra GS, et al (2011) Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. Arch Gen Psychiatry 68:71–78. doi: 10.1001/archgenpsychiatry.2010.116

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:

Harvey JA, Gormezano I, Cool-Hauser VA, Schindler CW (1988) Effects of LSD on classical conditioning as a function of CS-UCS interval: Relationship to reflex facilitation. Pharmacol Biochem Behav 30:433–441. doi: 10.1016/0091-3057(88)90477-7

Harvey JA (2003) Role of the Serotonin 5-HT2A Receptor in Learning. Learn Mem 10:355–362. doi: 10.1101/lm.60803

Iigaya K, Fonseca MS, Murakami M, et al (2018) An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. Nat Commun 9:10–12. doi: 10.1038/s41467-018-04840-2

Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR (2014) Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. J Psychopharmacol 28:983–992. doi: 10.1177/0269881114548296

Kandroodi MR, den Ouden HEM (2020) Effects of methylphenidate on reversal learning depend on working memory capacity. PsyArXiv

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

King AR, Martin IL, Melville KA (1974) Reversal learning enhanced by lysergic acid diethylamide (LSD): Concomitant rise in brain 5-hydroxytryptamine levels. Br J Pharmacol 52:419–426. doi: 10.1111/j.1476-5381.1974.tb08611.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

Lafrance A, Loizaga-Velder A, Fletcher J, et al (2017) Nourishing the Spirit: Exploratory Research on Ayahuasca Experiences along the Continuum of Recovery from Eating Disorders. J Psychoactive Drugs 49:427–435. doi: 10.1080/02791072.2017.1361559

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

Lawrence AD, Sahakian BJ, Rogers RD, et al (1999) Discrimination, reversal, and shift learning in Huntington's disease: Mechanisms of impaired response selection. Neuropsychologia 37:1359–1374. doi: 10.1016/S0028-3932(99)00035-4

Lee B, Groman S, London ED, Jentsch JD (2007) Dopamine D2/D3 receptors play a specific role in the reversal of a learned visual discrimination in monkeys. Neuropsychopharmacology 32:2125–2134. doi: 10.1038/sj.npp.1301337

Marona-Lewicka D, Nichols DE (2007) Further evidence that the delayed temporal dopaminergic effects of LSD are mediated by a mechanism different than the first temporal phase of action. Pharmacol Biochem Behav 87:453–461. doi: 10.1016/j.pbb.2007.06.001

Marona-Lewicka D, Thisted RA, Nichols DE (2005) Distinct temporal phases in the behavioral pharmacology of LSD: Dopamine D2 receptor-mediated effects in the rat and implications for psychosis. Psychopharmacology (Berl) 180:427–435. doi: 10.1007/s00213-005-2183-9

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

Moreno FA, Wiegand CB, Taitano EK, Delgado PL (2006) Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. J Clin Psychiatry 67:1735–1740. doi: 10.4088/JCP.v67n1110

Nichols DE (2004) Hallucinogens. Pharmacol Ther 101:131–181. doi: 10.1016/j.pharmthera.2003.11.002

Nichols DE (2016) Psychedelics. Pharmacol Rev 68:264–355

Nord M, Finnema SJ, Halldin C, Farde L (2013) Effect of a single dose of escitalopram on serotonin concentration in the non-human and human primate brain. Int J Neuropsychopharmacol 16:1577–1586. doi: 10.1017/S1461145712001617

Nutt DJ, Carhart-Harris RL (2020) The Current Status of Psychedelics in Psychiatry. JAMA Psychiatry E1–E2

Pokorny T, Duerler P, Seifritz E, et al (2019) LSD acutely impairs working memory, executive functions, and cognitive flexibility, but not risk-based decision-making. Psychol Med 1–10. doi: 10.1017/s0033291719002393

Rescorla RA, Wagner AR (1972) A theory of classical conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF (eds) Classical Conditioning II Current Research and Theory. Appleton-Century-Crofts, New York, pp 64–99

Romano AG, Quinn JL, Li L, et al (2010) Intrahippocampal LSD accelerates learning and desensitizes the 5-HT2A receptor in the rabbit. Psychopharmacology (Berl) 212:441–448. doi: 10.1007/s00213-010-2004-7

Ross S, Bossis A, Guss J, et al (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. J Psychopharmacol 30:1165–1180. doi: 10.1177/0269881116675512

