Methamphetamine-induced adaptation of learning rate dynamics depend on baseline performance.

Hans Kirschner[1], Hanna M. Molla[2], Matthew R. Nassar[3,4], Harriet de Wit[2*], and Markus Ullsperger[1,5,6*]

[1]Institute of Psychology, Otto-von-Guericke University, D-39106 Magdeburg, Germany.[2]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois, USA, [3]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence RI 02912-1821, USA. [4]Department of Neuroscience, Brown University, Providence RI 02912-1821, USA [5]Center for Behavioral Brain Sciences, D-39106 Magdeburg, Germany [6] German Center for Mental Health (DZPG), Center for Intervention and Research on Adaptive and Maladaptive Brain Circuits Underlying Mental Health (C-I-R-C), Halle-Jena-Magdeburg, Germany.

[*]shared senior authorship.

Abstract

The ability to calibrate learning according to new information is a fundamental component of an organism's ability to adapt to changing conditions. Yet, the exact neural mechanisms guiding dynamic learning rate adjustments remain unclear. Catecholamines appear to play a critical role in adjusting the degree to which we use new information over time, but individuals vary widely in the manner in which they adjust to changes. Here, we studied the effects of a low dose of methamphetamine (MA), and individual differences in these effects, on probabilistic reversal learning dynamics in a within-subject, double-blind, randomized design. Participants first completed a reversal learning task during a drug-free baseline session to provide a measure of baseline performance. Then they completed the task during two sessions, one with MA (20 mg oral) and one with placebo (PL). First, we showed that, relative to PL, MA modulates the ability to dynamically adjust learning from prediction errors. Second, this effect was more pronounced in participants who performed moderately low at baseline. These results present novel evidence for the involvement of catecholaminergic transmission on learning flexibility and highlights that baseline performance modulates the effect of the drug.

Introduction

Goal-directed behavior requires organisms to continually update predictions about the world to select actions in the light of new information. In environments that include discontinuities (changepoints) and noise (probabilistic errors), optimal learning requires increased weighting of surprising information during periods of change and ignoring surprising events during periods of stability. A burgeoning literature suggests that humans are able to calibrate learning rates according to the statistical content of new information (Behrens et al., 2007; Cook et al., 2019; Diederen et al., 2016; Nassar et al., 2019; Nassar et al., 2010; Razmi & Nassar, 2022), albeit to varying degrees (Kirschner et al., 2022; Kirschner et al., 2023; Nassar et al., 2016; Nassar et al., 2012; Nassar et al., 2021).

Although the exact neural mechanisms guiding dynamic learning adjustments are unclear, several neuro-computational models have been put forward to characterize adaptive learning. While these models differ in their precise computational mechanisms, they share the hypothesis that catecholamines play a critical role in adjusting the degree to which we use new information over time. For example, a class of models assumes that striatal dopaminergic prediction errors act as a teaching signal in cortico–striatal circuits to learn task structure and rules (Badre & Frank, 2012; Collins & Frank, 2013; Collins & Frank, 2016; Lieder et al., 2018; Pasupathy & Miller, 2005; Schultz et al., 1997). Another line of research highlights the role of dopamine in tracking the reward history with multiple learning rates (Doya, 2002; Kolling et al., 2016; Meder et al., 2017; Schweighofer & Doya, 2003). This integration of reward history over multiple time scales enables people to estimate trends in the environment through past and recent experiences and adjust actions accordingly (Wilson et al., 2013). Within the broader literature of cognitive control, it has been suggested that dopamine in the prefrontal cortex and basal ganglia is involved in modulating computational tradeoffs such as cognitive stability–flexibility balance (Cools, 2008; Dreisbach et al., 2005; Floresco, 2013; Goschke, 2013; Goschke & Bolte, 2014; Goschke & Bolte, 2018). In particular, it has been proposed that dopamine plays a crucial role in the regulation of meta-control parameters that facilitate dynamic switching between complementary control modes (i.e., shielding goals from distracting information vs. switching goals in response to significant changes in the environment) (Goschke, 2013; Goschke & Bolte, 2014; Goschke & Bolte,

2018). Finally, other theories highlight the importance of the locus coeruleus/norepinephrine system in facilitating adaptive learning and structure learning (Razmi & Nassar, 2022; Silvetti et al., 2018; Yu et al., 2021). Consistent with these neuro-computational models catecholaminergic drugs are known to affect cognitive performance including probabilistic reversal learning (Cook et al., 2019; Dodds et al., 2008; Repantis et al., 2010; Rostami Kandroodi et al., 2021; van den Bosch et al., 2022; Westbrook et al., 2020). Indeed, psychostimulants, such as methamphetamine, that increase extracellular catecholamine availability, can enhance cognition (Arria et al., 2017; Husain & Mehta, 2011; Smith & Farah, 2011) and are used to remediate cognitive deficits in attention deficit hyperactivity disorder (ADHD) (Arnsten & Pliszka, 2011; Prince, 2008). However, the cognitive enhancements vary across tasks and across individuals (Bowman et al., 2023; Cook et al., 2019; Cools & D'Esposito, 2011; Garrett et al., 2015; Rostami Kandroodi et al., 2021; van den Bosch et al., 2022; van der Schaaf et al., 2013) and the mechanisms underlying this variability remain poorly understood.

There is evidence that the effects of catecholaminergic drugs depend on an individual's baseline dopamine levels in the prefrontal cortex (PFC) and striatum (Cohen & Servan-Schreiber, 1992; Cools & D'Esposito, 2011; Dodds et al., 2008; Durstewitz & Seamans, 2008; Rostami Kandroodi et al., 2021; van den Bosch et al., 2022). Depending on individual baseline dopamine levels the administration of catecholaminergic drugs can promote states of cognitive flexibility or stability. For example, pushing dopamine from low to optimal (medium) levels may increase update thresholds in the light of new information (i.e., facilitating shielding/stability) , whereas if a drug pushes dopamine either too high or too low may decrease update thresholds (i.e., facilitating shifting/flexibility) (Durstewitz & Seamans, 2008; Goschke & Bolte, 2018).

Here, we argue that baseline performance should be considered when studying the behavioral effects of catecholaminergic drugs effects. For example, Rostami Kandroodi et al. (2021) reported that the re-uptake blocker methylphenidate did not alter reversal learning overall, but preferentially improved performance in participants with higher working memory capacity. To investigate the role of baseline performance in drug challenge studies, it is important to control for several factors. First, the order of drug and placebo sessions must be balanced to control for practice effects (Bartels et al., 2010; Garrett et al., 2015; MacRae et al., 1988; Servan-

Schreiber et al., 1998). Second, it is desirable to obtain an independent measure of baseline performance that is not confounded with the drug vs placebo comparison. Thus, participants may be stratified based on their performance on an independent session.

In the present study, we studied the effects of methamphetamine, a stimulant that increases monoaminergic transmission, on probabilistic reversal learning dynamics in a within-subject, double-blind, randomized design. The effects of the drug on a reversal learning task were examined in relation to participants' baseline level of performance. Baseline performance was determined during an initial drug-free session. Then, participants completed the task during two sessions after receiving placebo (PL) and 20 mg of methamphetamine (MA; order counterbalanced).

The task used to study adaptive learning dynamics was a reversal variant of an established probabilistic learning task (Fischer & Ullsperger, 2013; Jocham et al., 2014; Kirschner et al., 2022; Kirschner et al., 2023). On each trial, subjects made a choice to either gamble or avoid gambling on a probabilistic outcome, in response to a stimulus presented in the middle of the screen (see Figure 2A). A gamble could result in a gain or loss of 10 points, depending on the reward contingency associated with that stimulus. In choosing not to gamble, subjects avoided losing or winning points, but they were informed what would have happened if they had chosen to gamble. The reward contingency changed every 30-35 trials. By learning which symbols to choose and which to avoid, participants could maximize total points. A novel feature of this modified version of the task is that we introduced different levels of noise (probability) to the reward contingencies. Here, reward probabilities could be less predictable (30% or 70%, i.e., high outcome noise) or more certain (20% or 80%; i.e., low outcome noise). This manipulation allowed us to study the effect of MA on the dynamic balancing of updating and shielding beliefs about reward contingencies within different levels of noise in the task environment. To estimate learning rate adjustments, we fit a nested set of reinforcement learning models, that allowed for trial-by-trial learning rate adjustments.

We found that MA improved participants' performance in the task, but this effect was driven mainly by a greater improvement in performance in those participants who performed poorly during the baseline session. Modeling results suggested that MA helps performance by

adaptively shifting the relative weighting of surprising outcomes based on their statistical context. Specifically, MA facilitates down-weighting of probabilistic errors in stages of less predictable reward contingencies. Together, these results reveal novel insights into the role of catecholamines in adaptive learning behavior and highlights the importance to consider individual difference at baseline.

Results

97 healthy young adults completed the probabilistic learning task (Figure 2) (Fischer & Ullsperger, 2013; Jocham et al., 2014; Kirschner et al., 2022; Kirschner et al., 2023) on three separate sessions, an initial drug-free session, and after PL and MA. The study followed a double-blinded cross-over design, whereby 50% of participants received MA first, and 50% of participants PL first. Table 1 shows the demographic characteristics of the participants grouped by their task performance during the baseline session. The groups did not differ significantly on any of the variables measured. In a first analysis, we checked for general practice effects across the three task completions based on the total points earned in the task. We found a strong practice effect ($F(2,186) = 14.53$, $p < .001$) with better performance on session two and three compared to session one (baseline). There was no difference in the total scores between session two and three (see Figure 2B). These results suggest that the baseline session may have minimized order effects between MA and PL sessions (see also results and discussion below). The key findings detailed below are summarized in a schematic figure presented in the discussion section (Figure 10).

Table 1. Demographics and drug use characteristics of study participants (n = 94)

Demographic Categories	**low baseline performers** n*(%) or mean (*SD*)*	**high baseline performers** n*(%) or mean (*SD*)*
Sex (M/F)	17/30 (36/64%)	29/18 (62/38%)
Age (years)	24.3 (± 3.9)	24.7 (± 4.1)
BMI	23.1 (± 2.6)	23.3 (± 2.3)
Education (years)	15.8 (± 1.6)	15.7 (± 1.6)

Race/Ethnicity		
American Indian/Alaskan	1 (2%)	0 (0%)
Asian	10 (22%)	8 (17%)
Black or African American	2 (4%)	0 (0%)
Hispanic	7 (15%)	6 (13%)
White	24 (51%)	26 (55%)
More than one race	1 (2%)	6 (13%)
Not reported	2 (4%)	1 (2%)
Drug Use (past month)		
Caffeinated drinks/day	1.0 (± 0.8)	1.3 (± 1.2)
Alcoholic drinks/week	3.8 (± 3.0)	3.9 (± 2.6)
Cannabis uses/month	4.2 (± 6.7)	2.5 (± 4.6)
Daily nicotine users	2 (4%)	1 (2%)
Lifetime stimulant use		
People who previously used at least once (prescription)	3 (6%)	2 (4%)
People who previously used at least once (recreationally)	8 (17%)	9 (19%)
Other lifetime drug use (median number of times used)		
Cannabis	30.5	20.0
Opiates	0	0
Hallucinogens	1	0

Subjective drug effects

MA administration significantly increased 'feel drug effect' ratings compared to PL, at 30, 50, 135, 180, and 210 min post-capsule administration (see Figure 1; Drug x Time interaction $F(5,555) = 38.46$, $p < 0.001$). In the MA session, no differences in the 'feel drug effect' were observed between low and high baseline perfromer, including peak change-from-baseline ratings (rating at 50 min post-capsule: low baseline performer: 48.36(4.29) vs. high baseline performer: 47.21 (4.44); $t(91) = 0.18$, $p = 0.85$, $d = 0.03$; rating at 135 min post-capsule: low baseline performer: 37.27 (4.15) vs. high baseline performer: 45.38 (3.84); $t(91) = 1.42$, $p = 0.15$, $d = 0.29$).

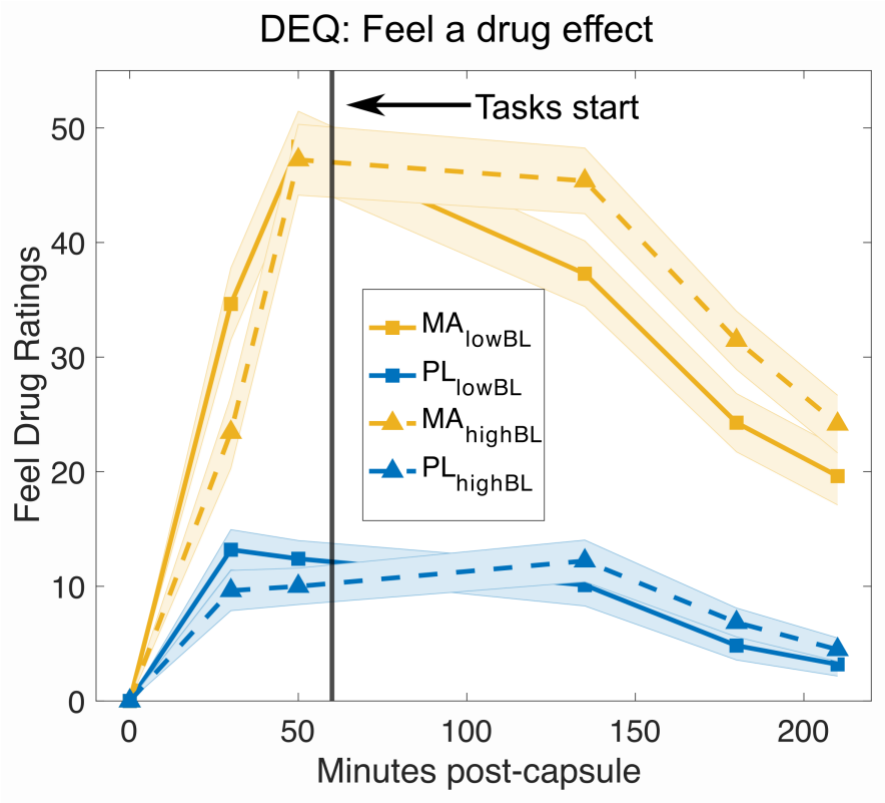


Figure 1. Subjective drug effects post-capsule administration.
MA increased 'feel drug effect' ratings compared to placebo. The scale for the ratings of Feeling a drug effect range from 0 to 100. The vertical black line indicates the time at which the task was started. Ratings of 'feeling' a drug effect did not differ significantly between low vs high baseline performers. (all p > .05). DEQ = Drug Effects Questionnaire (Morean et al., 2013).

Drug effects on overall performance and RT.

In general, participants learned the task well, based on the observation that their choice behavior largely followed the underlying reward probabilities of the stimuli across the sessions (see Figure 4D-F). When all subjects were considered together, we did not find a performance benefit under MA quantified by the total points scored in the task (MA: 736.59 (37.11) vs. PL: 702.02 (38.305); $t(93) = 1.38$, $p = 0.17$, d = 0.10). When participants were stratified by their baseline performance (median spilt on total points at baseline), we found a marginally significant Drug x Baseline Performance Group interaction (Drug x Baseline Performance Group interaction: $F(1,92) = 3.20$, $p = 0.07$; see Figure 2C and Figure 7A). Post hoc t tests revealed that compared to PL, MA improved performance marginally in participants with poor baseline performance (total points

MA: 522.55 (53.79) vs. PL: 443.61 (47.81); $t(46) = 1.85$, $p = 0.071$, d = 0.23). MA did not, however, improve performance in the high baseline performance group (total points MA: 950.63 (26.15) vs. PL: 960.42 (27.26); $t(46) = -0.38$, $p = 0.698$, d = 0.05). In control analyses we ensured that these effects are not driven by session-order effects (see also section on session control analyses below). Results showed no effect of Session ($F(1,92) = 0.71$, $p = 0.40$) and no Session x Baseline Performance Group interaction ($F(1,92) = 0.59$, $p = 0.44$; see Figure 1C). There was a trend for slightly faster RTs under MA (PL: 544.67ms (9.87) vs. MA: 533.84ms (11.51); $t(93) = 1.75$, $p = 0.08$, d = 0.10). This speed effect appeared to be independent of baseline performance (Drug x Baseline Performance Group interaction: $F(1,92) = 0.45$, $p = 0.50$). Moreover, MA was associated with reduced RT variability (average individual SD of RTs: PL: 193.74 (6.44) vs. MA: 178.98 (5.47); $t(93) = 2.54$, $p = 0.012$, d = 0.25). Reduced RT variability has previously been associated with increased attention and performance (Esterman et al., 2012; Karamacoska et al., 2018). Two-way ANOVA on RT variability revealed an effect of baseline performance ($F(1,92) = 4.52$, $p = 0.03$), with increased RT variability in low baseline performers across the drug sessions (low baseline performance: 197.27 (6.48) vs. high baseline performance: 175.45 (5.29)). Moreover, there was an effect of Drug ($F(1,92) = 6.87$, $p = 0.01$), and a Drug x Baseline Performance Group interaction ($F(1,92) = 6.97$, $p = 0.009$). Post hoc t tests indicated that the MA-related reduction in RT variability was specific to low baseline performers (PL: 212.07 (9.84) vs. MA: 182.46 (7.98); $t(46) = 3.04$, p = 0.003, d = 0.48), whereas MA did not affect high baseline performers RT variability (PL: 175.40 (7.51) vs. MA: 175.50 (7.55); $t(46) = -0.02$, $p = 0.98$, d < 0.01).

