Running Head: MUD Learning and Inflexibility

Word Count: 4803/5000

Are methamphetamine users compulsive? Faulty reinforcement learning, not inflexibility, underlies decision-making in people with methamphetamine use disorder

Alex H. Robinson[a], José C. Perales[b], Isabelle Volpe[c,d,e], Trevor T.-J. Chong[a],
Antonio Verdejo-Garcia[a*]

Addiction Biology

[a] Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia

[b] Department of Experimental Psychology, Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain

[c] Turning Point, Eastern Health, Melbourne, Australia

[d] Eastern Health Clinical School, Monash University, Melbourne, Australia

[e] Monash Addiction Research Centre, Monash University, Melbourne, Australia

***Corresponding Author:**
Antonio Verdejo-Garcia
Turner Institute for Brain and Mental Health
Monash University 18 Innovation Walk, Clayton VIC 3800 Australia
Antonio.Verdejo@monash.edu
Telephone: +61 3 9905 5374
Fax: +61 3 9905 3948

Abstract (250/250 words).

Methamphetamine Use Disorder involves continued use of the drug despite negative consequences. Such “compulsivity” can be measured by reversal learning tasks, which involve participants learning action-outcome task contingencies (acquisition-contingency), and then updating their behaviour when the contingencies change (reversal). Using these paradigms, animal models suggest that people with Methamphetamine Use Disorder (PwMUD) may struggle to avoid repeating actions that were previously rewarded but are now punished (inflexibility). However, difficulties in learning task contingencies (reinforcement learning) may offer an alternative explanation, with meaningful treatment implications. We aimed to disentangle inflexibility and reinforcement learning deficits in 35 PwMUD and 32 controls with similar sociodemographic characteristics, using novel trial-by-trial analyses on a probabilistic reversal learning task. Inflexibility was defined as 1) weaker reversal phase performance, compared to the acquisition-contingency phases; and 2) persistence with the same choice despite repeated punishments. Conversely, reinforcement learning deficits were defined as 1) poor performance across both acquisition-contingency and reversal phases; and 2) inconsistent post-feedback behaviour (i.e., switching after reward). Compared to controls, PwMUD exhibited weaker learning (*OR* = 0.69, *95% CI* [0.63 – 0.77], *p* <.001), though no greater accuracy reduction during reversal. Furthermore, PwMUD were more likely to switch responses after one reward/punishment (*OR* = 0.83, *95% CI* [0.77 – 0.89], *p* <.001; *OR* = 0.82, *95% CI* [0.72 – 0.93], *p* = .002) but just as likely to switch after repeated punishments (*OR* = 1.03, *95% CI* [0.73 – 1.45], *p* = .853). These results indicate that PwMUD’s reversal learning deficits are driven by weaker reinforcement learning, not inflexibility.

Key words: cognitive inflexibility, compulsivity, methamphetamine use disorder, reinforcement learning, reversal learning.

1. Introduction

People with Methamphetamine Use Disorder (PwMUD) continue to use methamphetamine despite experiencing growing negative consequences from their drug use (i.e., mental/physical illness, legal and financial problems, relationship loss).[1-2] This behaviour is defined as compulsive substance use and is a hallmark of addiction.[3] Internationally, there are concerns that the prevalence of Methamphetamine Use Disorder (MUD) is increasing, particularly due to the growing availability of high-purity crystal methamphetamine and its related harmful patterns of use.[4-5] Furthermore, although current treatments reduce short-term methamphetamine use and psychological distress,[6] most clients relapse within one year of treatment.[7-8] This difficulty in controlling drug use warrants a more nuanced understanding of "compulsivity" amongst PwMUD.

While several cognitive processes likely underlie compulsive behaviour (i.e., habit formation, avoidance),[9] contingency-based cognitive inflexibility appears the most prominent cognitive driver of compulsivity in addiction.[10] This is because inflexibility refers to a difficulty in updating behavioural responses that were initially associated with reward but are now associated with punishment.[10] It is defined separately from more basic reinforcement learning deficits, which involve the ability to learn which actions or stimuli are associated with obtaining reinforcers.[11] Contingency-based inflexibility has typically been measured using reversal learning tasks, whereby initial stimulus-outcome associations are learnt (either implicitly or explicitly; acquisition phases) and then changed throughout the measure (reversal phases).[10,12] Errors in the reversal phase (i.e. responding to previously reinforced stimuli post-reversal) are often used as the behavioural index for inflexibility on these paradigms.[12]

Previous research using reversal learning tasks has extensively examined the effects of methamphetamine exposure in rodents and non-human primates. Amongst these individuals, weaker reversal learning has been frequently reported and interpreted as an indication towards broader inflexibility and/or compulsivity.[13 – 19] In addition, abnormal performance on these paradigms has been linked to dysregulation of striatal dopamine (D2-type) and frontal serotonin systems likely caused by methamphetamine consumption.[15,19]

However, despite the consistent cross-species evidence for inflexibility, and the popularity of the compulsivity account of methamphetamine addiction, only one study has examined contingency-based reversal learning in humans.[20] This work also found worse reversal learning performance amongst PwMUD, when compared to healthy controls. However, closer inspection of learning rates also suggested that PwMUD made more errors in the tasks' acquisition phase, indicating that PwMUD may instead have fundamental deficits in reinforcement learning. Reappraisal of the animal literature also suggests reinforcement learning deficits versus (or in addition to) inflexibility.[17,19,21] For example, weaker reversal performance by methamphetamine-treated monkeys can be alleviated when given enough practice to master their knowledge of acquisition contingencies.[17]

Understanding whether PwMUD have deficits in reinforcement learning and/or cognitive inflexibility is important for several reasons. From a treatment perspective, the improvement of either construct requires different, targeted approaches. For example, reinforcement learning difficulties may benefit from the implementation of Contingency Management (CM) programs,[22] which may overcome participants' weakened learning responses via more immediate and/or obvious reinforcements. In contrast, cognitive inflexibility may require extinction/response prevention therapies, such as Cognitive Behavioural Therapy (CBT).[23] From an ethical standpoint, describing someone's learning as inflexible (or drug use as "compulsive") may be more likely to generate feelings of hopelessness among patients and clinicians, relative to a perspective of reduced learning.[24]

Unfortunately, prior research using reversal learning tasks may have used methods which conflated cognitive inflexibility and reinforcement learning. This is because appropriate reversal learning requires the ability to: 1) initially learn task contingencies (reinforcement learning); and 2) update behaviour when contingencies change (cognitive flexibility). Thus, traditional measures (i.e., errors in reversal phase)[12] are likely impacted by both processes.

To amend this, and thus disentangle inflexibility and reinforcement learning deficits, we conducted detailed novel trial-by-trial analysis of a reversal learning task amongst a cohort of PwMUD. We reasoned that inflexibility would manifest as: 1) a significant reduction in learning rates between the acquisition-contingency and reversal phases, whereby the latter shows significantly poorer learning; and 2) the maintenance of a certain action despite receiving multiple instances of punishment. In contrast, reinforcement learning deficits would manifest as: 1) consistently poor learning rates across both acquisition-contingency and reversal phases; and 2) an inconsistent pattern of behaviour after feedback (i.e., increased switching after reward/punishment). Based on prior clinical research in PwMUD[20] and detailed models of learning/inflexibility in non-human primates,[17] we hypothesized that PwMUD's behaviour on the reversal learning task would reflect reinforcement learning deficits, rather than cognitive inflexibility, compared to drug-naïve controls.

2. Materials and Methods

2.1. Design

Cross-sectional, observational design to characterise differences between PwMUD and drug-naïve controls on a reversal learning task.

2.2. Participants

Thirty-five PwMUD (M_{Age} = 33.26, *SD* = 7.78, 24 males) were compared to 32 drug-naïve controls (M_{Age} = 31.44, *SD* = 9.54, 20 males). PwMUD were recruited from public and private drug and alcohol treatment services across Melbourne, including inpatient detoxification/rehabilitation and outpatient counselling settings. The key criterion for inclusion of PwMUD was admission into treatment for methamphetamine use. However, two PwMUD had yet to formally commence treatment, and we based their inclusion on scores >4 on the Severity of Dependence Scale for methamphetamine dependence.[25,26] Table 1 presents PwMUD's patterns of methamphetamine consumption and treatment information, while Table 2 reports PwMUD's secondary substance and medication use. All PwMUD reported crystal methamphetamine as the predominant form used. Drug-naïve controls with similar socio-demographic characteristics (sex, age, education, IQ) were recruited using online and community advertisements. Exclusion criteria for all participants included diagnosis of schizophrenia or intellectual disability. PwMUD were required to have been abstinent for longer than 48 hours but less than 12 months.

Insert Table 1

Insert Table 2

2.3 Procedure

The Eastern Health Human Research Ethics Committee approved the study (E52/1213). Recruitment occurred between June 2017 and September 2018. We tested PwMUD at their treatment facility and controls at Monash University. However, when these premises were not convenient, we also used community libraries. Participants were screened before undergoing a standardised assessment session which lasted between 1 to 1.5 hours. Reimbursement included a $20 (AUD) gift card.

2.4. Measures

2.4.1. Sociodemographic and mental health characteristics

Participants self-reported their age and education, while IQ and intellectual disability were assessed/screened using the *National Adult Reading Test* (NART).[27] Depressive symptomatology was assessed using the *Centre for Epidemiologic Studies Depression Scale* (CES-D).[28]

2.4.2. Methamphetamine use

Frequency of methamphetamine consumption in the last month of use was collected using *The Timeline Follow Back interview* (TLFB).[29] PwMUD's degree of methamphetamine dependence was assessed using the *Severity of Dependence Scale* (SDS). [25,26]

2.4.3. Cognitive inflexibility vs. reinforcement learning

The *Probabilistic Reversal Learning Task* (PRLT)[30,31] is a computerised measure that requires participants to learn which of two different coloured squares is more rewarding (see Figure S1 for task diagram). On each trial, participants were presented with two coloured squares (one red and one green; all participants denied colour-blindness) on the left and right of the screen (randomly allocated). They were informed that, on any given trial, one square was "correct" (i.e., usually associated with a gain of two points and a positively-valenced "winning chime" sound), while the other was "incorrect" (i.e., usually associated with a loss of two points and a negatively-valenced "boh!" sound). Participants were instructed to select the square they believed was the more frequently rewarded stimulus based on the feedback that they received to that point. After making their response, they then received feedback on their choice, and the next trial would then be presented without an intertrial interval.