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

Schindler CW, Gormezano I, Harvey JA (1986) Effect of LSD on acquisition, maintenance, extinction and differentiation of conditioned responses. Pharmacol Biochem Behav 24:1293–1300. doi: 10.1016/0091-3057(86)90187-5

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science (80-) 275:1593–1599. doi: 10.1126/science.275.5306.1593

Schultz W (2019) Recent advances in understanding the role of phasic dopamine activity [version 1; peer review: 3 approved]. F1000Research 8:1–12. doi: 10.12688/f1000research.19793.1

Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. Science (80-) 321:848–851. doi: 10.1126/science.1160575

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

Swart JC, Froböse MI, Cook JL, et al (2017) Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. Elife 6:1–36. doi: 10.7554/eLife.22169

Vaidya VA, Marek GJ, Aghajanian GK, Duman RS (1997) 5-HT2A receptor-mediated regulation of brain-derived neurotrophic factor mRNA in the hippocampus and the neocortex. J Neurosci 17:2785–2795. doi: 10.1523/jneurosci.17-08-02785.1997

Vollenweider FX, Preller KH (2020) Psychedelic drugs: neurobiology and potential for treatment of psychiatric disorders. Nat Rev Neurosci. doi: 10.1038/s41583-020-0367-2

Walker SC, Robbins TW, Roberts AC (2009) Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. Cereb Cortex 19:889–898. doi: 10.1093/cercor/bhn136

Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. Nat Rev Neurosci 7:464–476. doi: 10.1038/nrn1919

Competing Interests Statement

T.W.R. discloses consultancy with Cambridge Cognition, Greenfields Bioventures and Unilever; he receives research grants from Shionogi & Co and GlaxoSmithKline and royalties for CANTAB from Cambridge Cognition and editorial honoraria from Springer Verlag and Elsevier. R.N.C. consults for Campden Instruments and receives royalties from Cambridge Enterprise, Routledge, and Cambridge University Press. H.E.M.d.O has consulted on task design and data analysis for Eleusis Benefit Corp but does not own stocks or shares. J.W.K., R.L.C-H, Q.L., and M.R.K. declare no conflicts of interest.

Acknowledgements

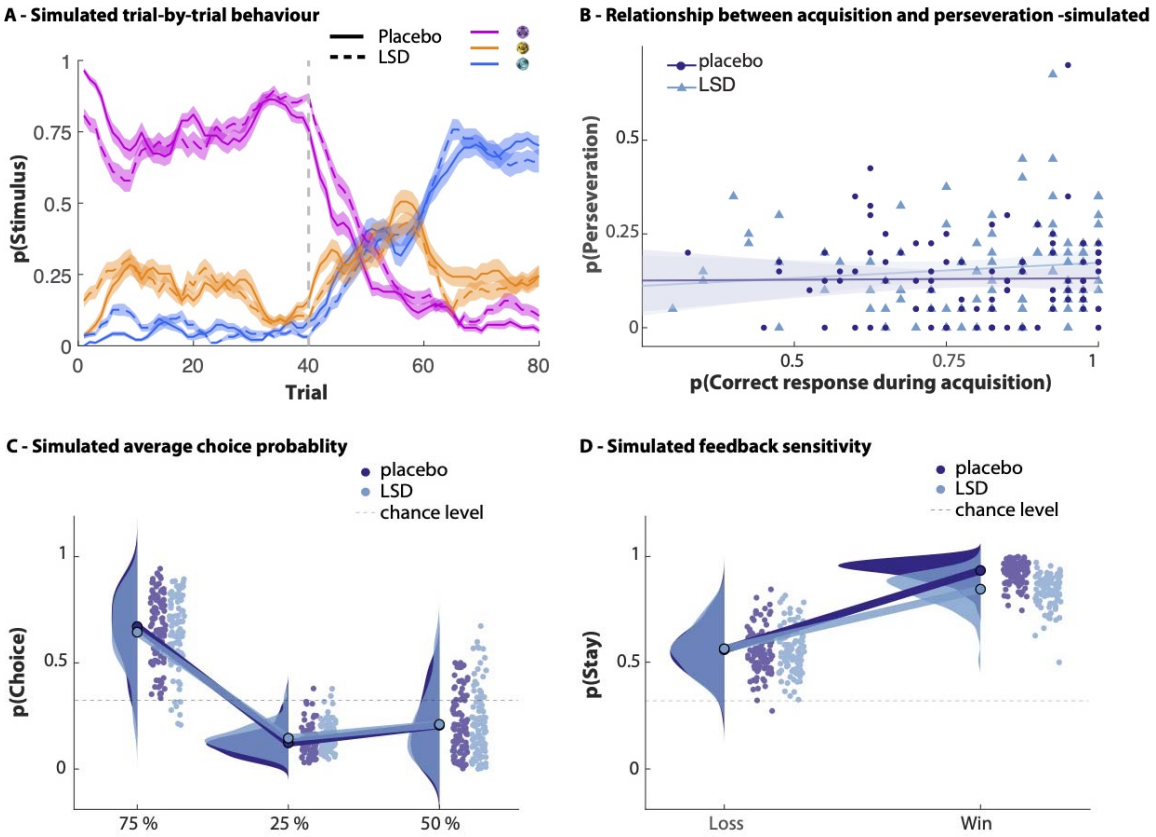
This study was funded by the Walacea.com crowdfunding campaign and the Beckley Foundation, awarded to R.L.C-H. J.W.K. was supported by a Gates Cambridge Scholarship and an Angharad Dodds John Bursary in Mental Health and Neuropsychiatry, T.W.R. by a Wellcome Trust Senior Investigator Grant 104631/Z/14/Z, and H.E.M.d.O. by the Netherlands Organisation for Scientific Research, NWO. 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.