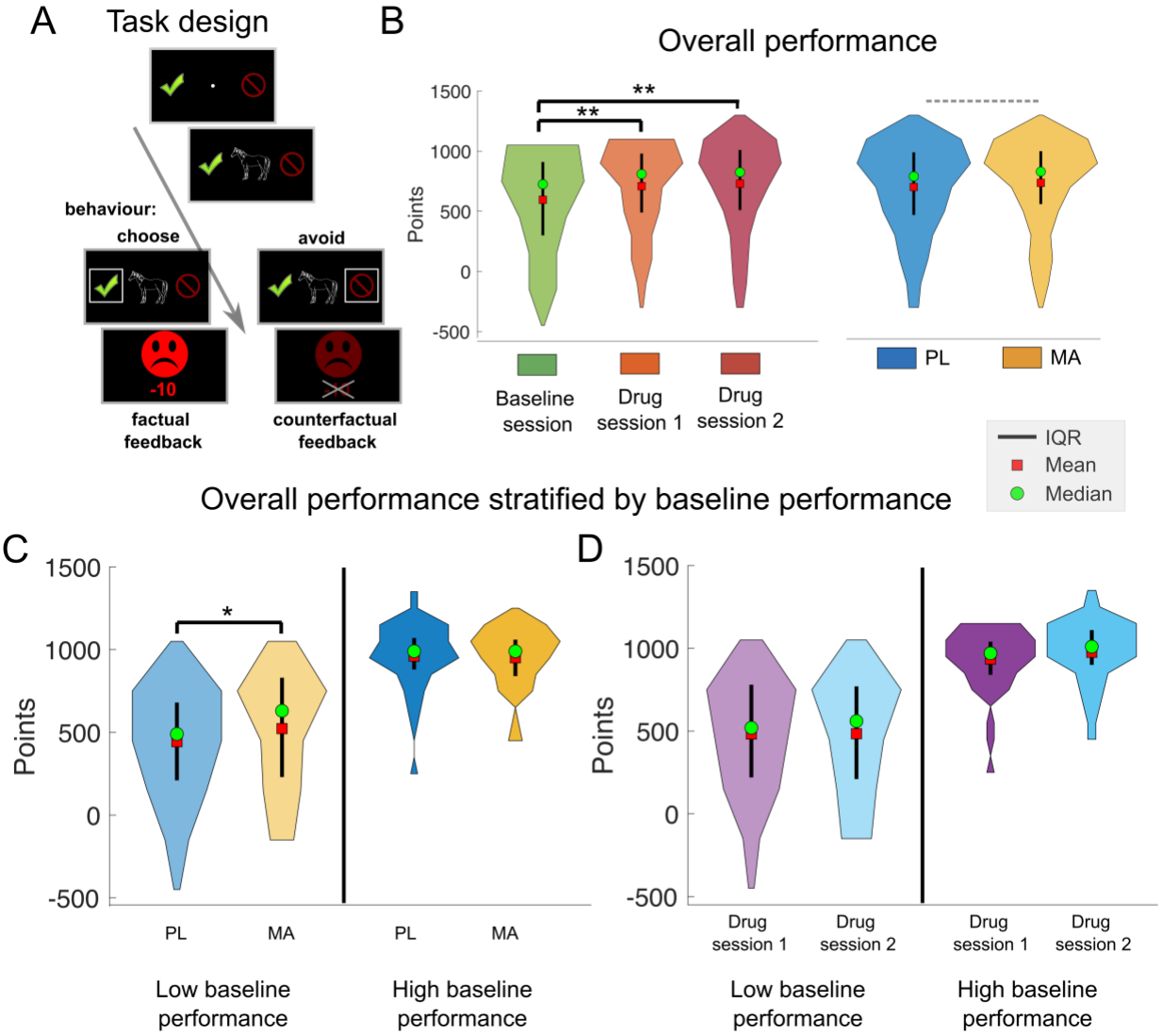


Figure 2. Methamphetamine improved performance in a modified probabilistic reversal learning task only in participants who performed the task poorly at baseline.

(A) Schematic of the learning task. Each trial began with the presentation of a random jitter between 300 ms and 500 ms. Hereafter, a fixation cross was presented together with two response options (choose – green tick mark; or avoid – red no-parking sign). After the fixation cross, the stimulus was shown centrally until the participant responded or for a maximum duration of 2000 ms. Thereafter, participants' choices were confirmed by a white rectangle surrounding the chosen option for 500 ms. Finally, the outcome was presented for 750 ms. If subjects chose to gamble on the presented stimuli, they received either a green smiling face and a reward of 10 points or a red frowning face and a loss of 10 points. When subjects avoided a symbol, they received the same feedback but with a slightly paler color and the points that could have been received were crossed out to indicate that the feedback was fictive and had no effect on the total score. A novel feature of this modified version of the task is that we introduced different levels of noise (probability) to the reward contingencies. Here, reward probabilities could be less predictable (30% or 70%), more certain (20% or 80%), or random (50%). **(B)** Total points earned in the task split up in

sessions (baseline, drug session 1 and 2) and drug condition (PL vs. MA). Results show practice effects but no differences between the two drug sessions (baseline vs. drug session 1: 595.85 (39.81) vs. 708.62 (36.93); $t(93) = -4.21$, $p = 5.95^{-05}$, d = 0.30; baseline vs. drug session 2: 595.85 (39.81) vs. 730.00 (38.53); $t(93) = -4.77$, $p = 6.66^{-06}$, d = 0.35; session 1 vs. session 2: $t(93) = -0.85$, $p = 0.399$, d = 0.05). Dashed gray indicates no significant difference on/off drug (Δ~35 points) **(C)** Interestingly, when we stratified drug effects by baseline performance (using median split on total points at baseline), we found that there was a trend towards better performance under MA in the low baseline performance group (n=47, p = .07). **(D)** Overall performance in drug session 1 and 2 stratified by baseline performance. Here, baseline performance appears not to affect performance in drug session 1 or 2. *Note.* IQR = inter quartile range; PL = Placebo; MA = methamphetamine.

Methamphetamine improves learning performance when reward contingencies are less predictable.

Next, to get a better understanding of how MA affects learning dynamics, we investigated the probability of correct choice (i.e., choosing the advantageous stimuli and avoiding disadvantageous stimuli) across successive reversals. As shown in Figure 3 the drug did not affect initial learning. However, the drug improved performance later in learning, particularly for stimuli with less predictable reward probabilities (see Figure 3B) and in subjects with low baseline performance. To quantify this observation, we first applied the Bai-Perron multiple break point test (see Methods) to find systematic breaks in the learning curves allowing us to divide learning into early and late stages. We applied the test to the reversal learning data across subjects. One break point was identified at 10 trials after a reversal (indexed by the vertical lines in Figure 3). We did not find drug differences when considering all reversals (PL: 0.84 (0.01) vs. MA 0.85 (0.01); t(93) = -1.14, p = 0.25, d = 0.07) and reversals to stimuli with high reward probability certainty (PL 0.86 (0.01) vs. MA 0.87 (0.01); t(93) = -0.25, p = 0.80, d = 0.02). Interestingly, we found a trend for increased learning under MA for stimuli with less predictable rewards (PL 0.80 (0.01) vs. 0.82 (0.01); $t(93) = -1.80$, $p = 0.07$, d = 0.14). Two-way ANOVA on the averaged probability of correct choice during the late stage of learning revealed a Drug x Baseline Performance Group interaction ($F(1,92) = 4.85$, $p = 0.03$; see Figure 7B). Post hoc *t* tests revealed that subjects performing lower at baseline appeared to benefit from MA (average accuracy late learning PL: 0.69 (0.02) vs. MA 0.74 (0.02); $t(46) = -2.59$, $p = 0.01$, d = 0.32), whereas there was no difference between MA and PL in the high baseline performance group (PL: 0.91 (0.01) vs. MA: 0.91 (0.01); $t(46) = 0.29$, p =

0.77, d = 0.04). We did not find other differences in reversal learning (all p > 0.1). In control analyses we split the learning curves into other possible learning situations in the task (i.e., acquisition, first reversal learning etc.). Here no drug related effects emerged (see Supplementary Figure1).

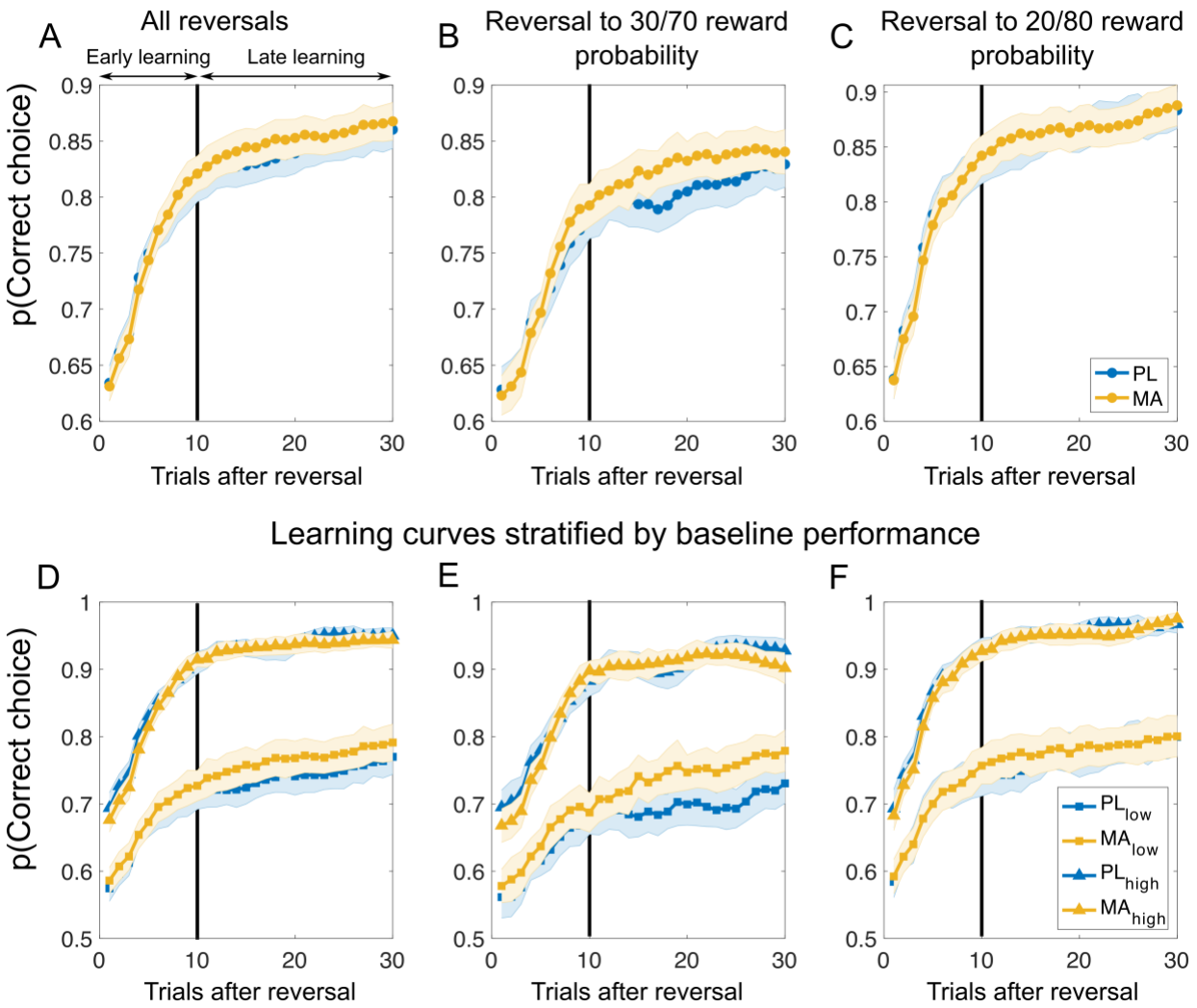


Figure 3. Learning curves after reversals suggest that methamphetamine improves learning performance in phases of less predictable reward contingencies in low baseline performer.

Top panel of the Figure shows learning curves after all reversals **(A)**, reversals to stimuli with less predictable reward contingencies **(B)**, and reversals to stimuli with high reward probability certainty **(C)**. Bottom panel displays the learning curves stratified by baseline performance for all reversals **(D)**, reversals to stimuli with less predictable reward probabilities **(E)**, and reversals to stimuli with high reward probability certainty **(F)**. Vertical black lines divide learning into early and late stages as suggested by the

Bai-Perron multiple break point test. Results suggest no clear differences in the initial learning between MA and PL. However, learning curves diverged later in the learning, particular for stimuli with less predictable rewards **(B)** and in subjects with low baseline performance **(E)**. *Note*. PL = Placebo; MA = methamphetamine; Mean/SEM = line/shading. Data plots are smoothed with a running average (+/- 2 trials), leading to slightly overestimated values after reversals. Additional analyses confirm that the probability of choosing the correct response after reversals is not above chance level (t-test against chance: all reversals: t(93)=1.64,p=0.10,d=0.17, 99% CI[0.49,0.55]; reversal to low outcome noise: t(93)=1.67,p=0.10,d=0.17, 99% CI [0.49,0.56]; Reversal to high outcome noise: t(93)=0.87,p=0.38,d=0.09, 99% CI [0.47,0.56]).

Computational modeling results

To gain a better mechanistical understanding of the trial-to-trial learning dynamics we constructed a nested model set built from RL models (see methods) that included the following features: (1) an inverse temperature parameter of the softmax function used to convert trial expected values to action probabilities (β), (2) a play bias term that indicates a tendency to attribute higher value to gambling behavior, and (3) an intercept term for the effect of learning rate on choice behavior. Additional parameters controlled trial-by-trial modulations of the learning rate including feedback confirmation (confirmatory feedback was defined as factual wins and counterfactual losses, disconfirmatory feedback was defined as factual losses and counterfactual wins), feedback modality (factual vs. counterfactual) and weighting of the learning rate as a function of the absolute value of previous prediction error (parameter Eta, determining the influence of surprise about the outcome on learning; Li et al., 2011). The winning model (as measured by lowest BIC and achieving protected exceedance probabilities of 100%) was one that allowed the learning rate to vary based on whether the feedback was confirmatory or not and the level of surprise of the outcome (see Figure 4A). Sufficiency of the model was evaluated through posterior predictive checks that matched behavioral choice data (see Figure 4D-F and Figure 5) and model validation analyses (see Supplementary Figure 2). Specifically, using individual maximum likelihood parameter estimates, we simulated task performance and confirmed that MA increases performance later in learning for stimuli with less predictable reward probabilities, particularly in subjects with low baseline performance (Figure 5A; mean ±

SD: simPL low performance: 0.69 ± 0.01 vs. simMA low performance: 0.72 ± 0.01; $t(46) = 2.00$, $p = 0.03$, $d = 0.23$). We did not find evidence for differences in model fit between the groups (avg. BIC PL: 596.77 (21.63) vs. MA: 599.66 (19.85); $t(93) = -0.25$, $p = 0.80$, $d = 0.01$).

Next, we compared MAs effect on best-fitting parameters of the winning model (see Figure 4B-C). We found that eta (the parameter controlling dynamic adjustments of learning rate according to recent absolute prediction errors) was reduced under MA (eta MA: 0.24 (0.01) vs. PL 0.30 (0.01); $t(93) = -3.005$, $p = 0.003$, $d = 0.43$). When we stratified drug effects by baseline performance, we found a marginally significant Drug x Baseline Performance Group interaction ($F(1,92) = 3.09$, $p = 0.08$; see Figure 7C)). Post hoc t tests revealed that compared to PL, MA affected eta depending on baseline performance in the task. Here, subjects performing less well at baseline showed smaller eta's (eta MA: 0.24 (0.01) vs. 0.33 (0.02); $t(46) = -3.06$, $p = 0.003$, $d = 0.67$), whereas there was no difference between MA and PL in the high baseline performance group MA: 0.23 (0.01) vs. 0.26 (0.01); $t(46) = -1.03$, $p = 0.31$, $d = 0.18$). We did not find drug-related differences in any other model parameters (all $p > 0.1$).