Overall, the task was separated into four, 40-trial phases. Phases one and three were acquisition-contingency phases, whereby participants attempted to learn the initial contingencies and reinforcement probabilities associated with each square. Phases two and four were reversal phases, whereby the rewarding/punishing elements of the stimuli were switched, and participants had to update their behaviour. In the first two phases, the reward/punishment rates were set at 80/20% for the correct square and vice-versa for the incorrect square. In the final two phases, the reward/punishment rates changed to 70/30%.[31,32] All phases immediately followed one-another, with no breaks or signalling to participants. The PRLT has been frequently used in prior addiction research.[31,32]

2.5. Statistical Analysis

To achieve our aim of disentangling reinforcement learning and inflexibility on the PRLT, our analyses focused on exploring behaviour related to accuracy (i.e., selecting the correct stimulus) and feedback (i.e., reward or punishment), across both shorter (i.e., individual trials) and longer time frames (i.e., across phases or multiple trials). As such, our main analyses are divided into two sections: 1) Trial-By-Trial Performance Across and Within

Phases; and, 2) Switch/Stay analyses. How we used each section to investigate inflexibility or reinforcement learning deficits are described in the relevant discussions below. For both of these approaches, we used a recommended series of Generalized Mixed-Effects Model stepwise (backward deletion) comparisons.[33] Briefly, this involved comparing models in a hierarchical order (using AIC/BIC as the measures of model fit measures), beginning with a "saturated model" (a model including all possible relevant effects). This "saturated model" was then compared to a simpler model which includes all the same predictors, except for the most complex effect/interaction. If the reduced model was of equal or better fit, it was then used for comparison with a further simplified version. We present results from both "saturated" and any "best-fitting" models to conservatively confirm results. We used the *lme4* package[34] in *R*[35] to create the models, while relevant quantitative predictors were zero-centred, and *p*-values obtained via *z*-test approximations. Alpha was set at $\alpha = .05$ for main analyses, and $\alpha = .017$ for post-hoc contrasts due to multiple comparisons. Potential confounders (i.e., age, education, IQ, and depression) were compared using Mann-Whitney *U*-tests and Bayes Factors in *JASP*.[36] To corroborate our main results, we applied a more traditional ANOVA and *t*-test approach, which is described in the relevant results section (3.4. Traditional Analyses of Reversal Learning). We also conducted control analyses to investigate if attention, motivation, severity of methamphetamine dependence, and severity of any potential comorbid cannabis dependence were impacting PwMUD's behaviour (3.5. Control Analyses).

2.5.1. Trial-By-Trial Performance Across and Within Phases

These analyses investigated group differences on trial-by-trial accuracy within and across phases. Model construction began by assuming that participants' within-phase learning curves were negatively accelerated (a trend previously observed on PwMUD's choice data when performing reversal learning[20]). Essentially, this theorises that early experiences contribute the most information when learning new behaviours.[37] To achieve this, we modelled Accuracy (choosing correctly or not on each trial) as a binomial variable (using a logit link function), and logarithmically transformed Trial (so that the underlying learning process is assumed to be linear relative to Log-trial). This transformation of the effect of trial outperformed both its original linear counterpart and a polynomial-based approach (another way to model negatively accelerating curves, see Supplementary Materials). Therefore, further mentions of "Trial" will reference the use of this logarithmic transformation.

After this, we defined Accuracy as the output variable for these analyses, and built models using a mixture of the following fixed effects: Phase (1-4), Trial within Phase (1-40), Group (PwMUD vs. drug-naïve controls), and any relevant covariates and interactions. Participant was always included as a random intercept.

To test inflexibility, we included a set of contrasts which investigated different patterns of change in accuracy rates across phases. The main contrast of interest, C1 (1, -1, 1, -1), compared phases one and three (acquisition contingencies) versus phases two and four (reversed contingencies). Contrast C2 (1, 1, -1, -1) modelled the effect of contingency degradation in the second half of the task (i.e., 80/20% versus 70/30% phases). Contrast C3 (1, -1, -1, 1) was theoretically irrelevant, but included to ensure the contrasts were comprehensive and orthogonal.[38] C1 is presented in the results section, while C2 and C3 are reported in Table 4. To aid in interpretation, a significant odds ratio (*OR*) > 1 for C1 would indicate that correct responses were more likely in phases one and three, relative to two and four (i.e., evidence that participants are showing reversal cost). In group interactions, greater inflexibility by PwMUD, relative to controls would be indicated by a C1 x Group *OR* > 1.

In contrast, to test reinforcement learning difficulties in these analyses we used the Group x Trial interaction. This investigates whether groups showed differences in their accuracy across the trials (i.e., accuracy should increase with the number of trials). An *OR* < 1 would indicate towards PwMUD exhibiting weaker learning, relative to controls.

2.5.2. Switch/Stay Analysis

These analyses aimed to investigate group differences on shifting after reward and punishment, the latter reflecting inflexibility. The dependent variable was defined dichotomously as Stay (0; repeating the previous choice), or Switch (1; making a different choice to the previous trial). The main predictor was Accumulated Feedback. This was coded as 3 (three consecutive punishments whereby participant did not change response in the last two trials), 2 (two consecutive punishments whereby participant did not change response in the last trial), 1 (previous trial punished), and 0 (previous trial rewarded). The best fitting model was selected using a combination of the following effects: Accumulated Feedback, Group, and any relevant covariates. Participant was entered as a random intercept.

Three new contrasts examined differences in Accumulated Feedback. C1 (-3. 1. 1. 1) compared behaviour after one reward to one/two/three instances of consecutive punishments. C2 (0, -2, 1, 1) compared behaviour after one punishment to two/three consecutive punishments. C3 (0, 0, -1, 1) compared behaviour after two consecutive punishments to three consecutive punishments. For interpretation, *OR*s > 1 for C1, C2 and C3 would indicate that participants are: C1) more likely to switch after any amount of accumulated negative feedback relative to reward (or less likely to switch after reward compared to any accumulated negative feedback); C2) more likely to switch after two or three instances of accumulated negative feedback relative to a single instance (or less likely to switch after one negative accumulated feedback compared to two or three); or, C3) more likely to switch after three instances of accumulated negative feedback relative to two (or less likely to switch after two negative feedbacks compared to three). In the Contrast x Group interactions, *OR*s > 1 would indicate that such contrast effects are larger for PwMUD, and *OR*s < 1 would indicate the contrast effects are larger in controls. Evidence towards inflexibility would be manifest if PwMUD were more likely to repeat actions despite multiple negative feedbacks, relative to controls (statistically, C3 *OR* > 1 and C3 x Group *OR* < 1). Evidence toward reinforcement learning deficits would be primarily manifest if the PwMUD group were more likely to switch after reward, relative to controls (statistically, C1 *OR* > 1, C1 x Group *OR* < 1).

3. Results

3.1. Descriptive Statistics Between Groups

Groups were matched in sex, age, education and IQ, but not for depression scores (Table 1) which were added as a covariate in subsequent analyses.

3.2. Trial-by-Trial Performance Across and Within Phases

Figure 1 displays the observed proportion of correct responses for each phase and trial of the task, and each group. Visually, drug-naïve controls showed steeper learning functions in all phases, as well as greater "reversal costs" at the start of each phase. Such observations likely reflect that controls were performing more accurately by the end of each phase, and therefore required a greater adjustment of their behaviour after reversal.

Insert Figure 1

To identify which variables best explained these group differences we began the model comparisons (see Table 3 for comparisons and Table 4 for statistics of saturated and best-fitting models). First, we checked whether there were indeed learning differences between PwMUD and controls. As such, a “Saturated No-Group-Learn” model and a “Saturated Group-Learn” model were compared. The Saturated No-Group-Learn Model included Accuracy as the outcome variable; Participant as a random intercept; and the following fixed-effects: Trial, Phase, Depression, Phase x Trial and Depression x Trial (the only Depression interaction identified in prior analysis). The Saturated Group-Learn Model included all the No-Group factors as well as all relevant Group effects and interactions (Group, Group x Phase, Group x Trial, and Group x Phase x Trial). When compared, the Saturated Group-Learn Model provided a better fit, indicating that group effects/interactions likely described our data.

We then attempted to remove any unnecessary effects from the Saturated Group-Learn Model, beginning with the most complex interactions. Simplified Group-Learn Model 1 removed the three-way interaction (Group x Phase x Trial) without losing fit and was thus used for the following comparisons. Removal of Group x Trial (Simplified Group-Learn Model 2.1) and Phase x Trial (Simplified Group-Learn Model 2.3) reduced fit, meaning these predictors were useful in explaining participants behaviour. In contrast, removal of Group x Phase (Simplified Group-Learn Model 2.2) did not reduce model fit and was removed due to parsimony. In this manner, Simplified Group-Learn Model 2.2 became the best-fitting model.

Insert Table 3

Because Simplified Group-Learn Model 2.2 did not include the interactions of Group x Phase and Group x Phase x Trial, it appeared there were no differences between PwMUD and controls in their performance between phases (i.e. no inflexibility). In contrast, the retention of the Group x Trial interaction (*OR* = 0.69, *95% CI* [0.63 – 0.77], *p* <.001) suggested that PwMUD had difficulties in learning action–outcome associations throughout the entire task (i.e. reinforcement learning deficits).

To confirm these results, we also examined the Saturated Group-Learn Model which retained the relevant contrast interactions within Group x Phase, and Group x Phase x Trial. While the primary contrast, C1, was significant (*OR* = 1.32, *95% CI* [1.23 – 1.41], *p* <.001), indicating a reduction in accuracy between acquisition-contingency and reversal phases across all participants, it did not interact with Group (*OR* = 0.91, *95% CI* [0.83 – .99], *p* = .032; α = .016 for post-hoc contrasts) or Group x Trial (*OR* = 1.00, *95% CI* [0.92 – 1.10], *p* = .939). Notably, the trending interaction between C1 x Group indicated in the opposite direction to inflexibility amongst PwMUD (indicating either a floor effect due to poor baseline learning, or that PwMUD were indeed more flexible).

Insert Table 4

3.3 Switch/Stay Analysis

Saturated Group Switch and Saturated No-Group Switch models were built and compared, based on a similar rationale to the previous section (see Table 5 for model comparisons and Supplementary Table S1 for statistics of saturated and best-fitting models). The Saturated No-Group Switch Model included Switch/Stay as the output variable; Participant as a random intercept; and the following fixed-effects predictors: Accumulated Feedback, Depression, and

Accumulated Feedback x Depression. The Saturated Group Switch Model included all these predictors plus all Group-relevant predictors/interactions and was again the better fit.

Simplified Group Switch Model 1 removed the three-way interaction (Accumulated Feedback x Group x Depression), retained model fit, and was used for further comparison. However, removal of the Accumulated Feedback x Group interaction did reduce fit (Simplified Group Switch Model 2), indicating a substantial contribution of this interaction and identifying Simplified Group Switch Model 1 as the best-fitting model.

Insert Table 5

In Simplified Group Switch Model 1, Group interacted with both C1 (comparing switch/stay after one reward to one/two/three punishments; *OR* = 0.83, *95% CI* [0.77 – 0.89], *p* <.001) and C2 (comparing switch/stay after one punishment to two/three punishment; *OR* = 0.82, *95% CI* [0.72 – 0.93], *p* = .002). There was no significant Group interaction when comparing two/three cumulative punishments (*OR* = 1.03, *95% CI* [0.73 – 1.45], *p* = .853).