Supplementary information

Simulation: Methods

We simulated behavioural data from the winning model to determine how behavioural patterns in the synthetic data compared to the raw data. Simulated data were analysed for win-stay probability, lose-stay probability, acquisition performance, and perseveration, as was done for the original raw data analysis. For each condition (placebo and LSD), we simulated 100 "virtual subjects" using the posterior mean parameters from that condition, from the winning model, per Kanen et al. (2019).

Simulation: Results

Simulated behavioural data, generated using parameter estimates from the winning model, were analysed using conventional methods in order to assess whether the winning model could capture the observed effects of LSD on raw behaviour. Simulated data are shown in Supplementary Figure 1. Consistent with the original data, lose-stay probability was unaffected by LSD in the simulated behaviour (t_{99} = –0.37, p = .71, d = .03) and acquisition performance was also unaffected (t_{99} = 0.25, p = .81, d = .03). Perseveration was enhanced by LSD in the simulation (t_{99} = –2.24, p = 0.03, d = .22), which differs slightly from, yet is in line with, the original analyses showing an enhanced relationship between acquisition and perseveration under LSD. Linear regression examining whether correct responses during the acquisition phase (LSD minus placebo) predicted more perseverative errors in the reversal stage (LSD minus placebo) was not significant in the simulated data (β = 0.15, p = 0.13). Separate regressions for each condition also showed no significant relationship between acquisition performance and perseveration for LSD (β = 0.12, p = 0.25) or for placebo (β = 0.01, p = .91). Win-stay probability was diminished under LSD in the simulated data (t_{99} = 11.91, $p = 8.21 \times 10^{-21}$, d = 1.19) whereas it was unaffected by LSD in the raw data analysis.

Supplementary Figure 1. Simulated data. **A)** Trial-by-trial average probability of choosing each stimulus, averaged over simulated subjects, separated by drug session. A sliding 5-trial window was used for smoothing. The vertical dotted line indicates the reversal of contingencies. Shading indicates 1 standard error of the mean (SE). **B)** Relationship between initial learning and perseveration on LSD versus placebo in simulated data. Shading indicates 1 SE. **C)** Distributions depicting the average per-subject probability (scattered dots) of simulated subjects choosing each stimulus while under placebo (shown in dark blue) and LSD (light blue). Mean value for each distribution is illustrated with a single dot at the base of each distribution, and the mean values for the probability of choosing different stimuli in each condition are connected by a line. One SE is shown by black error bars around the mean value. Horizontal dotted line indicates chance-level stay-behaviour (33%). **D)** Distributions depicting the average per-subject probability (scattered dots) of simulated subjects repeating a choice (staying) after receiving positive or negative feedback under placebo (dark blue) and LSD (light blue). Horizontal dotted line indicates chance-level stay-behaviour (33%).