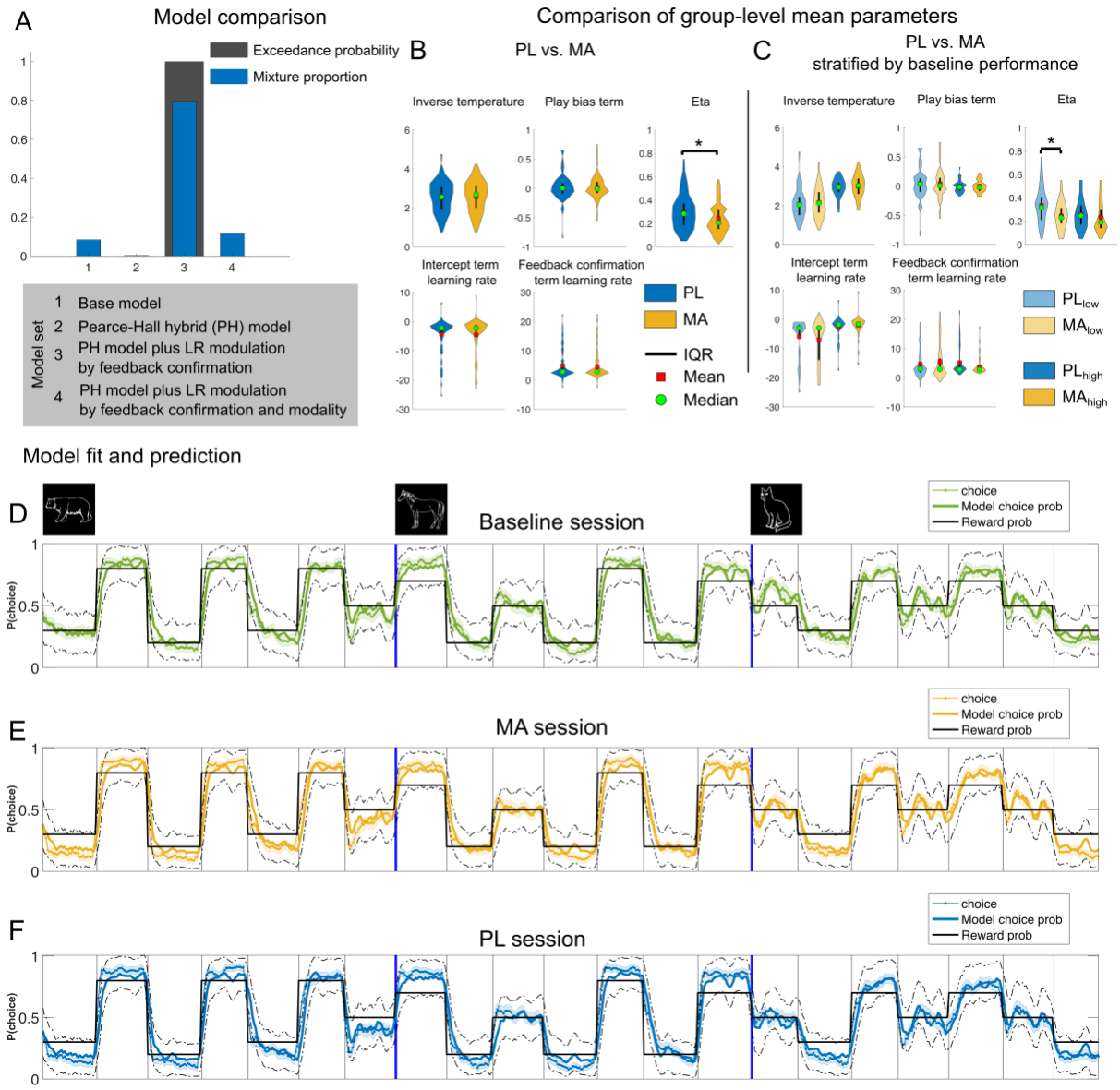


Figure 4. Computational modeling results reveal that methamphetamine affects the model parameter controlling dynamic adjustments of learning rate.

(A) Model comparison. Bayesian model selection was performed using -0.5*BIC as a proxy for model evidence (Stephan et al., 2009). The best fitting mixture model assigned proportions to each model based on the frequency with which they provided the "best" fit to the observed participant data (Mixture proportion; blue bars) and estimated the probability with which the true population mixture proportion for a given model exceeded that of all others (Exceedance probability; black bars). The hybrid model plus learning rate modulation by feedback confirmatory (model 3) provided the best fit to the majority of participants and had an exceedance probability near one in our model set. **(B-C)** Comparison of parameter estimates from the winning model on-/ off drug. Stars indicate significant difference for the respective parameter. Results suggest that only the parameter controlling dynamic adjustments of learning rate according to recent prediction errors, eta, was affected by our pharmacological manipulation. **(D-F)** Modelled and choice behavior of the participants in the task, stretched out for all stimuli. Note that in the

task the different animal stimuli were presented in an intermixed and randomized fashion, but this visualization allows to see that participants' choices followed the reward probabilities of the stimuli. Data plots are smoothed with a running average (+/- 2 trials). Ground truth corresponds to the reward probability of the respective stimuli (good: 70/80%; neutral: 50%; bad: 20/30%). Dashed black lines represent 95% confidence intervals derived from 1000 simulated agents with parameters that were best fit to participants in each group. Model predictions appear to capture the transitions in choice behavior well. Mean/SEM = line/shading. *Note.* IQR = inter quartile range; PL = Placebo; MA = methamphetamine;

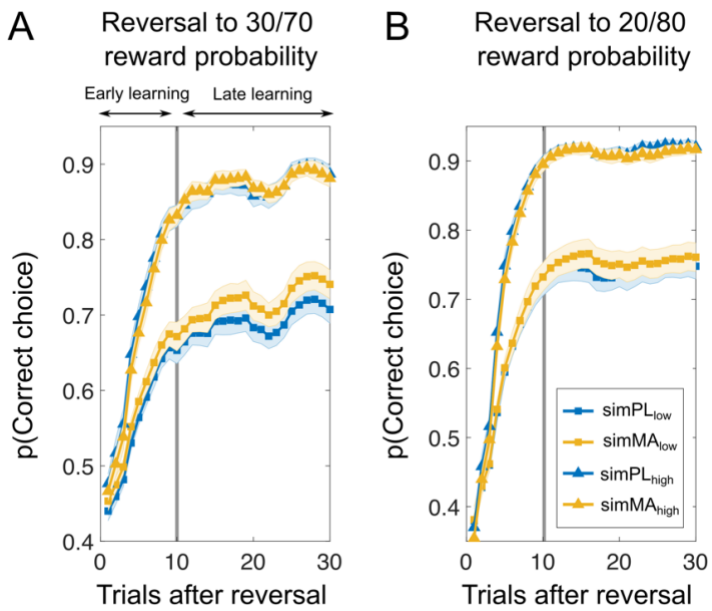


Figure 5. Simulated task performance based on individual maximum likelihood parameter estimates reflects drug-induced behavioral differences. Simulated performance (*y*-axis) is plotted against trials after reversals (*x*-axis) for low (blue) and high (yellow) baseline performers. Using each participant's estimated parameters, 100 artificial agents were simulated playing the task, and their choices were averaged to represent each participant's behavior. The simulation shows that MA increases performance later in learning for stimuli with high outcome noise, particularly in subjects with low baseline performance **(A)**. In contrast, no drug effect was observed for stimuli with low outcome noise **(B).**

Methamphetamine affects learning rate dynamics.

Next, we investigated how the model parameters fit with trial-by-trial modulations of the learning rate. Learning rates in our best fitting model were dynamic and affected by both model parameters and their interaction with feedback. Learning rate trajectories after reversals are depicted in Figure 6. As suggested by lower eta scores, MA appears to be associated with reduced learning rate dynamics in low-baseline performers. In contrast, low-baseline-performers in the PL condition exhibited greater variability in learning rate (and average LR throughout) rendering their choices more erratic. Consistent with this, on many trials their choices were driven by the most recent feedback, as their learning rates on a large subset of trials in later learning stages (on average 9 out of 11; Figure 6H) were greater than 0.5. Specifically, variability in learning rate

(average individual SD of learning rate) was reduced in both early and late stages of learning across all reversals (early PL: 0.20 (0.01) vs. MA: 0.17 (0.01); $t(93) = 2.72$, $p = 0.007$, $d = 0.36$; late PL: 0.18 (0.01) vs. MA: 0.15 (0.01); $t(93) = 2.51$, $p = 0.01$, $d = 0.33$), as were reversals to stimuli with less predictable rewards (early PL: 0.19 (0.01) vs. 0.16 (0.01); $t(93) = 2.98$, $p = 0.003$, $d = 0.39$; late PL: 0.18 (0.01) vs. MA: 0.16 (0.01); $t(93) = 2.66$, $p = 0.009$, $d = 0.35$). Reversals to stimuli with high outcome certainty were also associated with decreased learning rate variability after MA administration (early PL: 0.18 (0.01) vs. MA: 0.15 (0.01); $t(93) = 2.57$, $p = 0.01$, $d = 0.34$; late PL: 0.18 (0.01) vs. MA: 0.15 (0.01); $t(93) = 2.63$, $p = 0.009$, $d = 0.35$). Two-way ANOVA revealed that this effect depended on baseline performance across all reversals (Drug x Baseline performance: $F(1,92) = 3.47$, $p = 0.06$), reversals to stimuli with less predictable rewards (Drug x Baseline performance: $F(1,92) = 4.97$, $p = 0.02$), and stimuli with high outcome certainty (Drug x Baseline performance: $F(1,92) = 5.26$, $p = 0.03$). Here, reduced variability under MA was observed in low baseline performers (all $p < .006$, all $d > .51$) but not in high baseline performers (all $p > .1$). Together, these patterns of results suggest that people with high baseline performance show a large difference in learning rates after true reversals and during the rest of the task including misleading feedback. Specifically, they show a peak in learning after reversals and reduced learning rates in later periods of a learning block, when choice preferences should ideally be stabilized (see Figure 6C). This results in a better signal-to-noise ratio (SNR) between real reversals and misleading feedback (i.e., surprising outcomes in the late learning stage). In low baseline performers the SNR is improved after the administration of MA. This effect was particularly visible in stages of the task where rewards were less predictable. To quantify the SNR for less predictable reward contingencies for low baseline performers, we computed the difference between learning rate peaks on true reversals (signal) vs. learning rate peaks after probabilistic feedback later in learning (noise; SNR = signal -noise). The results of this analysis revealed that MA significantly increased the SNR for low baseline performers (PL: 0.01 (0.01) vs. MA: 0.04 (0.01); $t(46) = -2.81$, $p = 0.007$, $d = 0.49$). Moreover, learning rates were generally higher in later stages of learning, when choice preferences should ideally have stabilized (avg. learning rate during late learning for less predictable rewards: PL: 0.48 (0.01) vs. MA: 0.42 (0.01); $t(46) = 3.36$, $p = 0.001$, $d = 0.56$).

Thus far, our results suggest that (1) MA improved performance in subjects who performed poorly at baseline, and (2) that MA reduced learning rate variability in subjects with low baseline performance (driven by significantly lower eta parameter estimates, which improved the SNR between true reversals and misleading feedback particularly for less predictable rewards). Next, we aimed to test how these differences relate to each other. Given that eta causes increased learning after surprising feedback and that we found the biggest drug differences in later stages of learning for stimuli that have less predictable rewards, we tested the association between the probability of making the correct choice after two consecutive probabilistic errors (wins for bad stimuli and losses for good stimuli; in total this happened 8 times in the late learning stage for stimuli with 30/70% reward probability) and eta. We found a significant correlation across participants (see Figure 7), whereby higher etas scores were associated with fewer correct choices (r = .29, p = < .001). There was a trend toward a drug effect, with subjects in MA condition being more likely to make the correct choice after two misleading feedbacks (PL: 0.82 (0.02) vs. 0.84 (0.01); $t(93) = -1.92$, p = 0.06, d = 0.13). Two-way ANOVA revealed, that this effect depended on baseline performance (Drug x Baseline performance: $F(1,92) = 4.27$, $p = 0.04$). Post-hoc t tests indicated higher correct choice probabilities under MA in low baseline performers (PL: 0.70 (0.02) vs. MA: 0.75 (0.02); t(46) = -2.41, p = 0.02, d = 0.30) but not in high baseline performers (PL: 0.92 (0.01234) vs. MA: 0.92 (0.01); t(46) = 0.11, p = 0.91, d = 0.01).

Methamphetamine shifts learning rate dynamics closer to the optimum for low baseline performers.

To better understand the computational mechanism through which MA improved performance in low baseline performers, we first examined how performance in the task related to model parameters from our fits. To do so, we regressed task performance onto an explanatory matrix containing model parameter estimates across all conditions (see Figure 8A). The results of this analysis revealed that variability in several of the parameters was related to overall task performance, with the overall learning rate, feedback confirmation LR adjustments, and inverse

temperature all positively predicting performance and eta and the play bias term negatively predicting it.

While each of these parameters explained unique variance in overall performance levels, only the parameter controlling dynamic adjustments of learning rate according to recent prediction errors, eta, was affected by our pharmacological manipulation (Figure 6B). In particular, eta was reduced in the MA condition, specifically in the low baseline group, albeit to an extent that differed across individuals (Figure 8C). To better understand how changes in eta might have affected overall performance we conducted a set of simulations using the parameters best fit to human subjects, except that we equipped the model with a range of randomly chosen eta values, to examine how altering that parameter might affect performance. The results revealed that simulated agents with low to intermediate levels of eta achieved the best task performance, with models equipped with the highest etas performing particularly poorly (Figure 8D). To illustrate how this relationship between eta and performance could have driven improved performance for some participants under the methamphetamine condition, we highlight four participants with low-moderate eta values under methamphetamine, but who differ dramatically in their eta values in the placebo condition (Figure 8D, inset). Note that the participants who have the largest decreases in eta under the methamphetamines, resulting from the highest placebo levels of eta, would be expected to have the largest improvements in performance. It is noteworthy that low-baseline performers tended to have particularly high values of eta under the baseline condition (low-baseline performers: 0.33 (0.02) vs. high-baseline performers: 0.25 (0.01); $t(46) = 2.59$, $p = 0.01$ $d = 0.53$), explaining why these individuals saw the largest improvements under the methamphetamine condition. Taken together, these results suggest that MA alters performance by changing the degree to which learning rates are adjusted according to recent prediction errors (eta), in particular by reducing the strength of such adjustments in low-baseline performers to push them closer to task-specific optimal values.

While eta seemed to account for the differences in the effects of MA on performance in our low and high performance groups, it did not fully account for performance differences across the two

groups (see Figure 1C and Figure 10A/B). When comparing other model parameters between low and high baseline performer across drug sessions, we found that high baseline performer displayed higher overall inverse temperatures (2.97(0.05) vs. 2.11 (0.08); $t(93) = 7.94$, $p < .001$, d = 1.33). This suggests that high baseline performers displayed higher transfer of stimulus values to actions leading to better performance (as also indicated by the positive contribution of this parameter to overall performance in the GLM). Moreover, they tended to show a reduced play bias (-0.01 (0.01) vs. 0.04 (0.03); $t(93) = -1.77$, $p = 0.08$, d = 0.26) and increased intercepts in their learning rate term (-2.38 (0.364) vs. -6.48 (0.70); t(93) = 5.03, $p < .001$, d = 0.76). Both of these parameters have been associated with overall performance (see Figure 8A). Thus, overall performance difference between high and low baseline performed can be attributed to differences in model parameters other than eta. However, as described in the previous paragraph, differential effects of MA on performance on the two groups were driven by eta.

This pattern of results suggests that MA specifically affects the eta parameter while leaving other parameters, such as the inverse temperature, unaffected. This points to a selective influence on a single computational mechanism. To verify this conclusion, we extended the winning model by allowing each parameter, in turn, to be differentially estimated for MA and PL, while keeping the other parameters fixed at the group (low and high baseline performance) mean estimates of the winning model for the placebo session. These control analyses confirmed that MA affects only the eta parameter in the low-performer group (see Supplementary Table 1) and that parameters did not trade of in our model. A similar effect was observed in a previous study investigating the effects of catecholaminergic drug administration on a probabilistic reversal learning task (Rostami Kandroodi et al., 2021). In that study, methylphenidate was shown to influence the inverse learning rate parameter as a function of working memory span, assessed through a baseline cognitive task. Consistent with our findings, no drug effects were observed on other parameters in their model, including the inverse temperature.

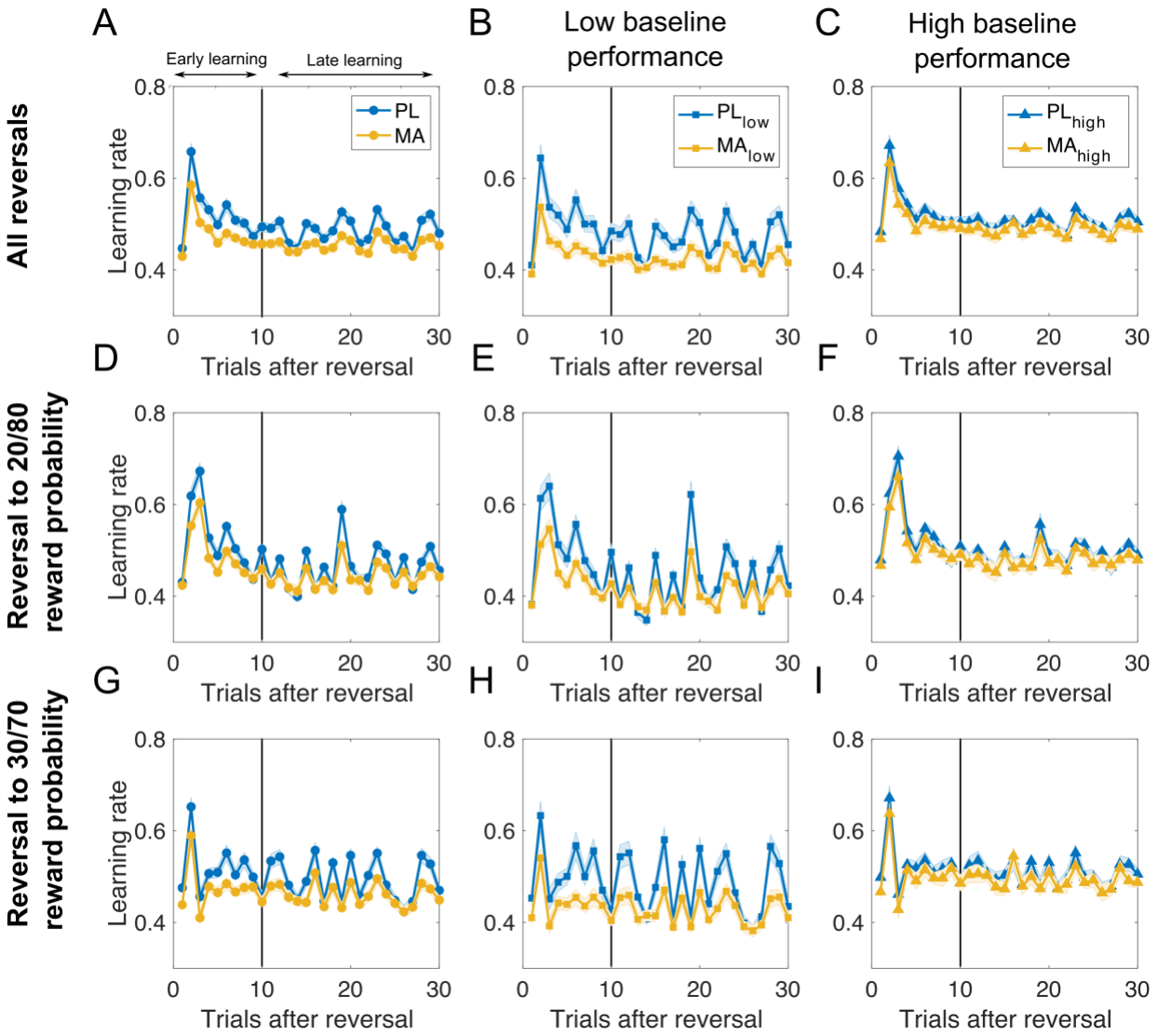


Figure 6. Methamphetamine boosts signal-to-noise ratio between real reversals and misleading feedback in late learning stages.