Such results indicated that PwMUD were more likely to switch after a single instance of reward/punishment compared to controls but were no more likely to switch after two or three consecutive punishments. This appeared to again rebuke inflexibility amongst PwMUD and further indicate towards reinforcement learning abnormalities.

When comparing these results with the original Saturated Group Switch Model, we also found similar findings, with significant interactions between C1 x Group (*OR* = 0.81, *95% CI*

[0.74 – 0.87], $p < .001$), C2 x Group (*OR* = 0.77, *95% CI* [0.67 – 0.90], *p* = .001) but not C3 x Group (*OR* = 0.91, *95% CI* [0.61 – 1.35], *p* = .629).

Figure 2 presents overall switch/stay behaviour by group, using predicted values from the Saturated Group Switch Model.

Insert Figure 2

3.4 Traditional Analyses of Reversal Learning

We also compared our original analyses to two "traditional" approaches for reversal learning data. The first compared the number of errors in both acquisition-contingency and reversal phases between PwMUD and controls. We found that PwMUD made significantly more errors in both acquisition-contingency (PwMUD: *M* = 26.23, *SD* = 9.64, Controls: *M* = 18.31, *SD* = 10.41; *t* (65) = 3.23, *p* = .002) and reversal phases (PwMUD: *M* = 32.09, *SD* = 9.74, Controls: *M* = 25.59, *SD* = 11.11; *t* (65) = 2.49, *p* = .013). These results mirror our original findings whereby PwMUD exhibited performance deficits throughout the task.

We also compared our analyses to a Mixed-ANOVA-based approach, previously used in people with Cocaine and Gambling Disorders on the PRLT.[39] Factors were the same as our modelling approach, except that Trial was replaced with Block (a grouping of 8 trials, 5 blocks per phase), and the dependent variable was the number of correct responses per block, ranging from 0 to 8. The results (see Table S2 for comprehensive statistics) again support our original findings, with the effect of Block differing across groups [*F*(2.50, 162.27) = 7.32, $p < .001$; akin to weaker learning in the PwMUD group], but with the effect of Phase not being different across groups [*F*(2.77, 180.31) = 0.15, *p* = .92; which reveals no difference in inflexibility between groups] or any three-way interaction [*F*(9.44, 613.75) = 1.07, *p* = .39].

3.5 Control Analyses

After obtaining these results, we then conducted further analyses to: 1) ensure our findings were not due to factors such as inattention or disengagement from our PwMUD participants; and, 2) investigate the impact of common clinical covariates (Severity of Dependence of methamphetamine and cannabis, and Time Since Last Use of methamphetamine) on PwMUD's performance. We found that: 1) PwMUD and controls had similar attention and motivation during the PRLT (indicated by non-significant differences in overall reaction times and choice-outcome behaviour consistent with learning); 2) the pattern identified in PwMUD (weak learning and increased switching) was exacerbated in more severe users, though time since last use had no significant effects on performance; and, 3) severity of any comorbid cannabis dependency were not impacting accuracy or stay/switch behaviour amongst PwMUD. These analyses are provided in the Supplementary Materials.

4. Discussion

This study aimed to disentangle the contribution of reinforcement learning and inflexibility (as a proxy of compulsivity) to reversal learning performance in PwMUD, compared to drug-naïve controls. We found that PwMUD deficits were due to weaknesses in reinforcement learning, as demonstrated by 1) poorer learning rates across the task; and, 2) a more inconsistent pattern of behaviour after feedback (i.e., greater switching after one instance of reward/punishment). In contrast, our results did not support inflexibility, as PwMUD 1) did not have a greater decline in their accuracy after the reversal of task contingencies; and, 2) did not perseverate after repeated punishments. Together, these findings challenge the prevailing view that MUD is associated with inflexibility.

While these findings conflict with reports of inflexibility in animal models of MUD,[13-16,18,] they do align with previous research in PwMUD. Specifically, our detection of overall weaker learning trajectories is consistent with the only other reversal learning study amongst PwMUD.[20] Furthermore, PwMUD also exhibit poor performance on other decision-making tasks that involve learning action-outcome relationships (i.e., Iowa Gambling Task),[40,41] as well as decreased dopaminergic populations in critical regions for reinforcement learning.[42] As such, we believe the discrepancy between our results and previous preclinical reports may be due to potential methodological oversights regarding the impact of reinforcement learning in non-human studies. This may have occurred because researchers: used reversal errors as the primary measure of inflexibility;[14] trained the acquisition phase before methamphetamine administration;[13,18] or, selected performance thresholds that may not capture “well-learnt” behaviour (i.e. 70% of trials correct).[15,16] Furthermore, due to the relatively small number of PRLT studies in clinical Substance Use Disorder populations, it is difficult to determine whether the deficits we have observed are unique to methamphetamine or generalise across other substances. For example, there is evidence supporting abnormal learning trajectories[32,39] and increased switching behaviour[43] in people with Cocaine and mixed Stimulant Use Disorders. However, support against the presence of inflexibility has been more varied, with mixed results in people with Cocaine Use Disorder,[39,44] and evidence against inflexibility in Amphetamine, and Opioid Use Disorders.[44]

Our identification of increased switching after both reward and punishment also provides clues towards which specific dysfunctional processes may underline PwMUD’s learning deficits, as well as when this may occur in the addiction process. For example, a recent computational analysis of reversal learning data in Stimulant Use Disorder also found greater win-switch and lose-switch behaviour. This was linked to lower reward sensitivity and higher punishment sensitivity, respectively.[43] In comparison, participants with Binge Eating Disorder (who share clinical characteristics with PwMUD) have also been shown to have greater overall switching, though this was instead associated with deficits in updating the value of alternative (non-chosen) options.[45] Furthermore, while animal models may provide support that learning deficits are the result of chronic methamphetamine use,[15,16,19] recent studies have also identified that such difficulties may predate substance use and play a role towards increased methamphetamine self-administration.[46] Although further clinical research is required, our finding of exacerbated learning deficits in PwMUD with more severe patterns of use appears compatible with both scenarios.

From a theoretical standpoint, our work outlines a new perspective of choice behaviour in PwMUD. Previously, behaviour amongst this population has been described as rigid, habitual, or perseverative.[47,48] However, our sample of PwMUD behaved contradictory to such descriptions, acting inconsistently and being overly eager to change responses. As such, it may be that what appears to be "compulsive" behaviour in PwMUD instead reflects difficulties in learning adaptive behaviour. At a therapeutic level, deficits in probabilistic reinforcement learning may explain why treatments such as Contingency Management (CM) are efficacious for PwMUD.[22] While this may seem counter-intuitive, due to CM's reliance on similar learning systems required for PRLT performance, these interventions may overcome PwMUD's deficits via increases in the immediacy/tangibility of reinforcement. This is supported by evidence identifying greater benefits amongst stimulant users when reinforcer magnitude and immediacy are increased.[49,50] Finally, at a psychological level, adopting a view that methamphetamine problems are partly due to an amenable learning difficulty may be more motivating to clients and clinicians, compared to a compulsivity narrative sometimes associated with hopelessness.[24]

Study strengths include the fine-grained analysis that allowed us to differentiate reinforcement learning and cognitive inflexibility. Furthermore, we recruited PwMUD from different treatment and sociodemographic settings, increasing the representativeness of our treatment-seeking sample. Finally, identifying controls with similar sociodemographic characteristics prevented the impact of age, sex, education, and IQ. Regarding limitations, one major consideration is that our task did not provide participants with any tangible positive rewards for accurate performance (i.e., beyond game points). Thus, it is possible that PwMUD may have been less interested in the task, compared to controls. Still, such a concern is mitigated by our control analyses, which identified adequate attention and motivation in the PwMUD group. Readers should also consider that the PRLT is a generalised measure that does not reference substance use. Therefore, while we identified domain-general learning deficits, it may be that PwMUD's inflexibility is restricted to methamphetamine-based contingencies. Relatedly, while these deficits were present in a novel, dynamic task (i.e. including learning and shifting components), it is possible that PwMUD may struggle adapting behaviours learnt prior to chronic methamphetamine consumption (as found in some rodent studies[13,18]). Furthermore, despite the comprehensiveness of our modelling procedure, our sample size may have had insufficient power to detect more subtle, three-way interactions. Finally, we allowed the inclusion of additional mental health diagnoses and secondary alcohol/drug use in our PwMUD group, without the aid of a standardised diagnostic interviews. Although this makes it difficult to ascribe group differences in task performance solely to MUD, such characteristics are also representative of treatment-seeking populations.[8]

5. Conclusion

We found that decision-making problems frequently ascribed to inflexibility in PwMUD were better explained by deficits in reinforcement learning. These findings challenge the "compulsive" stereotype often applied to PwMUD and support the use of treatment approaches targeting contingency-based learning.

Acknowledgements

We sincerely thank all participants, treatment centres, and staff who facilitated this research. Special thanks are particularly extended to those who gave their time during a difficult period in their lives. AVG was funded by an Australian Medical Research Future Fund Next Generation of Clinical Researchers Fellowship (MRF1141214). AHR was funded by an Australian Government Research Training Program Scholarship.

Contributions

AHR, IV and AVG designed the study. AHR and IV collected the data. JCP ran the analyses and provided interpretation for the results. AHR lead the manuscript writing process with all authors contributing to its final version. All authors also reviewed and approved the final version of this manuscript.

Data Availability Statement

The data that supports the findings of this study are available upon request from the corresponding author. It is not publicly available due to privacy and ethical restrictions.

References

1. Darke S, Kaye S, McKetin R, Duflou J. Major physical and psychological harms of methamphetamine use. *Drug Alcohol Rev.* 2008;27(3):253–262. https://doi.org/10.1080/09595230801923702

2. Maxwell JC. A new survey of methamphetamine users in treatment: Who they are, why they like "meth," and why they need additional services. *Subst. Use Misuse* 2014;49(6):639–644. https://doi.org/10.3109/10826084.2013.841244

3. Everitt BJ, Robbins TW. Drug addiction: Updating actions to habits to compulsions ten years on. *Annu. Rev. Psychol.* 2016;67(1):23–50. https://doi.org/10.1146/annurev-psych-122414-033457

4. McKetin R, Kelly E, McLaren J. The relationship between crystalline methamphetamine use and methamphetamine dependence. *Drug Alcohol Depend.* 2006;85(3):198–204. https://doi.org/10.1016/j.drugalcdep.2006.04.007

5. United Nations Office on Drugs and Labor. World Drug Report 2018 - Booklet 1, Executive Summary: Conclusions and Policy Implications. https://www.unodc.org/wdr2018/prelaunch/WDR18_Booklet_1_EXSUM.pdf. Published June 2018. Accessed 22 July, 2019.

6. Stuart AM, Baker AL, Denham AMJ, et al. Psychological treatment for methamphetamine use and associated psychiatric symptom outcomes: A systematic review. *J. Subst. Abuse Treat.* 2020;109:61–79. https://doi.org/10.1016/j.jsat.2019.09.005

7. Brecht M-L, Herbeck D. Time to relapse following treatment for methamphetamine use: A long-term perspective on patterns and predictors. *Drug Alcohol Depend*. 2014;139:18–25. https://doi.org/10.1016/j.drugalcdep.2014.02.702

8. McKetin R, Kothe A, Baker AL, Lee NK, Ross J, Lubman DI. Predicting abstinence from methamphetamine use after residential rehabilitation: Findings from the Methamphetamine Treatment Evaluation Study. *Drug Alcohol Rev.* 2018;37(1):70–78. https://doi.org/10.1111/dar.12528

9. Luigjes J, Lorenzetti V, de Haan S, et al. Defining compulsive behavior. *Neuropsychol. Rev.* 2019;29(1):4–13. https://doi.org/10.1007/s11065-019-09404-9

10. Fineberg NA, Chamberlain SR, Goudriaan AE, et al. New developments in human neurocognition: Clinical, genetic, and brain imaging correlates of impulsivity and compulsivity. *CNS Spectr.* 2014;19(1):69–89. https://doi.org/10.1017/S1092852913000801

11. Insel T, Cuthbert B, Garvey M, et al. Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 2010;167(7):748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

12. Izquierdo A, Jentsch J. Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology* 2012;219(2):607–620. https://doi.org/10.1007/s00213-011-2579-7

13. Cox B, Cope Z, Parsegian A, Floresco S, Aston-Jones G, See R. Chronic methamphetamine self-administration alters cognitive flexibility in male rats. *Psychopharmacology* 2016;233(12):2319–2327. https://doi.org/10.1007/s00213-016-4283-0

14. Diaz MP, Wilson ME, Howell LL. Effects of long-term high-fat food or methamphetamine intake and serotonin 2C receptors on reversal learning in female rhesus macaques. *Neuropsychopharmacology* 2019;44(3):478-486. https://doi.org/10.1038/s41386-018-0200-z

15. Groman SM, Lee B, Seu E, et al. Dysregulation of D2-mediated dopamine transmission in monkeys after chronic escalating methamphetamine exposure. *J. Neurosci.* 2012;32(17): 5843–5852. https://doi.org/10.1523/JNEUROSCI.0029-12.2012

16. Groman SM, Rich KM, Smith NJ, Lee D, Taylor JR. Chronic exposure to methamphetamine disrupts reinforcement-based decision making in rats. *Neuropsychopharmacology* 2018;43(4):770–780. https://doi.org/10.1038/npp.2017.159

17. Kangas B, Bergman J. Effects of self-administered methamphetamine on discrimination learning and reversal in nonhuman primates. *Psychopharmacology* 2016;233(3):373–380. https://doi.org/10.1007/s00213-015-4107-7

18. Kosheleff A, Rodriguez D, O'Dell SJ, Marshall J, Izquierdo A. Comparison of single-dose and extended methamphetamine administration on reversal learning in rats. *Psychopharmacology* 2012;224(3):459–467. https://doi.org/10.1007/s00213-012-2774-1

19. Stolyarova A, O'Dell SJ, Marshall JF, Izquierdo A. Positive and negative feedback learning and associated dopamine and serotonin transporter binding after methamphetamine. *Behav. Brain Res.* 2014;271:195–202. https://doi.org/10.1016/j.bbr.2014.06.031

20. Ghahremani DG, Tabibnia G, Monterosso J, Hellemann G, Poldrack RA, London ED. Effect of modafinil on learning and task-related brain activity in methamphetamine-dependent and healthy individuals. *Neuropsychopharmacology* 2011;36(5):950–959. https://doi.org/10.1038/npp.2010.233

21. Ye T, Pozos H, Phillips TJ, Izquierdo A. Long-term effects of exposure to methamphetamine in adolescent rats. *Drug Alcohol Depend.* 2014;138:17–23. https://doi.org/10.1016/j.drugalcdep.2014.02.021

22. Petry NM, Peirce JM, Stitzer ML, et al. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: A national drug abuse

treatment clinical trials network study. *Arch. Gen. Psychiatry* 2005;62(10):1148–1156. https://doi.org/10.1001/archpsyc.62.10.1148

23. Verdejo-García A, Alcázar-Córcoles MA, Albein-Urios N. Neuropsychological interventions for decision-making in addiction: A systematic review. *Neuropsychol. Rev.* 2018;29(1):79-92. https://doi.org/10.1007/s11065-018-9384-6

24. Heather N. Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addict. Behav. Rep.* 2017;6:15–38. https://doi.org/10.1016/j.abrep.2017.05.002

25. Gossop M, Darke S, Griffiths P, et al. The Severity of Dependence Scale (SDS): Psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. *Addiction* 1995;90(5):607–614. https://doi.org/10.1046/j.1360-0443.1995.9056072.x

26. Topp L, Mattick RP. Choosing a cut-off on the Severity of Dependence Scale (SDS) for amphetamine users. *Addiction* 1996;92(7):839–845. https://doi.org/10.1111/j.1360-0443.1997.tb02953.x

27. Nelson HE. *National Adult Reading Test (NART): For the Assessment of Premorbid Intelligence in Patients with Dementia: Test manual.* Windsor: NFER-Nelson; 1982.

28. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1977;1(3):385–401. https://doi.org/10.1177/014662167700100306

29. Sobell LC, Sobell MB. *Timeline Followback User's Guide: A Calendar Method for Assessing Alcohol and Drug Use*. Windsor: Addict Research Foundation; 1996.

30. Swainson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW. Probabilistic learning and reversal deficits in patients with parkinson's disease or frontal or temporal

lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia* 2000;38(5):596–612. https://doi.org/10.1016/S0028-3932(99)00103-7

31. Verdejo-García A, del Mar Sánchez-Fernández M, Alonso-Maroto LM, et al. Impulsivity and executive functions in polysubstance-using rave attenders. *Psychopharmacology* 2010;210(3):377–392. https://doi.org/10.1007/s00213-010-1833-8
32. Fernández-Serrano MJ, Perales JC, Moreno-López L, Pérez-García M, Verdejo-García A. Neuropsychological profiling of impulsivity and compulsivity in cocaine dependent individuals. *Psychopharmacology* 2012;219(2):673-83.
33. Rouder JN, Morey RD, Verhagen J, Swagman AR, Wagenmakers E-J. Bayesian analysis of factorial designs. *Psychol. Methods* 2017;22(2):304–321. https://doi.org/10.1037/met0000057
34. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01
35. R Core Team. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing; 2017.
36. *JASP* [computer program]. Version 0.11.1. Amsterdam: JASP Team, 2019
37. Bronfman ZZ, Ginsburg S, Jablonka E. Shaping the learning curve: Epigenetic dynamics in neural plasticity. *Front. Integr. Neurosci.* 2014;8:Article 55. https://doi.org/10.3389/fnint.2014.00055
38. Davis M. Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *J. Data. Sci.* 2010;8(1):61-73. http://doi.org/10.6339%2fJDS.2010.08(1).563
39. Torres A, Catena A, Cándido A, Maldonado A, Megías A, Perales JC. Cocaine dependent individuals and gamblers present different associative learning anomalies in feedback-

driven decision making: A behavioral and ERP study. *Front. Psychol.* 2013;4:Article 122. https://doi.org/10.3389/fpsyg.2013.00122

40. van der Plas EAA, Crone EA, van den Wildenberg WPM, Tranel D, Bechara A. Executive control deficits in substance-dependent individuals: A comparison of alcohol, cocaine, and methamphetamine and of men and women. J. Clin. Exp. *Neuropsychol.* 2009;31(6):706–719. https://doi.org/10.1080/13803390802484797
41. Wang G, Shi J, Chen N, et al. Effects of length of abstinence on decision-making and craving in methamphetamine abusers. *PLOS ONE* 2013;8(7):e68791. https://doi.org/10.1371/journal.pone.0068791
42. Volkow ND, Chang L, Wang G-J, et al. Low level of brain dopamine D2 receptors in methamphetamine abusers: Association with metabolism in the orbitofrontal cortex. *Am. J. Psychiatry* 2001;158(12):2015–2021. https://doi.org/10.1176/appi.ajp.158.12.2015
43. Kanen JW, Ersche KD, Fineberg NA, Robbins TW, Cardinal RN. 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* 2019;236(8):2337–2358. https://doi.org/10.1007/s00213-019-05325-w
44. Ersche KD, Roiser JP, Robbins TW, Sahakian BJ. Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. *Psychopharmacology* 2008;197(3):421–431. https://doi.org/10.1007/s00213-007-1051-1
45. Reiter AMF, Heinze H-J, Schlagenhauf F, Deserno L. Impaired flexible reward-based decision-making in binge eating disorder: Evidence from computational modeling and functional neuroimaging. *Neuropsychopharmacology* 2017;42(3):628–637. https://doi.org/10.1038/npp.2016.95

46. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR. Model-free and model-based influences in addiction-related behaviors. *Biol. Psychiatry*. 2019;85(11):936-945. doi:10.1016/j.biopsych.2018.12.017

47. Paulus MP, Hozack N, Frank L, Brown GG, Schuckit MA. Decision making by methamphetamine-dependent subjects is associated with error-rate-independent decrease in prefrontal and parietal activation. *Biol. Psychiatry* 2003;53(1):65–74. https://doi.org/10.1016/S0006-3223(02)01442-7

48. Voon V, Derbyshire K, Rück C, et al. Disorders of compulsivity: A common bias towards learning habits. *Mol. Psychiatry* 2015;20(3):345–352. https://doi.org/10.1038/mp.2014.44

49. Silverman K, Chutuape MA, Bigelow GE, Stitzer ML. Voucher-based reinforcement of cocaine abstinence in treatment-resistant methadone patients: effects of reinforcement magnitude. *Psychopharmacology* 1999;146(2):128–138. https://doi.org/10.1007/s002130051098

50. Packer RR, Howell DN, McPherson S, Roll JM. Investigating reinforcer magnitude and reinforcer delay: A contingency management analog study. *Exp. Clin. Psychopharmacol.* 2012;20(4):287–292. https://doi.org/10.1037/a0027802

Tables and Figure Legends

Table 1. Descriptive statistics of sociodemographic and methamphetamine use characteristics in PwMUD and drug-naïve controls.