Learning rate trajectories after reversal derived from the computational model. First column depicts learning rates across all subjects for all reversals **(A)**, reversal to stimuli with high reward probability certainty **(D)**, and reversal to stimuli with noisy outcomes **(G)**. Middle and right column shows learning rate trajectories for subjects stratified by baseline performance (**B, E, H** – low baseline performance; **C, F, I** – high baseline performance). Results suggest that people with high baseline performance show a large difference in learning rates after true reversals and during the rest of the task including misleading feedback. Specifically, they show a peak in learning after reversals and reduced learning rates in later periods of a learning block, when choice preferences should ideally be stabilized **(C)**. This results in a better signal-to-noise ratio (SNR) between real reversals and misleading feedback (i.e., surprising outcomes in the late learning stage). In low baseline performers the SNR is improved after the administration of MA. This effect was particularly visible in stages of the task where rewards were less predictable (**H**). *Note.* IQR = inter quartile range; PL = Placebo; MA = methamphetamine.

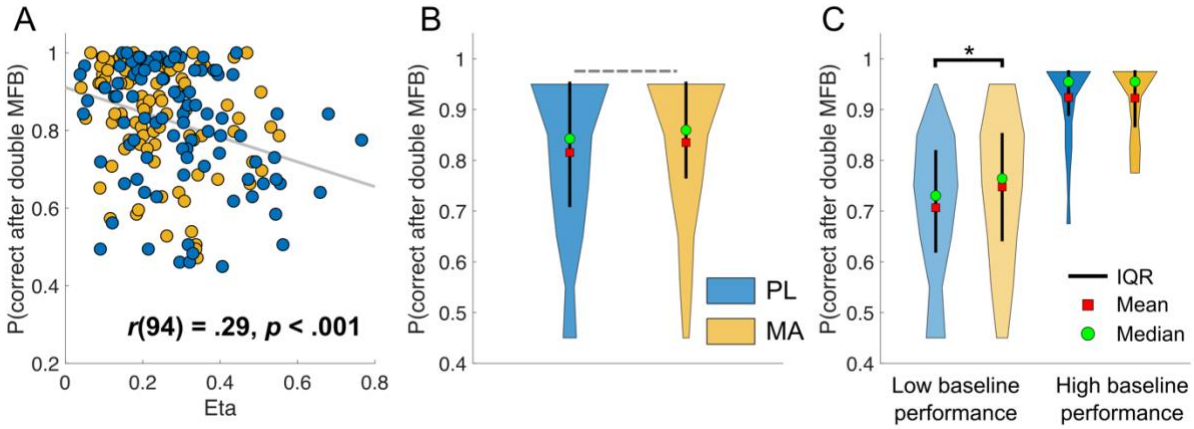


Figure 7. Misleading feedback effects on choice accuracy are modulated by eta and methamphetamine in low baseline performers. This figure shows the association between receiving misleading feedback later in learning (i.e., reward or losses that do not align with a stimulus' underlying reward probability) and the probability of making the correct choice during the next encounter of the same stimulus. Results indicate a negative correlation between the probability of a correct choice after double-misleading feedback and eta **(A)**. Here, the probability of a correct choice after double-misleading feedback decreases with increasing eta. There was a trend (p = .06) that subjects under MA were more likely to make the correct choice after two misleading feedback as compared to PL **(B)**. **(C)** This effect appeared to be dependent on baseline performance, whereby only subjects with low baseline performance seem to benefit from MA (p = 0.02). *Note.* IQR = inter quartile range; PL = Placebo; MA = methamphetamine; MFB = misleading feedback.

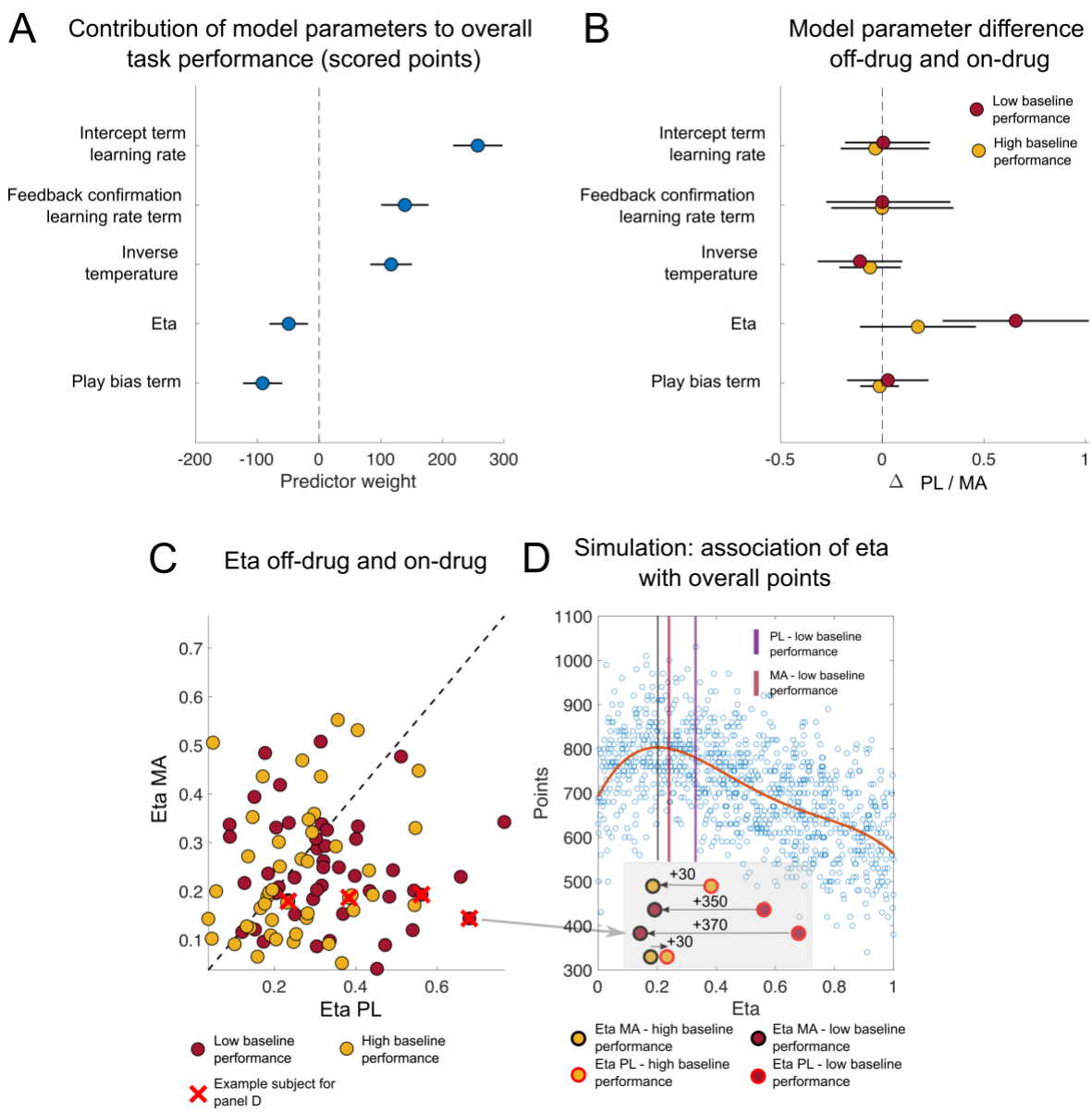


Figure 8. Changes in learning rate adjustment explain drug induced performance benefits in low baseline performers. (A) Regression coefficients and 95% confidence intervals (points and lines; sorted by value) stipulating the contribution of each model parameter estimate to overall participants task performance (i.e., scored points in the task). Play bias and eta (the parameter governing the influence of surprise on learning rate) both made a significant negative contribution to overall task performance, whereas inverse temperature and learning rates were positively related to performance. **(B)** Differences in parameter values for on- and off-drug sessions as quantified by regression coefficients and 95% confidence intervals are plotted separately for high (red) and low (yellow) baseline performers. Note that the drug predominately affected the eta parameter and did so to a greater extent in low baseline performers. **(C)** eta estimates on-drug (y-axis) are plotted against eta estimates off-drug (x-axis) for high baseline performer (yellow points) and low baseline performer (red points). Note that a majority of subjects showed a reduction in eta on-drug vs. off-drug (67.02%). This effect was more pronounced in low

baseline performers (low baseline performers: 74.47%; (low baseline performers: 59.57%). **(D)** To better understand how changes in eta might have affected overall performance we conducted a set of simulations using the parameters best fit to human subjects, except that we equipped the model with a range of randomly chosen eta values to examine how altering that parameter might affect performance (n=1000 agents). The results revealed that simulated agents with low to intermediate levels of eta achieved the best task performance, with models equipped with the highest etas performing particularly poorly. To illustrate how this relationship between eta and performance could have driven improved performance for some participants under the methamphetamine condition, we highlight four participants with low-moderate eta values under methamphetamine, but who differ dramatically in their eta values in the placebo condition (**D**, inset). PL = Placebo; MA = methamphetamine.

Methamphetamine may reduce misinterpretation of high outcome noise in low performer

In our task, outcomes are influenced by two distinct sources of noise: process noise (volatility) and outcome noise (stochasticity). Optimal learning rate should increase with volatility and decrease with stochasticity. Volatility was fairly constant in our task (change points around every 30-35 trials). However, misleading feedback (i.e., outcome noise) could be misinterpreted as indicating another change point because participants don't know the volatility beforehand. Strongly overinterpreting outcome noise as change points will hinder building a correct estimate of volatility and understanding the true structure of the task. Simultaneously estimating poses a challenge, as both contribute to greater outcome variance, making outcomes more surprising. A critical distinction, however, lies in their impact on generated outcomes: volatility increases the autocorrelation between consecutive outcomes, whereas stochasticity reduces it. Recent computational approaches have successfully utilised this fundamental difference to formulate a model of learning based on the joint estimation of stochasticity and volatility (Piray & Daw, 2021; Piray & Daw, 2024). They report evidence that humans successfully dissociate between volatility and stochasticity with contrasting and adaptive effects on learning rates, albeit to varying degrees. Interestingly they show that hypersensitivity to outcome noise, often observed in anxiety disorders, might arise from a misattribution of the outcome noise to volatility instead of stochasticity resulting in increased learning rates and overadjustments to misleading outcomes. It is noteworthy, that we observed a similar hypersensitivity to high outcome noise in low performer in our task that is partly reduced by MA. In an exploratory analysis, we fit two models

to our task structure using modified code provided by Piray and Daw (2021) (see Methods for formal Description of the model). The control model made inference about both the volatility and stochasticity. The lesioned model assumed stochasticity to be small and constant. We show the results of this analyses in Figure 9 and Supplementary Figure S7 and S8). We found that the inability to make inference about stochasticity, leads to misestimation of volatility, particularly for high outcome noise phases (Figure 9 A-B). Consistently, this led to reduced sensitivity of the learning rate to volatility (i.e., the first ten trials after reversals). The model shows similar behaviour to our low performer group, with reduced accuracy in later learnings stages for stimuli with high outcome noise (Figure 9 D). Finally, when we fit simulated data from the two models to our model, we see increased eta parameter estimates for the lesioned model. Together, these results may hint towards an overinterpretation of stochasticity in low performer of our task and that MA has beneficial effects for those individuals as it reduced the oversensitivity to volatility. It should be noted however, that we did not fit these models to our choice behaviour directly as this implementation is beyond the scope of our current study. Yet, our exploratory analyses make testable predictions for future research into the effect of catecholamines on the inference of volatility and stochasticity.

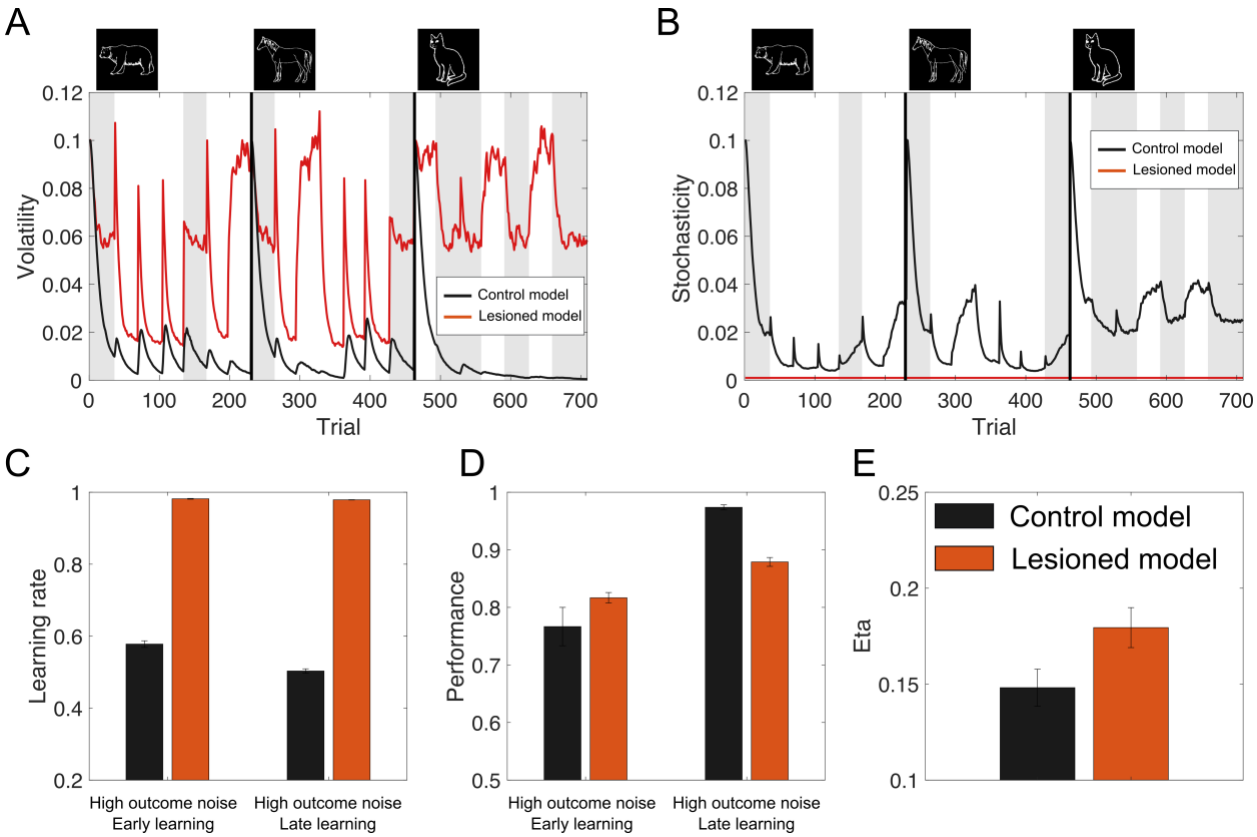


Figure 9. The stochasticity lesion model shows a pattern of learning deficits associated low performer in our task. Behaviour of the lesioned model, in which stochasticity is assumed to be small and constant, is shown along the control model that jointly estimates stochasticity and volatility***. (A-B)*** The inability to make inference about stochasticity, leads to misestimation of volatility, particularly for high outcome noise phases (grey patches show trials with high outcome noise (30/70% reward probability). **(C)** This led to reduced sensitivity of the learning rate to volatility (i.e., the first ten trials after reversals). **(D)** The lesioned model shows similar behaviour to our low performer group, with reduced accuracy in later learnings stages for stimuli with high outcome noise. **(E)** When we fit simulated data from the two models to our model, we see increased eta parameter estimates for the lesioned model.

Control analyses.

To control for the potentially confounding factor session order (i.e., PL first vs. MA first), we repeated the two-way mixed ANOVAs with significant Drug x Baseline Session interactions with session order as a between subject factor. Including session order did not alter the significance of the observed effects and did not interact with the effects of interest (all $p > .24$).