Demographics	PwMUD	Controls	Test Statistic	Bayes Factor
Sex (F/M)	11/24	12/20	$\chi^2 = .027, p = .60$	
Age	33.26 (7.78)	31.44 (9.54)	$U = 460.5, p = .21$	$BF_{10} = 0.44$
Years of Education	14.06 (2.15)	15.00 (2.11)	$U = 699, p = .078$	$BF_{10} = 1.75$
Verbal IQ	109.29 (5.46)	108.78 (6.29)	$U = 562.5, p = .98$	$BF_{10} = 0.27$
Depression (CES-D)	19.83 (11.79)	9.22 (6.66)	$U = 220, p < .001$	$BF_{10} = 615.90$

Meth. Use (PwMUD only)	*M* (or *N*)	*SD* (or %)	Range
Severity of Dependence (SDS)	9.97	3.05	[4 – 14]
Daily dose (grams)	0.44	0.29	[0.10 – 1.50]
Frequency (days/month)	19.54	10.34	[3 – 31]
Duration (years)	8.56	5.34	[0.58 – 22]
Last Use (days)	37.48	44.81	[3 – 180]
Treatment Type			
Inpatient Rehab	23	65.71%	-
Outpatient Counselling	6	17.14%	-
Multiple	4	11.43%	-
No treatment	2	5.71%	-
Route of Admin.			
Smoking	32	91.4%	-

Injecting	3	8.6%	-

Note: Two PwMUD also reported HIV+ status. CES-D: Centre for Epidemiologic Studies Depression Scale. SDS: Severity of Dependence Scale. SDS scores can range between 0 and 15; with those above 4 indicating likely MUD[26].

Table 2. Additional substance and medication use amongst PwMUD.

	N (or M)	% (or *SD*)
Other Illicit Subs.Use		
Cannabis	10	28.57%
GHB	8	22.86%
MDMA	6	17.14%
Cocaine	4	11.43%
Heroin	1	2.86%
SDS Alcohol	1	2.70
SDS Cannabis	1.6	3.20
Prescribed Medication		
Anti-Dep	11	31.43%
Anti-Psychotic	4	11.43%
Anti-Convulsive (lamotrigine)	1	2.86%
Benzodiazepine (diazepam)	1	2.86%

Note: Other Illicit Substance Use refers to substances taken more than 10 times in the past 12 months. SDS: Severity of Dependence Scale, scores can range between 0 and 15. Anti-Dep includes escitalopram, sertraline, fluoxetine, mirtazapine, venlafaxine, duloxetine. Anti-Psychotic includes aripiprazole and quetiapine.

Table 3. Fitting indices for models analysing trial-by-trial performance across and within phases (reversal learning inflexibility).

Model	*df*	*AIC*	*BIC*	$\chi 2$	*p*
Saturated No-Group-Learn	11	12276	12356		
Saturated Group-Learn	19	12214	12352	77.689	<.001 (Group > No-Group)
Simplified Group-Learn 1	16	12214	12343	5.818	.121 (1 ≥ Saturated)
Simplified Group-Learn 2.1	15	12266	12376	54.622	< .001 (2.1 < 1)
*Simplified Group-Learn 2.2**	13	12214	12330	6.494	.090 (2.2 ≥ 1)
Simplified Group-Learn 2.3	13	12262	12357	54.409	< .001 (2.3 < 1)

Note: p-values correspond to contrasts regarding the superiority of the more complex model relative to the simpler one. See text for a description of model compositions. * best-fitting model.

Table 4. Comprehensive statistics for the saturated and best-fitting model (trial-by-trial performance).

	Saturated model			**Best-fitting model**		
Predictors	***Odds Ratios***	***CI***	***p***	***Odds Ratios***	***CI***	***p***
Intercept	3.31	2.67 – 4.10	**<.001**	3.27	2.65 – 4.05	**<.001**
Phase						
C1	1.32	1.23 – 1.41	**<.001**	1.24	1.19 – 1.30	**<.001**
C2	1.31	1.22 – 1.41	**<.001**	1.27	1.21 – 1.33	**<.001**
C3	1.20	1.12 – 1.29	**<.001**	1.20	1.15 – 1.25	**<.001**
Trial	1.96	1.82 – 2.10	**<.001**	1.94	1.81 – 2.08	**<.001**
Depression	1.08	0.93 – 1.26	.311	1.08	0.93 – 1.26	.308
Group	0.54	0.39 – 0.73	**<.001**	0.54	0.40 – 0.74	**<.001**
Phase x Trial						
C1 x Trial	0.93	0.87 – 1.00	.049	0.93	0.89 – 0.98	**.002**
C2 x Trial	1.06	0.99 – 1.14	.075	1.06	1.02 – 1.11	**.005**
C3 x Trial	0.82	0.77 – 0.88	**<.001**	0.87	0.83 – 0.91	**<.001**
Trial x Depression	0.98	0.93 – 1.03	.412	0.98	0.93 – 1.03	.408
Group x Phase						
C1 x Group	0.91	0.83 – 0.99	.032			
C2 x Group	0.95	0.86 – 1.03	.225			
C3 x Group	0.99	0.91 – 1.09	.858			
Group x Trial	0.69	0.62 – 0.76	**<.001**	0.69	0.63 – 0.77	**<.001**
Group x Phase x Trial						
C1 x Group x Trial	1.00	0.92 – 1.10	.939			
C2 x Group x Trial	1.01	0.92 – 1.10	.886			
C3 x Group x Trial	1.11	1.02 – 1.22	**.016**			
Random Effects						
σ^2	3.29			3.29		
$\tau_{participant}$	0.29			0.28		
ICC	0.08			0.08		

Note: Bolded *p*-values are viewed as significant (≤.05 in typical analyses, ≤.0167 in contrasts). C1 (1, -1, 1, -1) compares acquisition-contingency to reversal phases; C2 (1, 1, -1, -1) compares easy and hard phases; C3 (1, -1, -1, 1) is theoretically irrelevant, necessary to complete contrasts and compares phases 1 and 4 with phases 2 and 3.

Table 5. Fitting indices for models involved in analyses of sensitivity to Accumulated Feedback (switch/stay analysis).

Model	***df***	***AIC***	***BIC***	***$\chi 2$***	***p***
Saturated No-Group Switch	9	11328	11394		
Saturated Group Switch	17	11301	11424	43.431	< .001 (Group >No-Group)
*Simplified Group Switch 1**	14	11297	11399	2.2828	.516 (1 > Saturated)
Simplified Group Switch 2	11	11322	11402	30.814	< .001 (1 > 2)

Note: p-values correspond to contrasts regarding the superiority of the more complex model relative to the simpler one. See text for a description of model compositions. * best-fitting model.

Figure 1. Observed percentage of correct responses as a function of Phase, Trial, and Group.

Figure 1 Legend: The dots represent the observed percentage of correct choices at each trial per group (statistically, a function of Phase, Trial and Group). The lines represent logarithmic trendlines maximizing the fitting for each Phase and Group.

Figure 2. Predicted percentage of changed response (and confidence intervals) in the current trial as a function of Accumulated Feedback.

Figure 2 Legend: This figure visualises the predicted percentage of each groups' switch/stay behaviour for each level of the Accumulated Feedback predictor, in the saturated model (see Table S1). This is achieved by tuning the saturated model's parameters using a maximum likelihood approach in order to best approximate participants observed choices. Results are similar across visualisations in observed, best-fitting, and saturated versions of this figure.

FIGURE 1

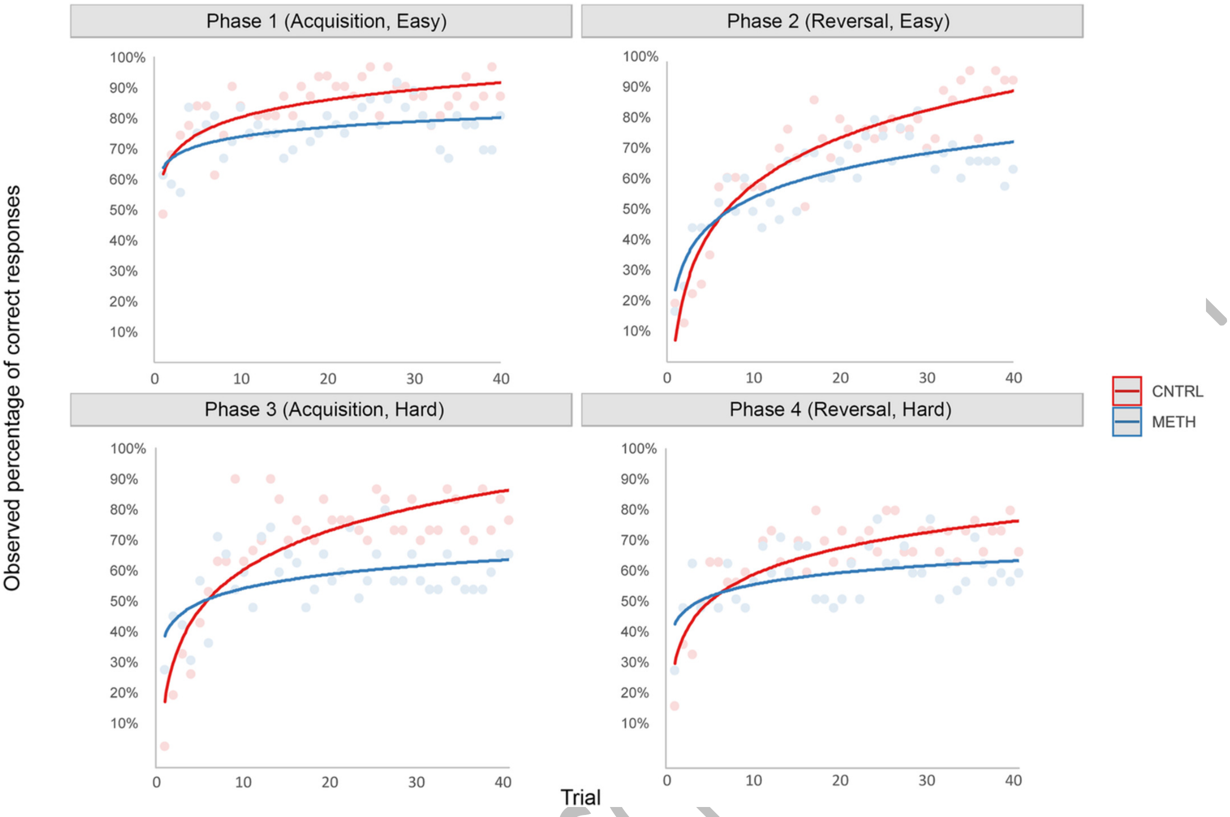
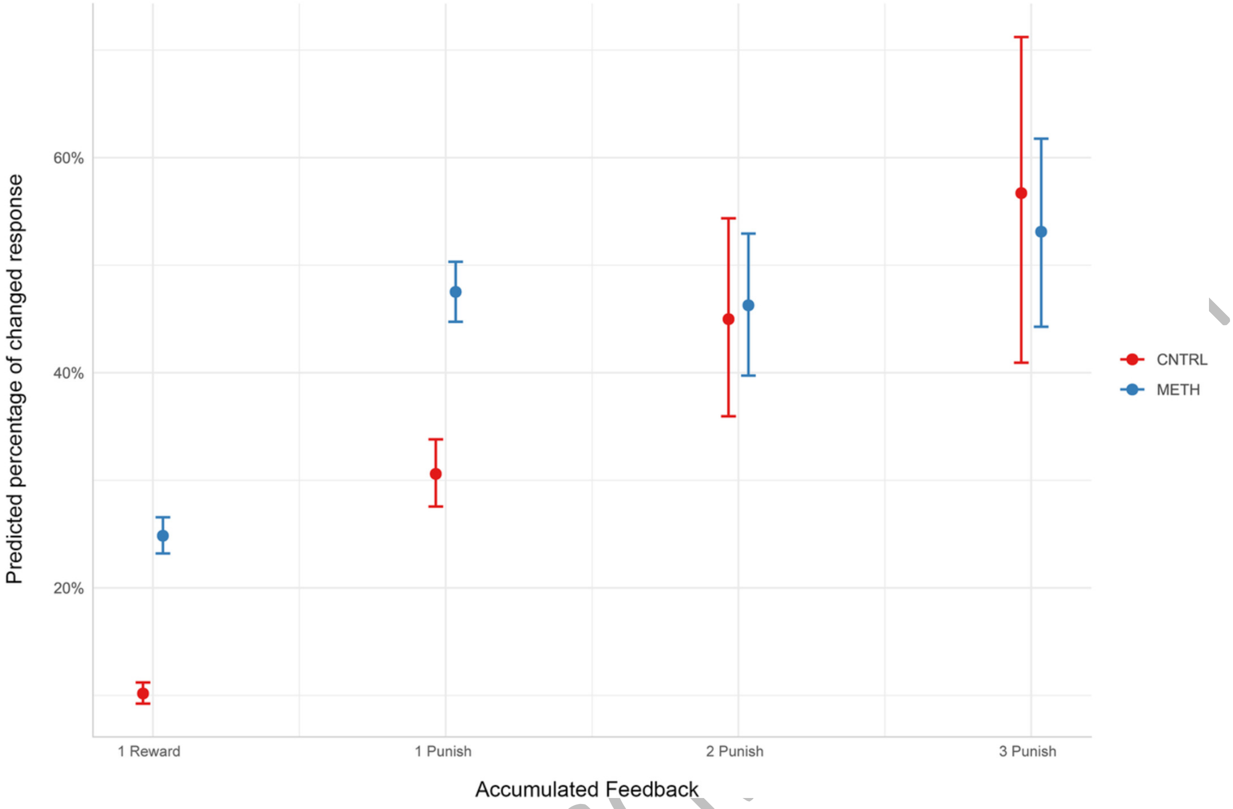