Discussion

To study learning dynamics participants completed a reversal variant of an established probabilistic learning task (Fischer & Ullsperger, 2013; Jocham et al., 2014; Kirschner et al., 2022; Kirschner et al., 2023). Participants completed the task three times: in a baseline session without drug, and after PL and after oral MA (20 mg) administration. We observed a trend towards a drug effect on overall performance, with improved task performance (total points scored in the task) selectively in low baseline performers. Follow-up analyses revealed that MA performance benefits were mainly driven by significantly better choices (i.e., choosing the advantageous stimuli and avoiding disadvantageous stimuli) at later stages after reversals for less predictable reward contingencies. Modeling results suggest that MA is helping performance by adaptively shifting the relative weighting of surprising outcomes based on their statistical context. Specifically, MA facilitated down-weighting of probabilistic errors in phases of less predictable reward contingencies. In other words, in low baseline-performers the SNR between true reversals and misleading feedback is improved after the administration of MA. Moreover, although existing literature has linked catecholamines to volatility-based learning rate adjustments (Cook et al., 2019), we show that these adjustments also relate to other context-dependent adjustments like levels of probabilistic noise. The key findings of this study are summarized in Figure 10.

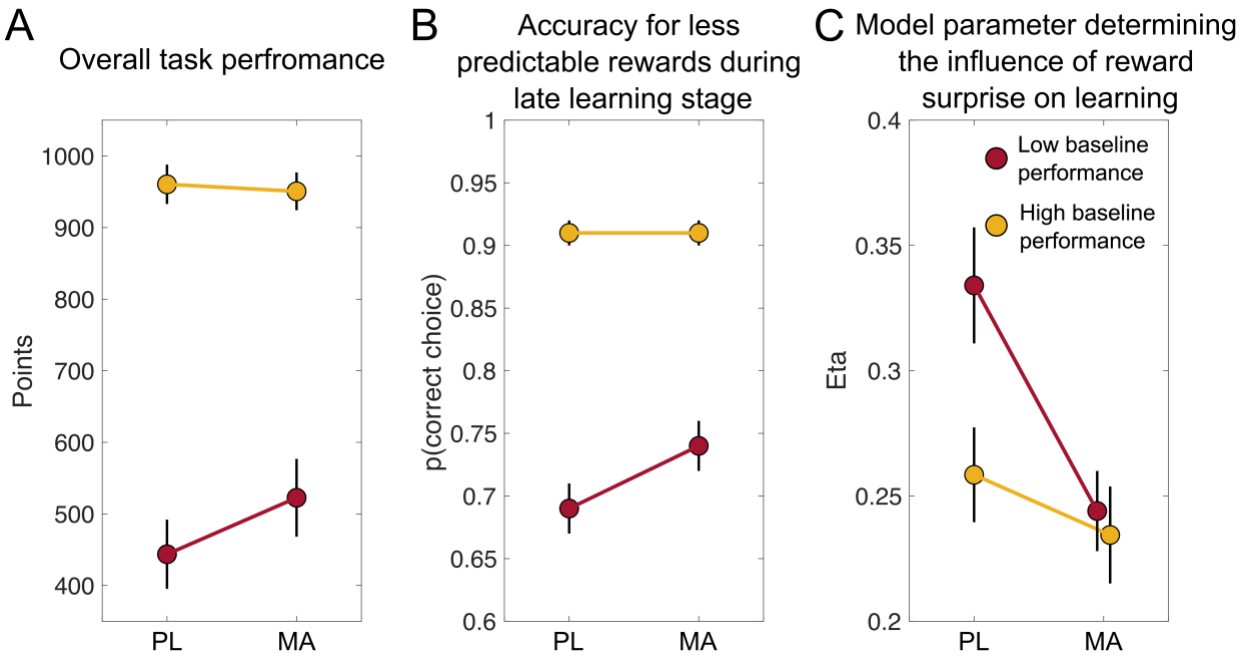


Figure 10. Summary of key findings.

Mean (SEM) scores on three measures of task performance after PL and MA, in participants stratified on low or high baseline performance**. (A)** There was a trend toward a drug effect, with boosted task performance (total points scored in the task) in low baseline performers (subjects were stratified via median split on baseline performance) after methamphetamine (20mg) administration. **(B)** Follow-up analyses revealed that on-drug performance benefits were mainly driven by significantly better choices (i.e., choosing the advantageous stimuli and avoiding disadvantageous stimuli) at later stages after reversals for less predictable reward contingencies (30/70% reward probability). **(C)** To understand the computational mechanism through which methamphetamine improved performance in low baseline performers we investigated how performance in the task related to model parameters from our fits. Our results suggest that methamphetamine alters performance by changing the degree to which learning rates are adjusted according to recent prediction errors (eta), in particular by reducing the strength of such adjustments in low-baseline performers to push them closer to task-specific optimal values.

Methamphetamine affects the relative weighting of reward prediction errors.

A key finding of the current study is that MA affected the relative weighting of reward prediction errors. In our model, adjustments in learning rate are afforded by weighting the learning rate as a function of the absolute value of the previous prediction error (Li et al., 2011). This associability-gated learning mechanism is empirically well supported (Le Pelley, 2004) and facilitates decreasing learning rates in periods of stability and increasing learning rates in periods of change.

MA was associated with lower weighting of prediction errors (quantified by lower eta parameters under MA). Our results comprise an important next step in understanding the neurochemical underpinnings of learning rate adjustments.

Neuro-computational models suggest that catecholamines play a critical role in adjusting the degree to which we use new information. One class of models highlights the role of striatal dopaminergic prediction errors as a teaching signal in cortico–striatal circuits to learn task structure and rules (Badre & Frank, 2012; Collins & Frank, 2013; Collins & Frank, 2016; Lieder et al., 2018; Pasupathy & Miller, 2005; Schultz et al., 1997). The implication of such models is that learning the structure of a task results in appropriate adjustments in learning rates. Optimal learning in our task with high level of noise in reward probabilities in combination with changing reward contingencies required increased learning from surprising events during periods of change (reversals) and reduced learning from probabilistic errors. Thus, neither too low learning adjustments after surprising outcomes (low eta), nor too high learning adjustments after surprising outcomes (high eta) are beneficial in our task structure. Interestingly, MA appears to shift eta closer to the optimum. Exploratory simulation studies using a model that jointly estimates stochasticity and volatility (Piray & Daw, 2021; Piray & Daw, 2024), revealed that MA might reduce the oversensitivity to volatility. In terms of the neurobiological implementation of this effect, MA may prolong the impact of phasic dopamine signals, which in turn facilitates better learning of the task structure and learning rate adjustments (Cook et al., 2019; Marshall et al., 2016; Volkow et al., 2002). Our data, in broad strokes, are consistent with the idea that dopamine in the prefrontal cortex and basal ganglia is involved in modulating meta-control parameters that facilitated dynamic switching between complementary control modes (i.e., shielding goals from distracting information vs. shifting goals in response to significant changes in the environment) (Cools, 2008; Dreisbach et al., 2005; Floresco, 2013; Goschke, 2013; Goschke & Bolte, 2014; Goschke & Bolte, 2018). A key challenge in our task is differentiating real reward reversals from probabilistic misleading feedback which is a clear shielding/shifting dilemma described in the meta-control literature. Our data suggest that MA might improve meta-control of when to shield and when to shift beliefs in low baseline performers.

Moreover, it is possible that MA's effect on learning rate adjustments is driven by its influence on the noradrenaline system. Indeed, a line of research is highlighting the importance of the locus coeruleus/norepinephrine system in facilitating adaptive learning and structure learning (Razmi & Nassar, 2022; Silvetti et al., 2018; Yu et al., 2021). In particular, evidence from experimental studies, together with pharmacological manipulations and lesion studies of the noradrenergic system suggest that noradrenaline is important for change detection (Muller et al., 2019; Nassar et al., 2012; Preuschoff et al., 2011; Set et al., 2014). Thus, the administration of MA may have increased participants' synaptic noradrenaline levels and, therefore, increased the sensitivity to salient events indicating true change points in the task.

It should be noted that other neuromodulators, such as acetylcholine (Marshall et al., 2016; Yu & Dayan, 2005), and serotonin (Grossman et al., 2022; Iigaya et al., 2018), have also been associated with dynamic learning rate adjustment. Future studies should compare the effects of neuromodulator-specific drugs for example a dopaminergic modulator, a noradrenergic modulator, a cholinergic modulator, and a serotonin modulator to make neuromodulator-specific claims (for example see Marshall et al., 2016). Taken together, it is likely that in our study MA effects on learning rate adjustments are driven by multiple processes that perhaps also work in concert. Moreover, because we only administered a single pharmacological agent, our results could reflect general effects of neuromodulation.

Our results are in line with recent studies that show improved performance under methylphenidate (MPH) by making learning more robust against misleading information. For example, Fallon et al. (2017) showed that (MPH) helped participants to ignore irrelevant information but impaired the ability to flexibly update items held in working memory. Another study showed that (MPH) improved performance by adaptively reducing the effective learning rate in participants with higher working memory capacity (Rostami Kandroodi et al., 2021). These studies highlight the complex effects of MPH on working memory and the role of working memory in reinforcement learning (Collins & Frank, 2012; Collins & Frank, 2018). It could be that the effect of MA on learning rate dynamics reflect a modulation of interactions between working memory and reinforcement learning strategies. However, it should be acknowledged that our

task was not designed to parse out specific contributions of the reinforcement learning system and working memory to performance.

Methamphetamine performance enhancement depends on initial task performance

Another key finding of the current study is that the benefits of MA on performance depend on the baseline task performance. Specifically, we found that MA selectively improved performance in participants that performed poorly in the baseline session. However, it should be noted, that all the drug x baseline performance interactions, including for the key computational eta parameter did not reach the statistical threshold, and only tended towards significance. We used a binary discretization of baseline performance to simplify the analysis and presentation. To parse out the relationship between methamphetamine effects and baseline performance into finer level of detail, we conducted additional linear mixed-effects model (LMM) analyses using a sliding window regression approach (see supplementary results and supplementary figure S4 and S5). A key thing to notice in the sliding regression results is that, while each regression reveals that drug effects depend on baseline performance, they do so non-linearly, with most variables of interest showing a saturating effect at low baseline performance levels and the strongest slope (dependence on baseline) at or near the median level of baseline performance, explaining why our median splits were able to successfully pick up on these baseline-dependent effects. Together, these results suggest that methamphetamine primarily affects moderately low baseline performer. It is noteworthy to highlight again that we had a separate baseline measurement from the placebo session, allowing us to investigate baseline-dependent changes while avoiding typical concerns in such analyses like regression to the mean (Barnett et al., 2004). This design enhances the robustness of our baseline-dependent effects. It is important to note, that MA did not bring performance of low baseline performers to the level of performance of high baseline performers. We speculate that high performers gained a good representation of the task structure during the orientation practice session, taking specific features of the task into account (change point probabilities, noise in the reward probabilities). This is reflected in a large signal to noise ratio between real reversals and misleading feedback. Because the high

performers already perform the task at a near-optimal level, MA may not further enhance performance (see Supplementary Figure S6 for additional evidence for this claim). Intriguingly, the data do not support an inverted-u-shaped effect of catecholaminergic action (Durstewitz & Seamans, 2008; Goschke & Bolte, 2018) given that performance of high performers did not decrease with MA. One could speculate that catecholamines are not the only factor determining eta and performance. Perhaps high performers have a generally more robust/resilient decision-making system which cannot be perturbed easily. Probably one would need even higher doses of MA (with higher side effects) to impair their performance.

These results have several interesting implications. First, a novel aspect of our design is that, in contrast to most pharmacological studies, participants completed the task during a baseline session before they took part in the two drug sessions. Drug order and practice effects are typical nuisance regressors in pharmacological imaging research. Yet, although practice effects are well acknowledged in the broader neuromodulator and cognitive literature (Bartels et al., 2010; MacRae et al., 1988; Servan-Schreiber et al., 1998), our understanding of these effects is limited. One of the few studies that report on drug administration effects, showed that d-amphetamine (AMPH) driven increases in functional-MRI–based blood oxygen level-dependent (BOLD) signal variability (SD_{BOLD}) and performance depended greatly on drug administration order (Garrett et al., 2015). In this study, only older subjects who received AMPH first improved in performance and SD_{BOLD}. Based on research in rats, demonstrating that dopamine release increases linearly with reward-based lever press practice (Owesson-White et al., 2008), the authors speculate that practice may have shifted participants along an inverted-U-shaped dopamine performance curve (Cools & D'Esposito, 2011) by increasing baseline dopamine release (Garrett et al., 2015). Interestingly, we did not see a modulation of the MA effects by drug session order (PL first vs. MA first). Thus, the inclusion of an orientation session might be a good strategy to control for practice and drug order effects.

Our results also illustrate the large interindividual variability of MA effects. Recently a large pharmacological fMRI/PET study (n=100) presented strong evidence that interindividual differences in striatal dopamine synthesis capacity explain variability in effects of methylphenidate on reversal learning (van den Bosch et al., 2022). They demonstrated that

methylphenidate improved reversal learning performance to a greater degree in participants with higher dopamine synthesis capacity, thus establishing the baseline-dependency principle for methylphenidate. These results are in line with previous research showing that methylphenidate improved reversal learning to a greater degree in participants with higher baseline working memory capacity, an index that is commonly used as an indirect proxy of dopamine synthesis capacity (Rostami Kandroodi et al., 2021; van der Schaaf et al., 2013; van der Schaaf et al., 2014). In the current study, we did not collect working memory capacity related information. However, our result that initial task performance strongly affected the effect of MA is in line with the pattern of results showing that individual baseline differences strongly influence drug effects and thus should be considered in pharmacological studies (Cools & D'Esposito, 2011; Durstewitz & Seamans, 2008; van den Bosch et al., 2022). Indeed, there is evidence from the broader literature on the effects of psychostimulants on cognitive performance, that suggest that stimulants improve performance only in low performers (Ilieva et al., 2013). Consistent with this, there is evidence in rats, that poor baseline performance was associated with greater response to amphetamine and increased performance in signal detection task (Turner et al., 2017).

Our findings can be contrasted to those of Rostami Kandroodi et al. (2021), who examined effects of methylphenidate on a reversal learning task, in relation to baseline differences on a cognitive task. Whereas Rostami Kandroodi et al. (2021) found that the methylphenidate improved performance mainly in participants with higher baseline working memory performance, we found that methamphetamine improved the ability to dynamically adjust learning from prediction errors to a greater extent in participants who performed poorly-to-medium at baseline. There are several possible reasons for these apparently different findings. First, MA and methylphenidate differ in their primary mechanisms of action: MPH acts mainly as a reuptake blocker whereas MA increases synaptic levels of catecholamines by inhibiting the vesicular monoamine transporter 2 (VMAT2) and inhibiting the enzyme monoamine oxidase (MAO). These differences in action could account for differential effects on cognitive tasks. Second, the tasks used by Rostami Kandroodi et al. (2021) and the present study differ in several ways. The Rostami Kandroodi et al. (2021) task assessed responses to a single reversal event during the session whereas the present study used repeated reversals with probabilistic outcomes. Third, the

measures of baseline function differed in the two studies: Rostami Kandroodi et al. (2021) used a working memory task that was not used in the drug sessions, whereas we used the probabilistic learning task as both the baseline measure and the measure of drug effects. Further research is needed to determine which of these factors influenced the outcomes.

Conclusion

The current data provide evidence that relative to placebo, methamphetamine facilitates the ability to dynamically adjust learning from prediction errors. This observation was seen to a greater degree in those participants who performed moderately low at baseline. These results advance existing literature by presenting evidence for a causal link between catecholaminergic modulation and learning flexibility and further highlights a baseline-dependency principle for catecholaminergic modulation.

Materials and methods

Design. The results presented here were obtained from the first two sessions of a larger four-session study (clinicaltrials.gov ID number NCT04642820). During the latter two sessions of the larger study, not reported here, participants participated in two fMRI scans. During the two 4-h laboratory sessions presented here, healthy adults received methamphetamine (20 mg oral; MA) or placebo (PL), in mixed order under double-blind conditions. One hour after ingesting the capsule they completed the 30-min reinforcement reversal learning task. The primary comparisons were on acquisition and reversal learning parameters of reinforcement learning after MA vs PL. Secondary measures included subjective and cardiovascular responses to the drug.

Drug sessions. The two drug sessions were conducted in a comfortable laboratory environment, from 9 am to 1 pm, at least 72 hours apart. Upon arrival, participants provided breath and urine samples to test for recent alcohol or drug use and pregnancy (CLIAwaived Inc,Carlsbad, CAAlcosensor III, Intoximeters; AimStickPBD, hCG professional, Craig Medical Distribution). Positive tests lead to rescheduling or dismissal from the study. After drug testing, subjects

completed baseline mood measures, and heart rate and blood pressure were measured. At 9:30 am they ingested capsules (PL or MA 20 mg, in color-coded capsules) under double-blind conditions. Oral MA (Desoxyn, 5 mg per tablet) was placed in opaque size 00 capsules with dextrose filler. PL capsules contained only dextrose. Subjects completed the reinforcement learning task 60 minutes after capsule ingestion. Drug effects questionnaires were obtained at multiple intervals during the session. They completed other cognitive tasks not reported here. Participants were tested individually and were permitted to relax, read or watch neutral movies when they were not completing study measures.