Phase 1 (Acquisition, Easy)
Phase 2 (Reversal, Easy)
Phase 3 (Acquisition, Hard)
Phase 4 (Reversal, Hard)
Observed percentage of correct responses
Trial
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
10
20
30
40
CNTRL
METH

FIGURE 2


Predicted percentage of changed response
60%
40%
20%
1 Reward
1 Punish
2 Punish
3 Punish
Accumulated Feedback
CNTRL
METH

Supplementary Materials for

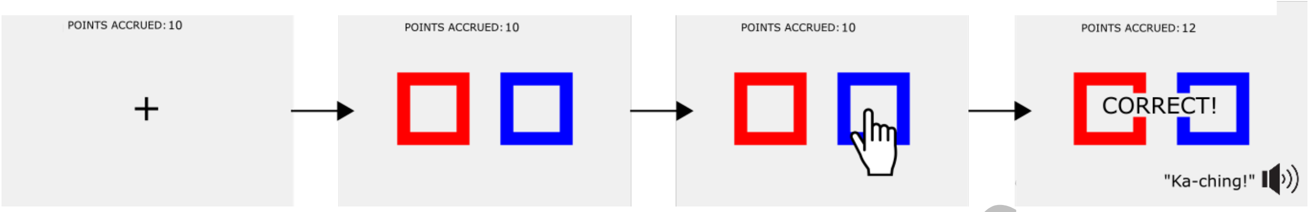
ARE METHAMPHETAMINE USERS COMPULSIVE? FAULTY REINFORCEMENT LEARNING, NOT INFLEXIBILITY, UNDERLIES DECISION-MAKING IN PEOPLE WITH METHAMPHETAMINE USE DISORDER

This document includes two sections. The first, *Supplementary Methods*, includes a visual diagram of the Probabilistic Reversal Learning Task (PRLT; Figure S1) and justification of the logarithmic transformation of Trial during our modelling process (Section S1.1). The second, *Supplementary Results*, includes comprehensive statistics for the Switch/Stay models in the manuscript (Table S1), comprehensive statistics for the traditional ANOVA (Table S2), and the Control Analyses. The Control Analyses include an investigation of inattention/motivation concerns amongst PwMUD (S2.1 and S2.2), and an investigation into potential clinical covariates of PRLT performance (severity of meth. dependency, time since last meth. use, and severity of cannabis dependency, S2.3, Tables S3 – S8, Figures S2 – S3).

S1. Supplementary Methods

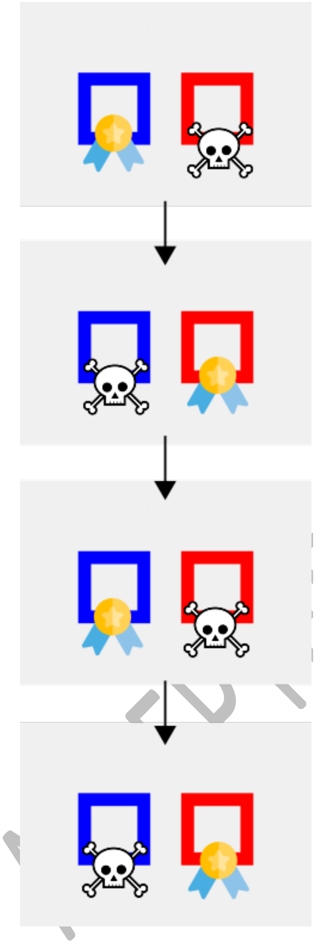
Figure S1. Event and Phase sequence of the Probabilistic Reversal Learning Task

1. Event sequence during individual trials of the *Probabilistic Reversal Learning Task*



1. Warning Cue *2. Options Presented* *3. Choice Selected* *4. Feedback*

2. Contingencies for each pase (40 trials) of the *Probabilistic Reversal Learning Task*



Acquisition-Contingency (Easy, Phase 1)
Blue Square: **"Correct"**, **80%** of choices are rewarded, **20%** are punished.
Red Square: **"Incorrect"**, **20%** of choices are rewarded, **80%** are punished.

Reversal (Easy, Phase 2)
Blue Square: **"Incorrect"**, **20%** of choices are rewarded, **80%** are punished.
Red Square: **"Correct"**, **80%** of choices are rewarded, **20%** are punished.

Acquisition-Contingency (Hard, Phase 3)
Blue Square: **"Correct"**, **70%** of choices are rewarded, **30%** are punished.
Red Square: **"Incorrect"**, **30%** of choices are rewarded, 7**0%** are punished.

Reversal (Hard, Phase 4)
Blue Square: **"Incorrect"**, **30%** of choices are rewarded, **70%** are punished.
Red Square: **"Correct"**, **70%** of choices are rewarded, **30%** are punished.

Note: Figure based on Verdejo-Garcia et al.[1] In the event sequence, coloured squares are randomly positioned in terms of horizontal location (left or right) for each trial. In the phase sequence, the medallion represents the "correct" square and the skull represents the "incorrect" square. In original task, square colours were red and green, with all participants denying the presence of colour-blindness.

S1.1 Evidence supporting logtrial transformation

We hypothesised that by logarithmically transforming Trial, we would be able to pragmatically model the 'negatively accelerating curve'. To confirm this, we compared the model described in the main text against: (a) a purely linear version of the model (with non-transformed Trial as predictor), and (b) a polynomial model, in which the effect of Trial was decomposed into a linear and a quadratic component. All models thus included Phase and Group as predictors, along with the corresponding version of Trial, and all possible interactions between predictors. As shown below, Logtrial was the best fit.

Linear: response ~ phase * trial * group + (1|participant) --> 17 degrees of freedom

AIC/BIC: 12383.35 / 12507.11

Logarithmic: response ~ phase * log-trial * group + (1|participant) --> 17 degrees of freedom

AIC/BIC: 12211.71/12335.4

Polynomial: response ~ phase * (trial + trial^2) * group + (1|participant) --> 25 degrees of freedom

AIC/BIC: 12241.20 / 12423.20

S2. Supplementary Results

Table S1. Switch/stay model results: Statistics for the saturated and best-fitting model.

	Saturated model			**Best fitting model**		
Predictors	*Odds Ratios*	*CI*	*p*	*Odds Ratios*	*CI*	*p*
Intercept	0.48	0.32 – 0.73	**<.001**	0.44	0.30 – 0.66	**<.001**
Accumulated feedback						
C1	1.62	1.51 – 1.74	**<.001**	1.57	1.49 – 1.66	**<.001**
C2	1.33	1.17 – 1.51	**<.001**	1.25	1.14 – 1.37	**<.001**
C3	1.27	0.90 – 1.79	.182	1.10	0.85 – 1.41	.473
Group	1.53	0.91 – 2.56	.107	1.64	0.99 – 2.73	.056
Depression	1.03	0.61 – 1.74	.921	0.89	0.55 – 1.46	.657
Acc. Feedback x Group						
C1 x Group	0.81	0.74 – 0.87	**<.001**	0.83	0.77 – 0.89	**<.001**
C2 x Group	0.77	0.67 – 0.90	**.001**	0.82	0.72 – 0.93	**.002**
C3 x Group	0.91	0.61 – 1.35	.629	1.03	0.73 – 1.45	.853
Acc. Feedback x Depression						
C1 x Depression	1.08	0.99 – 1.18	.095	1.03	0.99 – 1.07	.124
C2 x Depression	1.12	0.96 – 1.32	.153	1.01	0.94 – 1.07	.832
C3 x Depression	1.31	0.84 – 2.06	.234	1.03	0.86 – 1.23	.740
Group x Depression	0.94	0.52 – 1.70	.844	1.11	0.64 – 1.92	.716
Acc. Feedback x Group x Depression						
C1 x Group x Depression	0.95	0.86 – 1.04	.255			
C2 x Group x Depression	0.88	0.74 – 1.05	.147			
C3 x Group x Depression	0.76	0.46 – 1.23	.262			
Random Effects						
σ^2	3.29			3.29		
$\tau_{participant}$	0.66			0.66		
ICC	0.17			0.17		

Note: Bolded *p*-values are viewed as significant (<.05 in typical analyses, <.0167 in contrast). C1 (-3. 1. 1. 1) compares behaviour after one instance of reward to one/two/three instances of punishment; C2 (0, -2, 1, 1) compares behaviour after one instance of punishment and two/three instances of punishment; C3 (0, 0, -1, 1) compares behaviour after two instances of punishment with three instances of punishment.