Orientation session. Participants attended an initial orientation session to provide informed consent, and to complete personality questionnaires. They were told that the purpose of the study was to investigate the effects of psychoactive drugs on mood, brain, and behavior. To reduce expectancies, they were told that they might receive a placebo, stimulant, or sedative/tranquilizer. However, participants only received methamphetamine and placebo. They agreed not to use any drugs except for their normal amounts of caffeine for 24 hours before and 6 hours following each session. Women who were not on oral contraceptives were tested only during the follicular phase (1-12 days from menstruation) because responses to stimulant drugs are dampened during the luteal phase of the cycle (White et al., 2002). Most participants (N=97 out of 113) completed the reinforcement learning task during the orientation session as a baseline measurement. This measure was added after the study began. Participants who did not complete the baseline measurement were omitted from the analyses presented in the main text. We run the key analyses on the full sample (n=109). This sample included participants who completed the task only on the drug sessions. When controlling for session order and number (two vs. three sessions) effects, we see no drug effect on overall performance and learning. Yet, we found that eta was also reduced under MA in the full sample, which also resulted in reduced variability in the learning rate (see supplementary results for more details).

Drug sessions. The two drug sessions were conducted in a comfortable laboratory environment, from 9 am to 1 pm, at least 72 hours apart. Upon arrival, participants provided breath and urine

samples to test for recent alcohol or drug use and pregnancy (CLIAwaived Inc,Carlsbad, CAAlcosensor III, Intoximeters; AimStickPBD, hCG professional, Craig Medical Distribution). Positive tests lead to rescheduling or dismissal from the study. After drug testing, subjects completed baseline mood measures, and heart rate and blood pressure were measured. At 9:30 am they ingested capsules (PL or MA 20 mg, in color-coded capsules) under double-blind conditions. Oral MA (Desoxyn, 5 mg per tablet) was placed in opaque size 00 capsules with dextrose filler. PL capsules contained only dextrose. Subjects completed the reinforcement learning task 60 minutes after capsule ingestion. Drug effects questionnaires were obtained at multiple intervals during the session. They completed four other cognitive tasks not reported here. Participants were tested individually and were permitted to relax, read or watch neutral movies when they were not completing study measures.

Dependent measures.

Reinforcement Learning Task. Participants performed a reversal variant of an established probabilistic learning task (Fischer & Ullsperger, 2013; Jocham et al., 2014; Kirschner et al., 2022; Kirschner et al., 2023). On each trial participants were presented one of three different stimuli and decided to either gamble or avoid gambling with that stimulus with the goal to maximize the final reward (see Figure 1A). A gamble resulted in winning or losing points, depending on reward contingencies associated with the particular stimulus. If participants decided not to gamble, they avoided any consequences, but were still able to observe what would have happened if they had gambled by receiving counterfactual feedback. The three stimuli – white line drawings of animals on black background - were presented in a pseudo random series that was the same for all participants. Reward contingencies for every stimulus could be 20%, 30%, 50%, 70%, or 80% and stayed constant within one block of 30-35 trials. After every block, reward contingency changed without notice. The experiment consisted of 7 blocks per stimulus, leading to 18 reversals and 714 trials in total (see supplementary results for additional information on time-on-task effects). Presentation 22.0 (Neurobehavioral Systems) was used for task presentation. Every trial of the task began with a central fixation cross, presented for a variable time between 300 and 500 ms. After fixation, the stimulus was presented together with the two choice alternatives (a green

checkmark for choosing and a red no-go sign for avoiding, sides counterbalanced across subjects) for a maximum of 2000 ms or until a response was given. If participants failed to respond in time, a question mark was shown and the trial was repeated at the end of the block. When a response was made, the stimulus stayed on screen and feedback was given after 500 ms. The outcome was then presented for 750 ms depending on the subject's choice. Choosing to gamble led to either a green smiley face and a reward of 10 points or a red frowning face and a loss of 10 points according to the reward probability of the stimulus. An avoided gamble had no monetary consequences: the outcome was always 0. Counterfactual/fictive outcomes, indicating what would have happened had the participant chosen to gamble, were shown on screen using the same smileys in a paler color, but the reward or punishment was crossed out to indicate that the outcome was fictive.

Drug Effects Questionnaire (DEQ) (Morean et al., 2013) The DEQ consists of 5 questions in total. In this paper we only reported the ratings of the "Do you feel any drug effect?" question which was rated on a 100 mm visual analog scale. Participants completed this at regular intervals throughout the session.

Reinforcement learning model fitting.

We fit variants of reinforcement learning models to participants' choice behavior using a constrained search algorithm (fmincon in MATLAB 2021b), which computed a set of parameters that maximized the total log posterior probability of choice behavior. The base model (M1) was a standard Q-learning model with three parameters: (1) an inverse temperature parameter of the softmax function used to convert trial expected values to action probabilities, (2) a play bias term that indicates a tendency to attribute higher value to gambling behavior (Jang et al., 2019), and (3) an intercept term for the effect of learning rate on choice behavior (Jang et al., 2019). On each trial the expected value (Q_t) of a stimulus (X_t) was calculated according to the following formula:

$$Q_{t+1}(X_t) = Q_t(X_t) + \alpha * \delta_t \; with \; \delta_t = R_t - Q_t(X_t)$$

Q values represent the expected value of an action at trial t. α reflects the learning rate. δ_t represents the prediction error with R_t being the reward magnitude of that trial. On each trial, this value term was transferred into a "biased" value term ($V_B(X_t) = B_{play} + Q_t(X_t)$, where B_{play} is the play bias term) and converted into action probabilities (P(play$|V_{B\,play}(t)(X_t)$; P(pass$|V_{B\,pass}(t)(X_t)$) using a softmax function with an inverse temperature (β):

$$\mathrm{P(play}|V_{B\,play}(t)(X_t)) = \frac{\exp\left(V_{\mathrm{B\,play}}(t)(X_t)\cdot\beta\right)}{\exp\left(V_{\mathrm{play}}(t)(X_t)\cdot\beta\right)+\exp\left(V_{\mathrm{pass}}(t)(X_t)\cdot\beta\right)}$$

$$\mathrm{P(pass}|V_{B\,pass}(t)(X_t)) = \frac{\exp\left(V_{\mathrm{pass}}(t)(X_t)\cdot\beta\right)}{\exp\left(V_{\mathrm{play}}(t)(X_t)\cdot\beta\right)+\exp\left(V_{\mathrm{pass}}(t)(X_t)\cdot\beta\right)}$$

This was our base model (M1). Next, we fit further reinforcement models by complementing the base model with additional parameters. These additional parameters controlled trial-by-trial modulations of the learning rate. Note that our base model treats the learning rate for value updates as a constant. However, previous studies have been shown that people are able to adjust their learning rate according to the volatility of the environment (Behrens et al., 2007; Nassar et al., 2010). In the Pearce-Hall hybrid model, adjustments in learning rate are afforded by weighting the learning rate as a function of the absolute value of previous prediction error (Li et al., 2011). This associability-gated learning mechanism is empirically well supported (Le Pelley, 2004) and facilitates decreasing learning rates in periods of stability and increasing learning rates in periods of change. Previous work has shown that the hybrid model can approximate normative learning rate adjustments (Li et al., 2011; Piray et al., 2019). In this hybrid model, the learning rate is updated as follows:

$$\alpha_t = \kappa A_t$$

$$A_{t+1}(X_t) = \eta * |\delta_t| + (1-\eta) * A_t(X_t)$$

Here, κ is scale of learning rate (α_t) and η determines the step size for updating associability (A_t) as a function of the absolute RPE ($|\delta_t|$). On each trial, the learning rate (α_t) depends on the absolute RPE from past trial. Note that the initial learning rate is defined by κ, whereby κ is determined by a logistic function of a weighted predictor matrix that could include an intercept term (Pearce-Hall hybrid model (M2)) and task variables that may additionally affected trial-by-trial learning rate adjustments. In the Pearce-Hall hybrid feedback confirmatory model (M3) the predictor matrix included an intercept term and feedback confirmatory information (i.e., was the feedback on a given trial confirmatory (factual wins and counterfactual losses) or disconfirmatory (factual losses and counterfactual wins)). Finally, in the Pearce-Hall hybrid feedback confirmatory and modality model (M4) the predictor matrix included an intercept term, feedback confirmatory information and feedback modality (factual vs. counterfactual feedback) information. Two additional learning rate terms—feedback confirmation and modality—were added to the model set, as these factors have been shown to influence learning in similar tasks (Kirschner et al., 2023; Schüller et al., 2020). The best-fitting model was determined by computing the Bayesian Information Criterion (BIC) for each model (Schwarz, 1978). Moreover, we computed protected exceedance probabilities, which gives the probability that one model was more likely than any other model of the model space (Rigoux et al., 2014). To compare participant behavior to model-predicted behavior, we simulated choice behavior using the best fitting model (Pearce-Hall hybrid feedback confirmatory model; see Figure 3A). For each trial, we used the expected trial value ($Q_t(X_t)$) computed above, and the parameter estimates of the temperature variable as inputs to a softmax function to generate choices. Validation of model selection and parameter recovery is reported in the supplementary materials (Figure S1).

Description of the joint estimation of stochasticity and volatility model

In an exploratory analysis, we fit two models to our task structure using modified code provided by Piray and Daw (2021) the following description of the model is adapted therefrom.

In this model, the outcome on trial t, denoted as o_t, is determined by three latent variables: reward probability rate, stochasticity, and volatility. The reward rate on trial t, represented as x_t, follows a Markov process with the following dynamics:

$$x_t = x_{t-1} + \epsilon_t$$

where ϵ_t is Gaussian noise with zero mean and a variance governed by volatility. Consequently, the probability distribution of $x_t$given x_{t-1} and volatility v_t is:

$$p(x_t \mid x_{t-1}, v_t) = \mathcal{N}(x_t \mid x_{t-1}, v_t)$$

where v_t represents the volatility. For analytical convenience, the inverse volatility is defined as:

$$z_t = v_t^{-1}$$

a formulation commonly used in previous studies due to its computational advantages (Piray & Daw, 2020). Outcome values are generated based on the reward rate and stochasticity, following a Gaussian distribution:

$$p(o_t \mid x_t, s_t) = \mathcal{N}(o_t \mid x_t, s_t)$$

where s_t represents stochasticity, and the inverse stochasticity is denoted as:

$$y_t = s_t^{-1}$$

For volatility and stochasticity, a multiplicative noise model was applied to their inverse, an approach previously shown to facilitate analytical inference when considered in isolation (Gamerman et al., 2013; West, 1987). Specifically, the dynamics governing these variables follow:

$$z_t = \eta_v^{-1} z_{t-1} \epsilon_t$$

where 0 <η_v< 1 is a constant, and ϵ_t is a random variable within the unit range, following a Beta distribution:

$$p(\epsilon_t) = B(\epsilon_t \mid j, 0.5\eta_v(1-\eta_v)^{-1}, 0.5)$$

The conditional expectation of z_t is given by z_{t-1}, since $E(\epsilon_t) = \eta_v$. A similar and independent dynamic is assumed for y, parameterized by the constant η_s, such that:

$$y_t = \eta_s^{-1} y_{t-1} \varepsilon_t$$

where ε_t follows a distribution similar to ε_t, parameterized by η_s.

In this implementation, the model was parameterized using $\lambda_v = 1 - \eta_v$ and $\lambda_s = 1 - \eta_s$, respectively. These parameters are interpreted as the update rate for volatility and stochasticity. In other words, larger values of λ_v and λ_s lead to faster updates of volatility and stochasticity. Intuitively, a smaller λ_v increases the mean of ε_t, resulting in a larger update of z_t. Since volatility is the inverse of z_t, a smaller λ_v consequently leads to a slower update of volatility. This relationship has been formally demonstrated in recent work (Piray & Daw, 2020). In addition to these two parameters, the generative process depends on the initial values of volatility and stochasticity, denoted as v_0 and s_0. For our simulations, we assumed v_0 = 0.1 and s_0 = 0.1. For inference, a Rao-Blackwellized Particle Filtering approach (Murphy & Russell, 2001) was employed, where inference about v_t and s_t was performed using a particle filter (Doucet & Johansen, 2011), and conditional on these, inference over x_t was carried out using the Kalman filter.

The Kalman filter maintains beliefs about the reward probability at each trial in the form of a Gaussian distribution, characterized by a mean m_t and a variance w_t, which represents uncertainty about the true value. On each trial, the update is driven by a prediction error signal δ_t and a learning rate α_t This results in the following update rules upon observing the outcome o_t:

$$\delta_t = o_t - m_t$$

$$\alpha_t = \frac{w_t + v_t}{w_t + v_t + s_t}$$

$$m_{t+1} = m_t + \alpha_t \delta_t$$

$$w_{t+1} = (1 - \alpha_t)(w_t + v_t)$$

These equations describe how the belief about the reward rate is adjusted dynamically based on observed outcomes and the associated uncertainty.

The particle filter is a Monte Carlo sequential importance sampling method that maintains a set of particles (i.e., samples). The algorithm consists of three steps per trial. First, in the prediction step, each particle transitions to the next state based on the generative process. Second, the weight of each particle is updated based on the probability of the observed outcome:

$$b_t^l \propto N\left(o_t \mid m_{t-1}^l, w_{t-1}^l + v_t^l + s_t^l \right)$$

where b_t^l is the weight of particle l on trial t, and m_{t-1}^l and w_{t-1}^l) are the estimated mean and variance from the Kalman filter on the previous trial. The terms v_t^l and s_t^l represent the sampled values of volatility and stochasticity (i.e., the inverse of z_t^l and y_t^l). During this step, particles are resampled using the systematic resampling procedure if the ratio of effective to total particles falls below 0.5. In the third step, the Kalman filter updates the mean and variance. Specifically, for every particle, Equations of the Kalman filter are used to define alpha_t^l and update m_t^l and w_t^l. The learning rate and estimated reward probability on trial t are then computed as the weighted average of all particles, with the weights given by b_t^l.

In the simulations the reward probability followed those of our task with outcome varience set to .01 for low outcome nosie (20/80% reward probability), .02 for high outcome noise (30/70%

reward probability) and .05 for random reward probabilites. The stochasticity for the lesioned models the simulations were assumed to be 0.001. For simulating choice, we used the softmax with a decision noise of 3.

Data analysis.

We analyzed drug effects on behavioral performance, and model parameters using paired *t* tests. Given the effects of initial performance and practice in pharmacological imaging research (Garrett et al., 2015), we additionally stratified MA effects by task performance in the orientation using median split. These data were analyzed using a two-way repeated-measures ANOVA with the factors Drug (two levels) and Baseline Performance (two levels). Paired *t* tests were used as post hoc tests. Moreover, we investigated reversal learning by calculating learning curves. Post hoc, we observed that drug effects on learning became only apparent in the second phase of learning. We therefore used the Bai-Perrin multiple break point test (Bai & Perron, 2003) to identify the number and location of structural breaks in the learning curves. In broad strokes, the test detects whether breaks in a curve exists, and if so, how many there are, based on the regression slope in predefined segments (here, we set the segment length to 5 trials). In our case, the test could reveal between 0 and 5 breaks (number of trials / segment length - 1). We run this test using data from all subjects and all sessions. The test detected one break that cut the learning curves into two segments (see results). We then calculated an index of learning performance after reversals by averaging the number of correct choices over the second learning phase. The index was then subjected to a two-way repeated ANOVA with the factors Drug (two levels) and Baseline Performance (two levels).

Data Availability Statement

All raw data and analysis scripts can be accessed at:
https://github.com/HansKirschner/REFIT_Chicago_public.git

Acknowledgements

We thank all our participants who took part in this research for the generosity of their time and commitment. This research was supported by the National Institute on Drug Abuse DA02812. HM was supported by the National Institutes of Health T32 GM07019. MU was supported by the Deutsche Forschungsgemeinschaft, Grant/Award Number: SFB 1436; and the European Research Council, Grant/Award Number: 101018805".

Competing interests

HdW is on the Board of Directors of PharmAla Biotech, and on scientific advisory committees of Gilgamesh Pharmaceuticals and MIND Foundation. These activities are unrelated to the present study. The other authors report no competing interests.

References

Arnsten, A. F., & Pliszka, S. R. (2011). Catecholamine influences on prefrontal cortical function: relevance to treatment of attention deficit/hyperactivity disorder and related disorders. *Pharmacol Biochem Behav*, *99*(2), 211-216. https://doi.org/10.1016/j.pbb.2011.01.020

Arria, A. M., Caldeira, K. M., Vincent, K. B., O'Grady, K. E., Cimini, M. D., Geisner, I. M., Fossos-Wong, N., Kilmer, J. R., & Larimer, M. E. (2017). Do college students improve their grades by using prescription stimulants nonmedically? *Addict Behav*, *65*, 245-249. https://doi.org/10.1016/j.addbeh.2016.07.016

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI. *Cereb Cortex*, *22*(3), 527-536. https://doi.org/10.1093/cercor/bhr117

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1-22. https://doi.org/https://doi.org/10.1002/jae.659

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215-220. https://doi.org/10.1093/ije/dyh299

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci*, *11*, 118. https://doi.org/10.1186/1471-2202-11-118

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci*, *10*(9), 1214-1221. https://doi.org/10.1038/nn1954

Bowman, E., Coghill, D., Murawski, C., & Bossaerts, P. (2023). Not so smart? "Smart" drugs increase the level but decrease the quality of cognitive effort. *Sci Adv*, *9*(24), eadd4165. https://doi.org/10.1126/sciadv.add4165

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol Rev*, *99*(1), 45-77. https://doi.org/10.1037/0033-295x.99.1.45

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol Rev*, *120*(1), 190-229. https://doi.org/10.1037/a0030852

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024-1035. https://doi.org/https://doi.org/10.1111/j.1460-9568.2011.07980.x

Collins, A. G. E., & Frank, M. J. (2016). Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition*, *152*, 160-169. https://doi.org/10.1016/j.cognition.2016.04.002

Collins, A. G. E., & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proc Natl Acad Sci U S A*, *115*(10), 2502-2507. https://doi.org/10.1073/pnas.1720963115

Cook, J. L., Swart, J. C., Froböse, M. I., Diaconescu, A. O., Geurts, D. E., den Ouden, H. E., & Cools, R. (2019). Catecholaminergic modulation of meta-learning. *Elife*, *8*. https://doi.org/10.7554/eLife.51439

Cools, R. (2008). Role of Dopamine in the Motivational and Cognitive Control of Behavior. *The Neuroscientist*, *14*(4), 381-395. https://doi.org/10.1177/1073858408317009

Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol Psychiatry*, *69*(12), e113-125. https://doi.org/10.1016/j.biopsych.2011.03.028

Diederen, K. M., Spencer, T., Vestergaard, M. D., Fletcher, P. C., & Schultz, W. (2016). Adaptive Prediction Error Coding in the Human Midbrain and Striatum Facilitates Behavioral

Adaptation and Learning Efficiency. *Neuron*, *90*(5), 1127-1138. https://doi.org/10.1016/j.neuron.2016.04.019

Dodds, C. M., Müller, U., Clark, L., van Loon, A., Cools, R., & Robbins, T. W. (2008). Methylphenidate has differential effects on blood oxygenation level-dependent signal related to cognitive subprocesses of reversal learning. *J Neurosci*, *28*(23), 5976-5982. https://doi.org/10.1523/jneurosci.1153-08.2008

Doucet, A., & Johansen, A. M. (2011). *A tutorial on particle filtering and smoothing: fifteen years later*. Oxford University Press.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Netw*, *15*(4-6), 495-506. https://doi.org/10.1016/s0893-6080(02)00044-8

Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K.-P., & Brocke, B. (2005). Dopamine and Cognitive Control: The Influence of Spontaneous Eyeblink Rate and Dopamine Gene Polymorphisms on Perseveration and Distractibility. *Behavioral Neuroscience*, *119*(2), 483-490. https://doi.org/10.1037/0735-7044.119.2.483

Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biol Psychiatry*, *64*(9), 739-749. https://doi.org/10.1016/j.biopsych.2008.05.015

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2012). In the Zone or Zoning Out? Tracking Behavioral and Neural Fluctuations During Sustained Attention. *Cerebral Cortex*, *23*(11), 2712-2723. https://doi.org/10.1093/cercor/bhs261

Fallon, S. J., van der Schaaf, M. E., ter Huurne, N., & Cools, R. (2017). The Neurocognitive Cost of Enhancing Cognition with Methylphenidate: Improved Distractor Resistance but Impaired Updating. *Journal of Cognitive Neuroscience*, *29*(4), 652-663. https://doi.org/10.1162/jocn_a_01065

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243-1255. https://doi.org/10.1016/j.neuron.2013.07.006

Floresco, S. B. (2013). Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions [Review]. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00062

Gamerman, D., dos Santos, T. R., & Franco, G. C. (2013). A NON-GAUSSIAN FAMILY OF STATE-SPACE MODELS WITH EXACT MARGINAL LIKELIHOOD. *Journal of Time Series Analysis*, *34*(6), 625-645. https://doi.org/https://doi.org/10.1111/jtsa.12039

Garrett, D. D., Nagel, I. E., Preuschhof, C., Burzynska, A. Z., Marchner, J., Wiegert, S., Jungehülsing, G. J., Nyberg, L., Villringer, A., Li, S. C., Heekeren, H. R., Bäckman, L., & Lindenberger, U. (2015). Amphetamine modulates brain signal variability and working memory in younger and older adults. *Proc Natl Acad Sci U S A*, *112*(24), 7593-7598. https://doi.org/10.1073/pnas.1504090112

Goschke, T. (2013). Volition in Action: Intentions, Control Dilemmas, and the Dynamic Regulation of Cognitive Control. In W. Prinz, M. Beisert, & A. Herwig (Eds.), *Action Science: Foundations of an Emerging Discipline* (pp. 0). The MIT Press. https://doi.org/10.7551/mitpress/9780262018555.003.0024

Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: the role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia*, *62*, 403-423. https://doi.org/10.1016/j.neuropsychologia.2014.07.015

Goschke, T., & Bolte, A. (2018). A dynamic perspective on intention, conflict, and volition: Adaptive regulation and emotional modulation of cognitive control dilemmas. In *Why people do the things they do: Building on Julius Kuhl's contributions to the psychology of motivation and volition.* (pp. 111-129). Hogrefe. https://doi.org/10.1027/00540-000

Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2022). Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, *32*(3), 586-599.e587. https://doi.org/https://doi.org/10.1016/j.cub.2021.12.006

Husain, M., & Mehta, M. A. (2011). Cognitive enhancement by drugs in health and disease. *Trends Cogn Sci*, *15*(1), 28-36. https://doi.org/10.1016/j.tics.2010.11.002

Iigaya, K., Fonseca, M. S., Murakami, M., Mainen, Z. F., & Dayan, P. (2018). An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature Communications*, *9*(1), 2477. https://doi.org/10.1038/s41467-018-04840-2

Ilieva, I., Boland, J., & Farah, M. J. (2013). Objective and subjective cognitive enhancing effects of mixed amphetamine salts in healthy people. *Neuropharmacology*, *64*, 496-505. https://doi.org/10.1016/j.neuropharm.2012.07.021

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), 719-732. https://doi.org/10.1038/s41562-019-0597-3

Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. *PLoS One*, *15*(2), e0229345. https://doi.org/10.1371/journal.pone.0229345

Jocham, G., Klein, T. A., & Ullsperger, M. (2014). Differential modulation of reinforcement learning by D2 dopamine and NMDA glutamate receptor antagonism. *J Neurosci*, *34*(39), 13151-13162. https://doi.org/10.1523/jneurosci.0757-14.2014

Karamacoska, D., Barry, R. J., & Steiner, G. Z. (2018). Electrophysiological underpinnings of response variability in the Go/NoGo task. *International Journal of Psychophysiology*, *134*, 159-167. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2018.09.008

Kirschner, H., Fischer, A. G., & Ullsperger, M. (2022). Feedback-related EEG dynamics separately reflect decision parameters, biases, and future choices. *Neuroimage*, *259*, 119437. https://doi.org/10.1016/j.neuroimage.2022.119437

Kirschner, H., Nassar, M. R., Fischer, A. G., Frodl, T., Meyer-Lotz, G., Froböse, S., Seidenbecher, S., Klein, T. A., & Ullsperger, M. (2023). Transdiagnostic inflexible learning dynamics explain deficits in depression and schizophrenia. *Brain*, *147*(1), 201-214. https://doi.org/10.1093/brain/awad362

Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B., & Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat Neurosci*, *19*(10), 1280-1285. https://doi.org/10.1038/nn.4382

Le Pelley, M. E. (2004). The Role of Associative History in Models of Associative Learning: A Selective Review and a Hybrid Model. *The Quarterly Journal of Experimental Psychology Section B*, *57*(3b), 193-243. https://doi.org/10.1080/02724990344000141

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci*, *14*(10), 1250-1252. https://doi.org/10.1038/nn.2904

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS Comput Biol*, *14*(4), e1006043. https://doi.org/10.1371/journal.pcbi.1006043

MacRae, P. G., Spirduso, W. W., & Wilcox, R. E. (1988). Reaction time and nigrostriatal dopamine function: the effects of age and practice. *Brain Res*, *451*(1-2), 139-146. https://doi.org/10.1016/0006-8993(88)90758-5

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190. https://doi.org/https://doi.org/10.1016/j.jneumeth.2007.03.024

Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., & Bestmann, S. (2016). Pharmacological Fingerprints of Contextual Uncertainty. *PLoS Biol*, *14*(11), e1002575. https://doi.org/10.1371/journal.pbio.1002575

Meder, D., Kolling, N., Verhagen, L., Wittmann, M. K., Scholl, J., Madsen, K. H., Hulme, O. J., Behrens, T. E. J., & Rushworth, M. F. S. (2017). Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nature Communications*, *8*(1), 1942. https://doi.org/10.1038/s41467-017-02169-w

Morean, M. E., de Wit, H., King, A. C., Sofuoglu, M., Rueger, S. Y., & O'Malley, S. S. (2013). The drug effects questionnaire: psychometric support across three drug types. *Psychopharmacology (Berl)*, *227*(1), 177-192. https://doi.org/10.1007/s00213-012-2954-z

Muller, T. H., Mars, R. B., Behrens, T. E., & O'Reilly, J. X. (2019). Control of entropy in neural models of environmental state. *Elife*, *8*. https://doi.org/10.7554/eLife.39404

Murphy, K., & Russell, S. (2001). Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Sequential Monte Carlo methods in practice* (pp. 499-515). Springer.

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *Elife*, *8*. https://doi.org/10.7554/eLife.46975

Nassar, M. R., Bruckner, R., Gold, J. I., Li, S. C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nat Commun*, *7*, 11609. https://doi.org/10.1038/ncomms11609

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci*, *15*(7), 1040-1046. https://doi.org/10.1038/nn.3130

Nassar, M. R., Waltz, J. A., Albrecht, M. A., Gold, J. M., & Frank, M. J. (2021). All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain*, *144*(3), 1013-1029. https://doi.org/10.1093/brain/awaa453

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci*, *30*(37), 12366-12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Owesson-White, C. A., Cheer, J. F., Beyene, M., Carelli, R. M., & Wightman, R. M. (2008). Dynamic changes in accumbens dopamine correlate with learning during intracranial self-stimulation. *Proc Natl Acad Sci U S A*, *105*(33), 11957-11962. https://doi.org/10.1073/pnas.0803896105

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*(7028), 873-876. https://doi.org/10.1038/nature03287

Piray, P., & Daw, N. D. (2020). A simple model for learning in volatile environments. *PLoS Comput Biol*, *16*(7), e1007963. https://doi.org/10.1371/journal.pcbi.1007963

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587. https://doi.org/10.1038/s41467-021-26731-9

Piray, P., & Daw, N. D. (2024). Computational processes of simultaneous learning of stochasticity and volatility in humans. *Nat Commun*, *15*(1), 9073. https://doi.org/10.1038/s41467-024-53459-z

Piray, P., Ly, V., Roelofs, K., Cools, R., & Toni, I. (2019). Emotionally Aversive Cues Suppress Neural Systems Underlying Optimal Learning in Socially Anxious Individuals. *J Neurosci*, *39*(8), 1445-1456. https://doi.org/10.1523/jneurosci.1394-18.2018

Preuschoff, K., t Hart, B. M., & Einhäuser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Front Neurosci*, *5*, 115. https://doi.org/10.3389/fnins.2011.00115

Prince, J. (2008). Catecholamine dysfunction in attention-deficit/hyperactivity disorder: an update. *J Clin Psychopharmacol*, *28*(3 Suppl 2), S39-45. https://doi.org/10.1097/JCP.0b013e318174f92a

Razmi, N., & Nassar, M. R. (2022). Adaptive Learning through Temporal Dynamics of State Representation. *J Neurosci*, *42*(12), 2524-2538. https://doi.org/10.1523/jneurosci.0387-21.2022

Repantis, D., Schlattmann, P., Laisney, O., & Heuser, I. (2010). Modafinil and methylphenidate for neuroenhancement in healthy individuals: A systematic review. *Pharmacol Res*, *62*(3), 187-206. https://doi.org/10.1016/j.phrs.2010.04.002

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. *Neuroimage*, *84*, 971-985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Rostami Kandroodi, M., Cook, J. L., Swart, J. C., Froböse, M. I., Geurts, D. E. M., Vahabie, A. H., Nili Ahmadabadi, M., Cools, R., & den Ouden, H. E. M. (2021). Effects of methylphenidate on reinforcement learning depend on working memory capacity.

Psychopharmacology (Berl), *238*(12), 3569-3584. https://doi.org/10.1007/s00213-021-05974-w

Schüller, T., Fischer, A. G., Gruendler, T. O. J., Baldermann, J. C., Huys, D., Ullsperger, M., & Kuhn, J. (2020). Decreased transfer of value to action in Tourette syndrome. *Cortex*, *126*, 39-48. https://doi.org/10.1016/j.cortex.2019.12.027

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599. https://doi.org/10.1126/science.275.5306.1593

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461-464. http://www.jstor.org/stable/2958889

Schweighofer, N., & Doya, K. (2003). Meta-learning in reinforcement learning. *Neural Netw*, *16*(1), 5-9. https://doi.org/10.1016/s0893-6080(02)00228-9

Servan-Schreiber, D., Carter, C. S., Bruno, R. M., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part II. D-amphetamine effects in human subjects performing a selective attention task. *Biol Psychiatry*, *43*(10), 723-729. https://doi.org/10.1016/s0006-3223(97)00449-6

Set, E., Saez, I., Zhu, L., Houser, D. E., Myung, N., Zhong, S., Ebstein, R. P., Chew, S. H., & Hsu, M. (2014). Dissociable contribution of prefrontal and striatal dopaminergic genes to learning in economic games. *Proc Natl Acad Sci U S A*, *111*(26), 9615-9620. https://doi.org/10.1073/pnas.1316259111

Silvetti, M., Vassena, E., Abrahamse, E., & Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Comput Biol*, *14*(8), e1006370. https://doi.org/10.1371/journal.pcbi.1006370

Smith, M. E., & Farah, M. J. (2011). Are prescription stimulants "smart pills"? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychol Bull*, *137*(5), 717-741. https://doi.org/10.1037/a0023825

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, *46*(4), 1004-1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Turner, K. M., Peak, J., & Burne, T. H. (2017). Baseline-dependent effects of amphetamine on attention are associated with striatal dopamine metabolism. *Sci Rep*, *7*(1), 297. https://doi.org/10.1038/s41598-017-00437-9

van den Bosch, R., Lambregts, B., Määttä, J., Hofmans, L., Papadopetraki, D., Westbrook, A., Verkes, R. J., Booij, J., & Cools, R. (2022). Striatal dopamine dissociates methylphenidate effects on value-based versus surprise-based reversal learning. *Nat Commun*, *13*(1), 4962. https://doi.org/10.1038/s41467-022-32679-1

van der Schaaf, M. E., Fallon, S. J., Ter Huurne, N., Buitelaar, J., & Cools, R. (2013). Working memory capacity predicts effects of methylphenidate on reversal learning. *Neuropsychopharmacology*, *38*(10), 2011-2018. https://doi.org/10.1038/npp.2013.100

van der Schaaf, M. E., van Schouwenburg, M. R., Geurts, D. E., Schellekens, A. F., Buitelaar, J. K., Verkes, R. J., & Cools, R. (2014). Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. *Cereb Cortex*, *24*(3), 633-642. https://doi.org/10.1093/cercor/bhs344

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Franceschi, D., Maynard, L., Ding, Y. S., Gatley, S. J., Gifford, A., Zhu, W., & Swanson, J. M. (2002). Relationship between blockade of dopamine transporters by oral methylphenidate and the increases in extracellular dopamine: therapeutic implications. *Synapse*, *43*(3), 181-187. https://doi.org/10.1002/syn.10038

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, *74*(3), 646-648. https://doi.org/10.1093/biomet/74.3.646

Westbrook, A., van den Bosch, R., Määttä, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science*, *367*(6484), 1362-1366. https://doi.org/10.1126/science.aaz5891

White, T. L., Justice, A. J., & de Wit, H. (2002). Differential subjective effects of D-amphetamine by gender, hormone levels and menstrual cycle phase. *Pharmacol Biochem Behav*, *73*(4), 729-741.