Table S2. Traditional mixed ANOVA approach predicting Accuracy.

Within subjects' effects

Predictors	***df***	***F***	***p***
Phase	2.77	24.114	<.001
Group x Phase	2.77	0.15	.92
Residual	180.31		
Block	2.50	42.29	<.001
Block x Group	2.50	7.32	<.001
Residual	162.27		
Block x Phase	9.44	5.01	<.001
Block x Group x Phase	9.44	1.07	.39
Residual	613.75		

Between subjects' effects

Predictors	***df***	***F***	***p***
Group	1	13.09	<.001
Residual	65		

S2.1 Testing PwMUD's inattention on PRLT

We hypothesised that if PwMUD were inattentive on the PRLT they would either show significantly faster or slower responses, compared to controls. In contrast, if attention was intact, response times would be similar. We collected each participants' median response time across the task, using a recommended approach.[2] After this, a Mann-Whitney *U* test was applied and found no significant difference between groups' responses (PwMUD = 995ms, Controls = 908.5ms; *U* =444.5, *p* = .15). This indicated similar task attention between groups.

S2.2 Testing PwMUD's motivation on PRLT

To exclude lack of motivation/engagement, we examined if the performance of PwMUD was better explained by a 'learning' model relative to a 'null' model. The 'learning' model used Trial and Phase to predict choices, and the 'null' model included only the intercept. Results showed that the 'learning' model yielded a significantly better fit (*AIC* = 6998.2, *BIC* = 7057.9) than the null model (*AIC* = 7228.8, *BIC* = 7242.1; *p* <.001). Thus, PwMUD were indeed using information from their previous learning to guide their choices.

S2.3. Clinical covariate analyses in PwMUD

S2.3.1. Methamphetamine dependency and Time Since Last Use

Here, we evaluated whether severity of methamphetamine dependence (measured by the Severity of Dependence Scale, SDS Meth) or Time Since Last Use impacted performance. This began by conducting preliminary 2 x 4 Mixed ANCOVAs (see Tables S3 and S4) to identify whether SDS or Time Since Last Use explained performance differences among participants in the PwMUD group on a relatively simple model of accuracy. Here, SDS Meth was associated with overall worse performance amongst PwMUD [*F*(1,33) = 4.22, *p* = .048], while Time Since Last Use was not [*F*(1, 29) = 0.01, *p* = .91]. As a result, Time Since Last Use was discontinued from further analyses.

Following this, we entered SDS Meth as a covariate into our more complex trial-by-trial performance and stay/switch model building (with analyses restricted solely to PwMUD; comprehensive tables of best-fitting models included below). Here, the best-fitting trial-by-trial model (also the saturated model, statistics presented in Table S5, visually presented in Figure S2) identified poorer learning in more severe methamphetamine users, where those with higher SDS scores showed flatter/weaker learning curves, relative to those with lower SDS scores (*OR* = 0.91, *p* = .002). Relatedly, those with lower SDS Meth scores actually showed a greater reversal cost (i.e., support towards greater inflexibility) compared to those with higher SDS Meth scores (C1 x SDS Meth: *OR* = 0.89, *p* < .001; C1 x SDS Meth x Trial: *OR* = .93, *p* = .010). We hypothesize that this is due to a person's severity of dependence impacting their ability to learn the initial acquisition contingencies. In this manner, more severe users struggle to learn the initial contingencies (and are thus unaffected when they are reversed), while less severe users can partly learn contingencies (and are thus more affected when these are reversed).

In the switch/stay models, the best-fitting model (statistics presented in Table S6) found that PwMUD with higher SDS were more prone to switch behaviour (*OR* = 1.32, *p* = .013). Nonetheless, the SDS x Accumulated Feedback interactions did not survive model comparisons, nor were they significant in the saturated model (lowest *p* = .12), indicating that the increased switching from SDS Meth was not significantly interacting with specific forms of feedback (i.e., one reward, one/two/three punishment). Interestingly, however, is that despite the absence of these contrast interactions, a visual representation of the saturated model's predicted responses (Figure S3) showed a trend whereby more severe PwMUD were predicted to exhibit greater switching after one instance of reward/punishment.

S2.3.2. Cannabis dependency

We also investigated whether level of any comorbid cannabis dependency (SDS Cannabis) was impacting our PwMUD group's performance. Results were again obtained from both a best-fitting model and a saturated model. Descriptive statistics for SDS Cannabis are available in Table 2.

Regarding accuracy across and within phases (see Table S7), the interaction investigating changes between acquisition and reversal phases (SDS Cannabis x C1) did not survive into the best-fitting model. Furthermore, in the saturated model, SDS Cannabis did not have a significant effect on trial-by-trial learning (SDS Cannabis x Trial: *p* = .68), did not affect adaption to reversal (i.e., no evidence of inflexibility, C1 x SDS Cannabis: *p* = .62), and was not involved in any three-way interactions (C1 x SDS Cannabis x Trial: *p* = .38).

Similarly, in the switch/stay analyses (see Table S8), the interactions between SDS Cannabis and Accumulated Negative Feedback did not survive into the best-fitting model. Furthermore, in the saturated model, SDS Cannabis did not significantly predict overall switch/stay behaviour (SDS Cannabis: *p* = .30), nor were there any significant differences in responses after reward or punishments (C1 x SDS Cannabis: *p* = .08, C2 x SDS Cannabis: *p* = .65; C3 x SDS Cannabis: *p* = .65).

Taken together, these results indicate that any comorbid dependency on cannabis was not providing a significant impact towards the behaviour reported in the manuscript.

Table S3. Mixed ANCOVA predicting Accuracy including SDS Meth as covariate

Within subjects' effects of PwMUD			
Predictors	***df***	***F***	***p***
Phase	2.82	4.22	.009
Phase x SDS Meth	2.82	1.27	.29
Residual	93.02		
Block	2.43	2.60	.070
Block x SDS Meth	2.43	0.63	.57
Residual	80.03		
Block x Phase	8.25	0.80	.60
Block x Phase x SDS Meth	8.25	0.92	.51
Residual	272.15		
Between subjects' effects			
Predictors	***df***	***F***	***p***
SDS Meth	1	4.22	.048
Residual	33		

Table S4. Mixed ANCOVA predicting Accuracy including Time Since Last Use as covariate

Within subjects' effects

Predictors	***df***	***F***	***p***
Phase	2.75	7.97	<.001
Phase x Last Use	2.75	1.59	.20
Residual	79.74		
Block	2.23	5.91	.003
Block x Last Use	2.23	0.31	.76
Residual	64.63		
Block x Phase	8.29	1.72	.092
Block x Phase x Last Use	8.29	0.97	.46
Residual	240.27		

Between subjects' effects

Predictors	***df***	***F***	***p***
Last Use	1	0.01	.91
Residual	29		

Table S5. Statistics for the best-fitting model (SDS Meth, trial-by-trial performance)

Predictors	***Odds Ratios***	***CI***	***p***
Intercept	1.85	1.60 – 2.14	**<.001**
Phase			
C1	1.21	1.14 – 1.28	**<.001**
C2	1.25	1.18 – 1.32	**<.001**
C3	1.20	1.14 – 1.27	**<.001**
Trial	1.34	1.27 – 1.42	**<.001**
Phase x Trial			
C1 x Trial	0.95	0.89 – 1.00	.054
C2 x Trial	1.07	1.01 – 1.13	**.016**
C3 x Trial	0.92	0.87 – 0.97	**.002**
SDS Meth	0.84	0.72 – 0.97	**.017**
Phase x SDS Meth			
C1 x SDS Meth	0.89	0.84 – 0.95	**<.001**
C2 x SDS Meth	0.94	0.88 – 1.00	.038
C3 x SDS Meth	0.95	0.90 – 1.01	.092
SDS Meth x Trial	0.91	0.86 – 0.97	**.002**
Phase x SDS Meth x Trial			
C1 x SDS Meth x Trial	0.93	0.87 – 0.98	**.010**
C2 x SDS Meth x Trial	1.01	0.96 – 1.07	.635
C3 x SDS Meth x Trial	1.05	0.99 – 1.11	.109
σ^2	3.29		
τ_{00}	0.16 $_{\text{code}}$		
ICC	0.05		
AIC	6972.286		

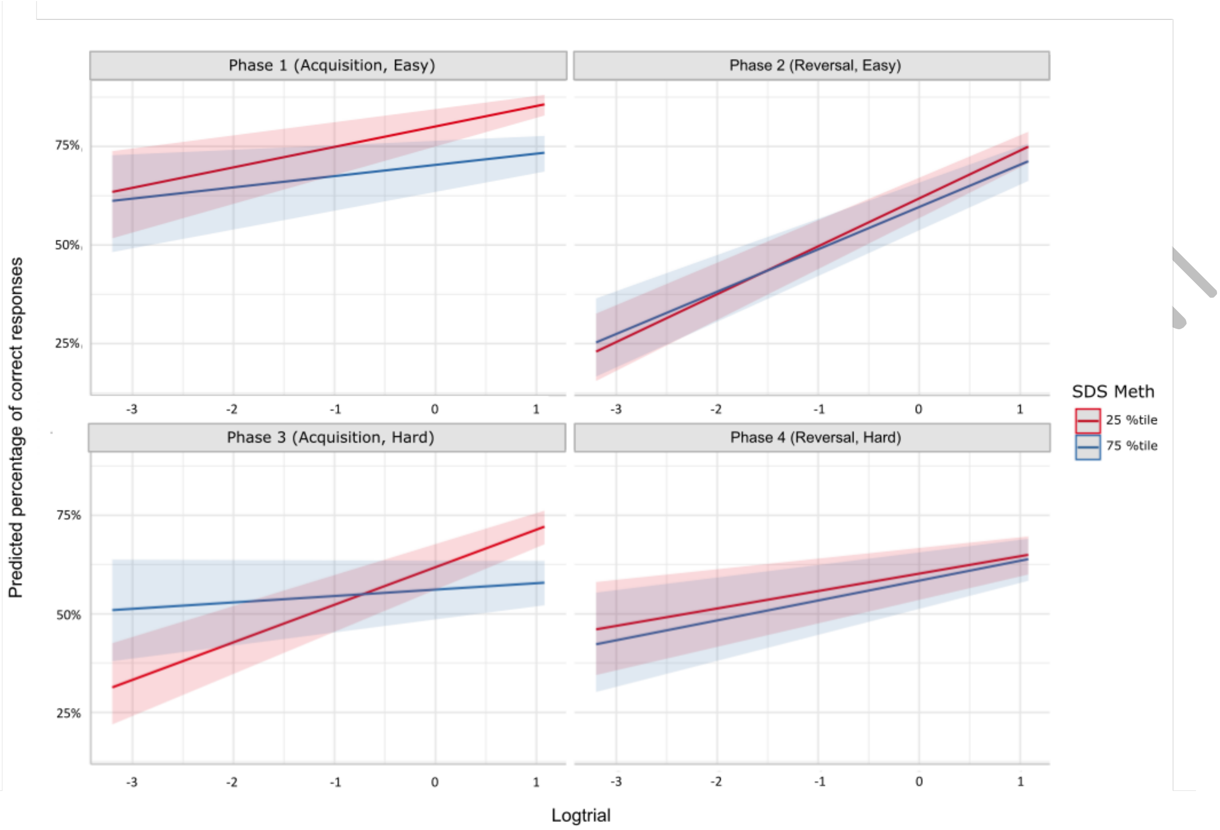
Note: Bolded p-values are viewed as significant (≤.05 in typical analyses, ≤.0167 in contrasts). C1 (1, -1, 1, -1) compares acquisition-contingency to reversal phases; C2 (1, 1, -1, -1) compares easy and hard phases; C3 (1, -1, -1, 1) is theoretically irrelevant, necessary to complete contrasts and compares phases 1 and 4 with phases 2 and 3. Trial remains log-transformed as per original analyses.

Table S6. Statistics for the best-fitting and saturated models (SDS Meth, switch/stay)

	Saturated model			**Best fitting model**		
Predictors	*Odds Ratios*	*CI*	*p*	*Odds Ratios*	*CI*	*p*
Intercept	0.73	0.57 – 0.93	**.011**	0.73	0.57 – 0.93	**.010**
Accumulated feedback						
C1	1.31	1.26 – 1.37	**<.001**	1.31	1.26 – 1.37	**<.001**
C2	1.03	0.95 – 1.11	.501	1.02	0.95 – 1.10	.521
C3	1.10	0.90 – 1.35	.352	1.11	0.90 – 1.35	.326
SDS Meth	1.25	0.98 – 1.58	.069	1.32	1.06 – 1.65	**.013**
C1 x SDS Meth	0.97	0.93 – 1.01	.116			
C2 x SDS Meth	0.96	0.89 – 1.03	.242			
C3 x SDS Meth	1.05	0.86 – 1.28	.648			
σ^2		3.29			3.29	
$\tau_{participant}$		0.66			0.66	
ICC		0.17			0.17	
AIC		5815.339			5812.705	

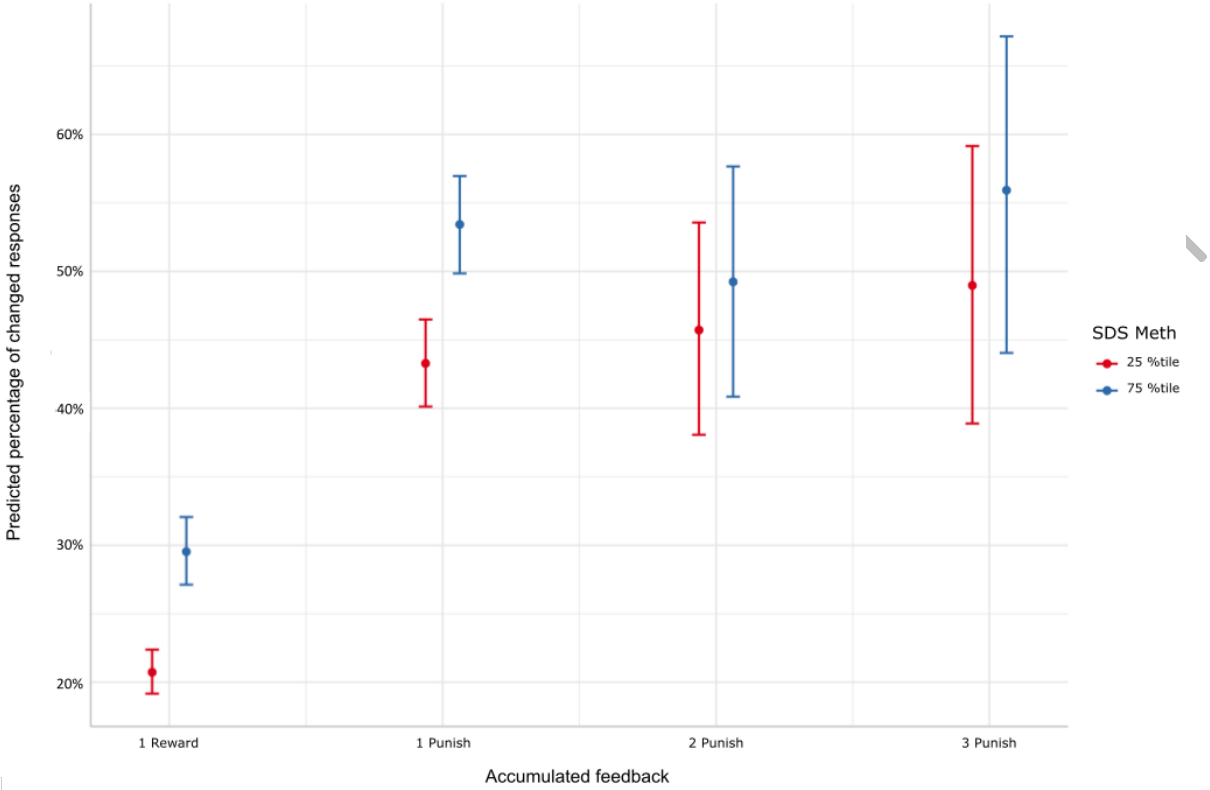
Note: Bolded *p*-values are viewed as significant (<.05 in typical analyses, <.0167 in contrast). C1 (-3. 1. 1. 1) compares behaviour after one instance of reward to one/two/three instances of punishment; C2 (0, -2, 1, 1) compares behaviour after one instance of punishment and two/three instances of punishment; C3 (0, 0, -1, 1) compares behaviour after two instances of punishment with three instances of punishment.

Figure S2. Predicted percentage of correct responses amongst PwMUD with high (75th percentile, blue) and low (25th percentile, red) SDS Meth scores.



Note: This figure visualises the predicted percentage of correct responses for given values of SDS Meth at across each phase, in the saturated model (see Table S5). This is achieved by tuning the saturated model's parameters using a maximum likelihood approach in order to best approximate participants' observed data.

Figure S3. Predicted percentage of changed responses amongst PwMUD with high (75th percentile, blue) and low (25th percentile, red) SDS Meth scores.



Note: This figure visualises the predicted percentage of switch/stay behaviour for given values of SDS Meth at each level of the Accumulated Feedback predictor, in the saturated model (see Table S6). This is achieved by tuning the saturated model's parameters using a maximum likelihood approach in order to best approximate participants' observed choices.

Table S7. Statistics for best-fitting and saturated model (SDS Cannabis, trial-by-trial performance)

	Saturated model			**Best-fitting model**		
Predictors	*Odds Ratios*	*CI*	*p*	*Odds Ratios*	*CI*	*p*
Intercept	1.83	1.57 – 2.13	**<0.001**	1.83	1.57 – 2.13	**<0.001**
Phase C1	1.19	1.13 – 1.26	**<0.001**	1.19	1.13 – 1.26	**<0.001**
Phase C2	1.24	1.17 – 1.31	**<0.001**	1.24	1.17 – 1.31	**<0.001**
Phase C3	1.19	1.13 – 1.26	**<0.001**	1.19	1.13 – 1.26	**<0.001**
Trial	1.33	1.26 – 1.41	**<0.001**	1.33	1.26 – 1.41	**<0.001**
SDS Cannabis	0.98	0.85 – 1.14	0.823	0.98	0.84 – 1.14	0.799
Trial x C1	0.94	0.89 – 0.99	0.028	0.94	0.89 – 0.99	0.027
Trial x C2	1.07	1.01 – 1.13	0.020	1.07	1.01 – 1.13	0.019
Trial x C3	0.91	0.86 – 0.97	**0.002**	0.91	0.86 – 0.97	**0.002**
C1 x SDS Cannabis	1.01	0.96 – 1.07	0.624			
C2 x SDS Cannabis	1.02	0.97 – 1.08	0.424			
C3 x SDS Cannabis	0.99	0.94 – 1.05	0.838			
Trial x SDS Cannabis	0.99	0.93 – 1.05	0.677			
C1 x SDS Cannabis x Trial	0.98	0.92 – 1.03	0.376			
C2 x SDS Cannabis x Trial	1.02	0.96 – 1.07	0.580			
C3 x SDS Cannabis x Trial	1.04	0.98 – 1.10	0.211			
Random Effects						
σ^2		3.29			3.29	
τ_{00}		0.18			0.18	
ICC		0.05			0.05	
AIC		7010			7000	

Note: Bolded p-values are viewed as significant (≤.05 in typical analyses, ≤.0167 in contrasts). C1 (1, -1, 1, -1) compares acquisition-contingency to reversal phases; C2 (1, 1, -1, -1) compares easy and hard phases; C3 (1, -1, -1, 1) is theoretically irrelevant, necessary to complete contrasts and compares phases 1 and 4 with phases 2 and 3. Trial remains log-transformed as per original analyses.

Table S8. Statistics for best-fitting and saturated model (SDS Cannabis, switch/stay)

	Saturated model			**Best-fitting model**		
Predictors	*OR*	*CI*	*p*	*OR*	*CI*	*p*
Intercept	0.71	0.55 – 0.92	**0.009**	0.71	0.55 – 0.93	**0.011**
Acc. feedback C1	1.31	1.26 – 1.37	**<0.001**	1.31	1.26 – 1.37	**<0.001**
Acc. feedback C2	1.02	0.95 – 1.10	0.558	1.02	0.95 – 1.10	0.522
Acc. feedback C3	1.10	0.90 – 1.35	0.355	1.11	0.90 – 1.35	0.327
SDS Cannabis	1.14	0.89 – 1.45	0.299			
C1 x SDS Cannabis	1.03	1.00 – 1.07	0.079			
C2 x SDS Cannabis	1.02	0.95 – 1.09	0.646			
C3 x SDS Cannabis	1.04	0.87 – 1.24	0.654			
Random Effects						
σ^2		3.29			3.29	
τ_{00}		0.44			0.45	
ICC		0.12			0.12	
AIC		5820			5816	

Note: Bolded *p*-values are viewed as significant (<.05 in typical analyses, <.0167 in contrast). C1 (-3. 1. 1. 1) compares behaviour after one instance of reward to one/two/three instances of punishment; C2 (0, -2, 1, 1) compares behaviour after one instance of punishment and two/three instances of punishment; C3 (0, 0, -1, 1) compares behaviour after two instances of punishment with three instances of punishment.

References

1. Verdejo-García A, del Mar Sánchez-Fernández M, Alonso-Maroto LM, et al. Impulsivity and executive functions in polysubstance-using rave attenders. *Psychopharmacology* 2010;210(3):377–392. https://doi.org/10.1007/s00213-010-1833-8
2. Harald Baayen R, Milin P. Analyzing reaction times. *Int. J. Psychol. Res.*;2010;3(2):12–28 https://doi.org/10.21500/20112084.807