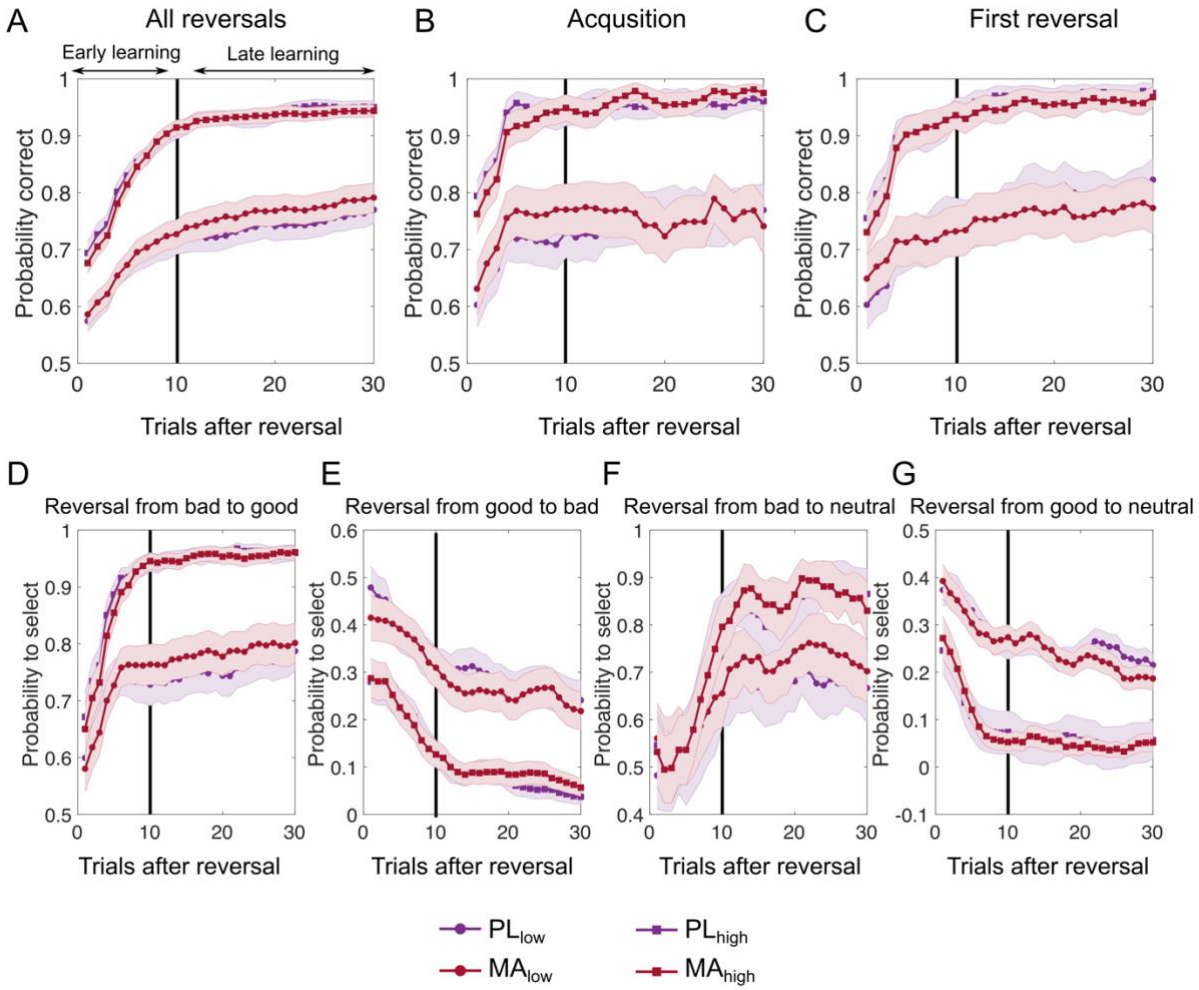
Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Comput Biol*, *9*(7), e1003150. https://doi.org/10.1371/journal.pcbi.1003150

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681-692. https://doi.org/10.1016/j.neuron.2005.04.026

Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neurosci Biobehav Rev*, *128*, 270-281. https://doi.org/10.1016/j.neubiorev.2021.06.024

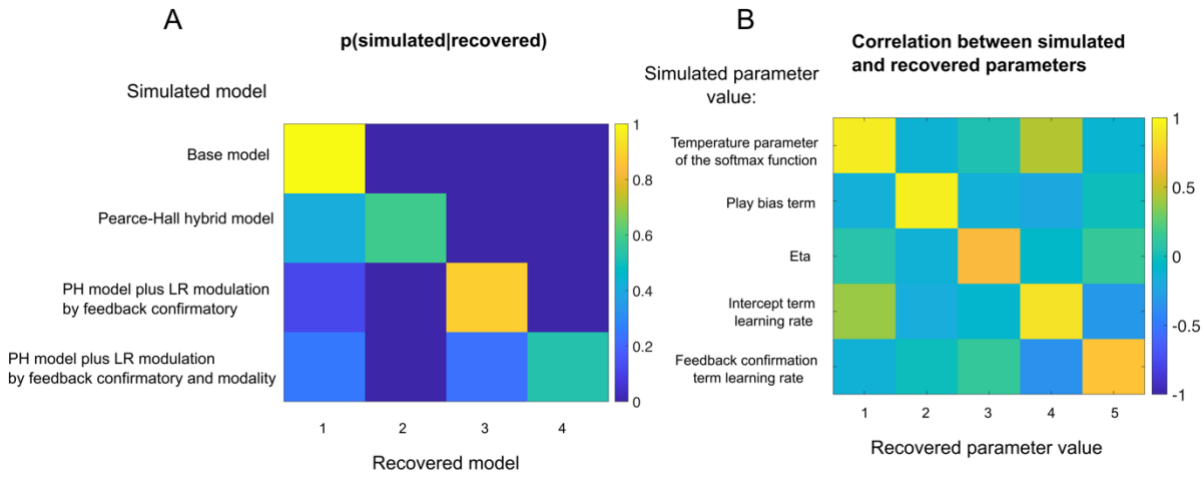
Methamphetamine-induced adaptation of learning rate dynamics depend on baseline performance.

Supplementary Information



Supplementary Figure S1. Learning curves

Top part shows learning curves quantified as the probability to select the correct choice (choosing the advantageous stimuli and avoiding disadvantageous stimuli) stratified by orientation performance. Two-way ANOVAs with the factors Drug (two levels) and Baseline Performance (two levels) on the averaged probability of correct choice during the early and late stage of learning were used to investigate drug effects. **(A)** No differences in the learning curves between MA and PL became evident when considering all reversals (all $p > .1$). **(B)** There was no drug related difference in the acquisition phase of the task between (all $p > .05$) or **(C)** in the first reversal learning (all $p > .1$). In the bottom part of the figure, learning curves are defined as the probability to select a stimulus. **(D)** No drug effect emerged for reversal learning from a bad stimulus to a good stimulus (all $p > .09$) or **(E)** good to bad stimuli (all $p > .09$). Moreover, there was no difference in reversal learning to neutral stimuli **(F and G)**. Note. PL = Placebo; MA = methamphetamine.



Supplementary Figure S2. Validation of model selection and parameter recovery.
After model-fitting, each model was used to simulate data for each participant using the best-fitting parameters for that participant. Each model was fit to each synthetic dataset and BIC was used to determine which model provided best fit to synthetic data. **(A)** Inverse confusion matrix. The frequency with which a recovered model (abscissa, determined by lowest BIC) corresponded to a given simulation model (ordinate) is depicted in color. Recovered models correspond to the same models labeled on the ordinate, with recovered model 1 corresponding to the base model, and so on. The results of the model recover analyses suggest that the recovered model typically corresponded to a synthetic dataset produced by that model. **(B)** Parameter values that were used to simulate data from the hybrid model with additional modulation of the learning rate by feedback confirmatory (ordinate) tended to correlate (color) with the parameter values best fit to those synthetic datasets (abscissa). Recovered parameter values correspond to the labels on the ordinate, with parameter 1 reflecting temperature parameter of the softmax function, and so on.

Supplementary Table 1. Analysis of parameter differences between mathampehtamine (MA) and placebo (PL).

Model	Parameter PL Mean (SD)	Parameter MA Mean (SD)	Inferential statistic	Signed-Rank Test
Model $\Delta_{\text{inverse temperature}}$				
Low performer	1.90 (0.16)	1.83 (0.11)	t(46) = 0.60, p = 0.55	p = 0.899
High performer	3.12 (0.08)	3.06 (0.08)	t(46) = 1.00, p = 0.32	p = 0.088
Model $\Delta_{\text{play bias}}$				
Low performer	0.04 (0.05)	0.02 (0.03)	t(46) = 0.26, p = 0.79	p = 0.168
High performer	-0.03 (0.01)	-0.01 (0.02)	t(46) = -1.03, p = 0.30	p = 0.189
Model $\Delta_{\text{LR intercept}}$				
Low performer	-7.17 (0.58)	-7.24 (0.64)	t(46) = 0.17, p = 0.86	p = 0.611
High performer	-4.86 (0.15)	-4.92 (0.15)	t(46) = 0.47, p = 0.63	p = 0.227
Model Δ_{eta}				
Low performer	0.35 (0.02)	0.28 (0.02)	t(46) = 2.90, p = 0.005	p = 0.006
High performer	0.26 (0.02)	0.24 (0.01)	t(46) = 1.55, p = 0.12	p = 0.114

Here, we we extended the winning model by allowing each parameter, in turn, to be differentially estimated for MA and PL, while keeping the other parameters fixed at the group (low and high baseline performance) mean estimates of the winning model for the placebo session.

Time-on-Task Effects

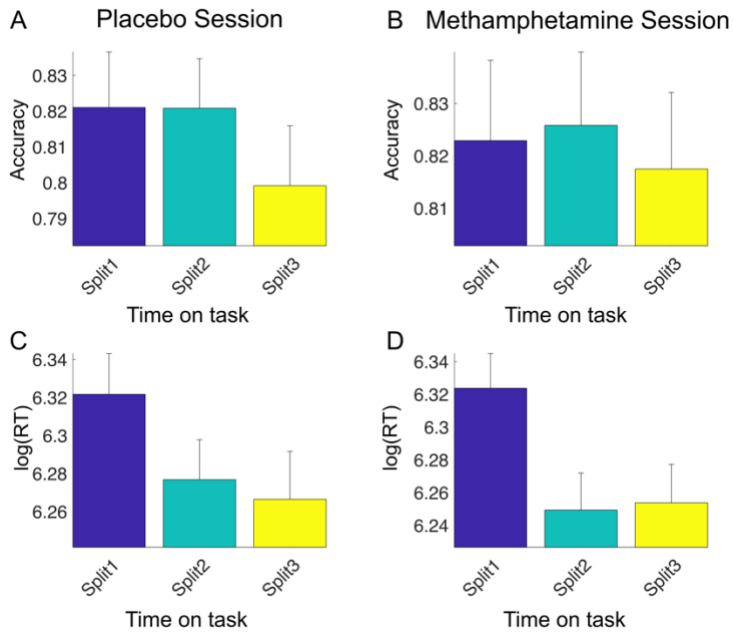
Given the length of our task, we investigated whether fatigue effects or changes in behavior occurred over time. Specifically, we regressed each participant's single-trial log-scaled reaction times (RT) and accuracy (a binary variable reflecting whether participants displayed stimulus-appropriate behavior on each trial) onto trial number, which served as a proxy for time on task. The resulting t-values for the time-on-task regressor were analyzed at the group level using two-sided t-tests against zero and compared across sessions and baseline performance groups. The results of these regression models are presented in Supplementary Table S2, with raw data splits shown in Supplementary Figure S7.

Our findings indicate that choice behavior was not systematically affected over the course of the task. This effect did not differ between low and high baseline performers and was not influenced by the drug. In contrast, reaction times decreased over the course of the task, with this speeding effect being enhanced by MA, particularly in the low-performance group.

Supplementary Table S2. Time-on-Task Effect on choice and reaction time.

Model	ToT Coefficient PL Mean (SD)	ToT Coefficient MA Mean (SD)	Group comparison
Logistic regression on accuracy			
All	-0.05(0.17)	-0.31(0.23)	t(93) = 0.90, p = 0.37
Low performer	-0.18(0.35)	0.41(0.25)	t(46) = 1.61, p = 0.11
High performer	-0.43(0.32)	-0.51(0.22)	t(46) = 0.16, p = 0.86
Regression on log(RT)			
All	-1.77(0.42)*	-3.09 (0.48)*	t(93) = 2.28, p = 0.02
Low performer	-2.11(0.65)*	-4.32(0.65)*	t(46) = -2.59, p = 0.01
High performer	--1.43(0.54)*	-1.85(0.64)*	t(46) = -0.55, p = 0.58

Note: ToT = Time-on-Task; *= average within participant t-values signficantly differe from zero



Supplementary Figure S3. Speeding over the course of the task is enhanced by methamphetamine. This figure shows raw data splits for the effect of time on task (i.e., trial number) on accuracy (panels A & B; a binary variable indicating stimulus-appropriate behavior on each trial) and log-scaled reaction times (RT; panels C & D) across different drug sessions. Results indicate that overall choice accuracy was not affected by time on task. However, participants exhibited faster reaction times over the course of the task, with this speeding effect being more pronounced in the methamphetamine condition.

Linear mixed effects model analyses for key findings

In the manuscript, we used a binary discretization of baseline performance to simplify the analysis and presentation. To parse out the relationship between methamphetamine effects and baseline performance into finer level of detail, we conducted additional linear mixed-effects model (LMM) analyses, focusing on the key findings reported in the manuscript. Specifically, we examined drug effects:

- On overall total points.
- On the probability of correct choice in the late learning phase under high outcome noise.
- On the learning rate parameter eta
- On learning rate variability.
- On choices following double misleading feedback.
- On the signal-to-noise ratio in high outcome noise conditions.

We first fit the following model:

$$\text{Outcome}_i = \beta_0 + \beta_1 \text{SessionOrder}_i + \beta_2 \text{DrugSession}_i + \beta_3 \text{BLPerformance}_i + \beta_4 (\text{DrugSession}_i \times \text{BLPerformance}_i) + u_{s(i)} + \varepsilon_i$$

where:

- $Outcome_i$ is the dependent variable (i.e., total points, P(correct choice) late learing phase high outcome noise; parameter estimate for eta; learning rate variance; P(correct choice) after double misleading outcomes; Signal-to-noise ratio high outcome noise).
- $\beta_0$is the fixed intercept.
- β_1 $\beta_2$are the fixed-effect coefficients.
- Session Order_i (drug first vs. placebo first) is a fixed effect to control for possible session order effects.
- Drug Session_i is a fixed effect for drug session.

- $u_{s(i)} \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept for subject $s(i)$.
- $\varepsilon_{s(i)} \sim \mathcal{N}(0, \sigma_u^2)$ is the residual error term.

To our surprise, none of these models revealed a significant drug x baseline performance interaction. To further investigate this, we plotted the difference scores (i.e., the outcome variable under the drug condition minus placebo condition) against baseline performance scores. To better visualize potential nonlinearities, we applied a sliding window approach (step size of one subject and a window size of 25 subjects based on recent recommendations (Jenkins & Quintana-Ascencio, 2020)). This revealed nonlinear relationships between baseline performance and outcome variables (see Supplementary Figure S6). Specifically, we observed that drug effects were maximal in moderately low baseline performer.

To formally test this pattern, we applied the same sliding window regression approach, fitting a set of LMMs separately for each step size. These models followed the same structure as before but excluded the baseline performance term, allowing us to examine within-window effects:

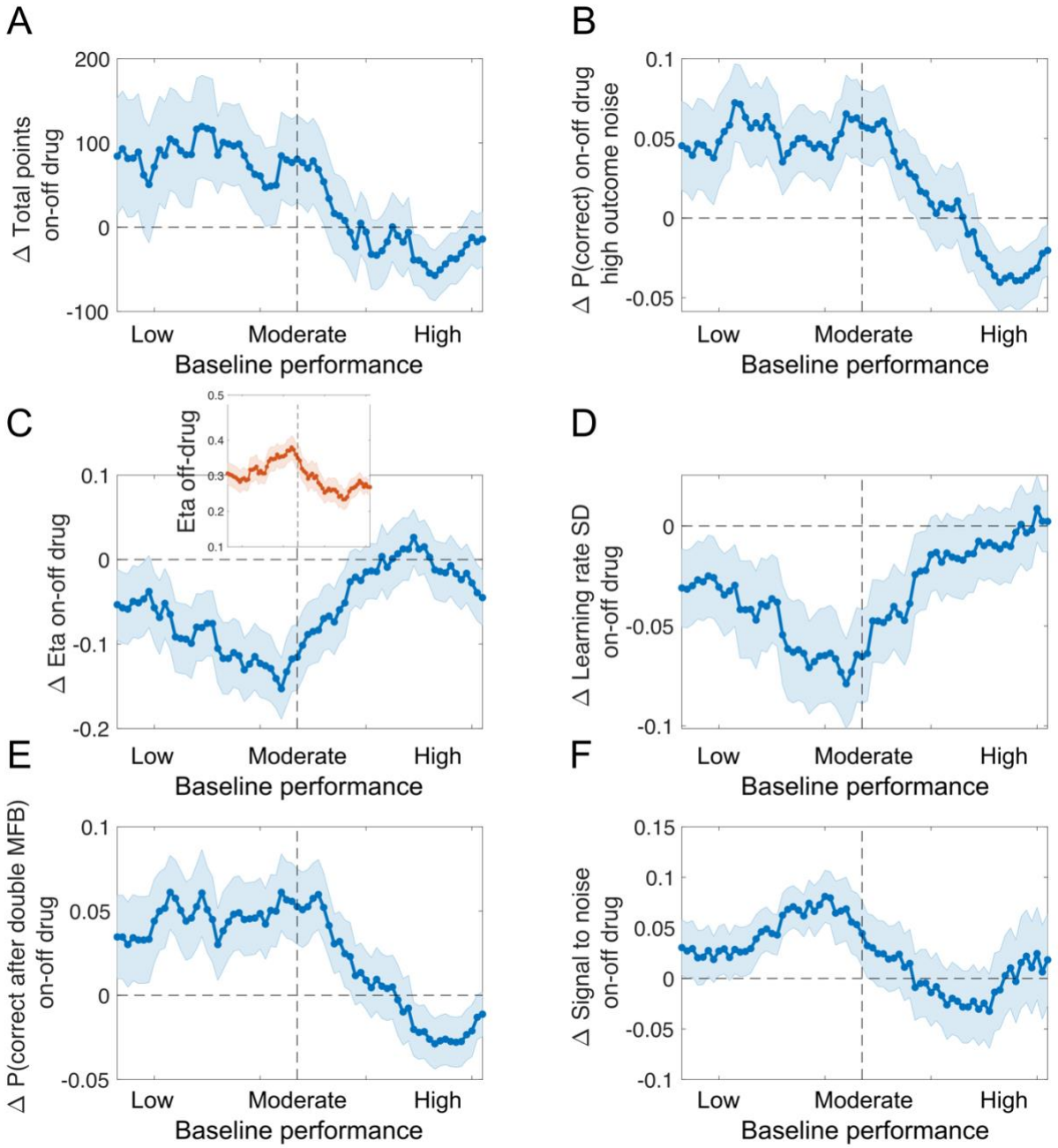
$$Outcome\ Variable_i = \beta_0 + \beta_1 \text{SessionOrder}_i + \beta_2 \text{DrugSession}_i + u_{s(i)} + \varepsilon_i$$

where:

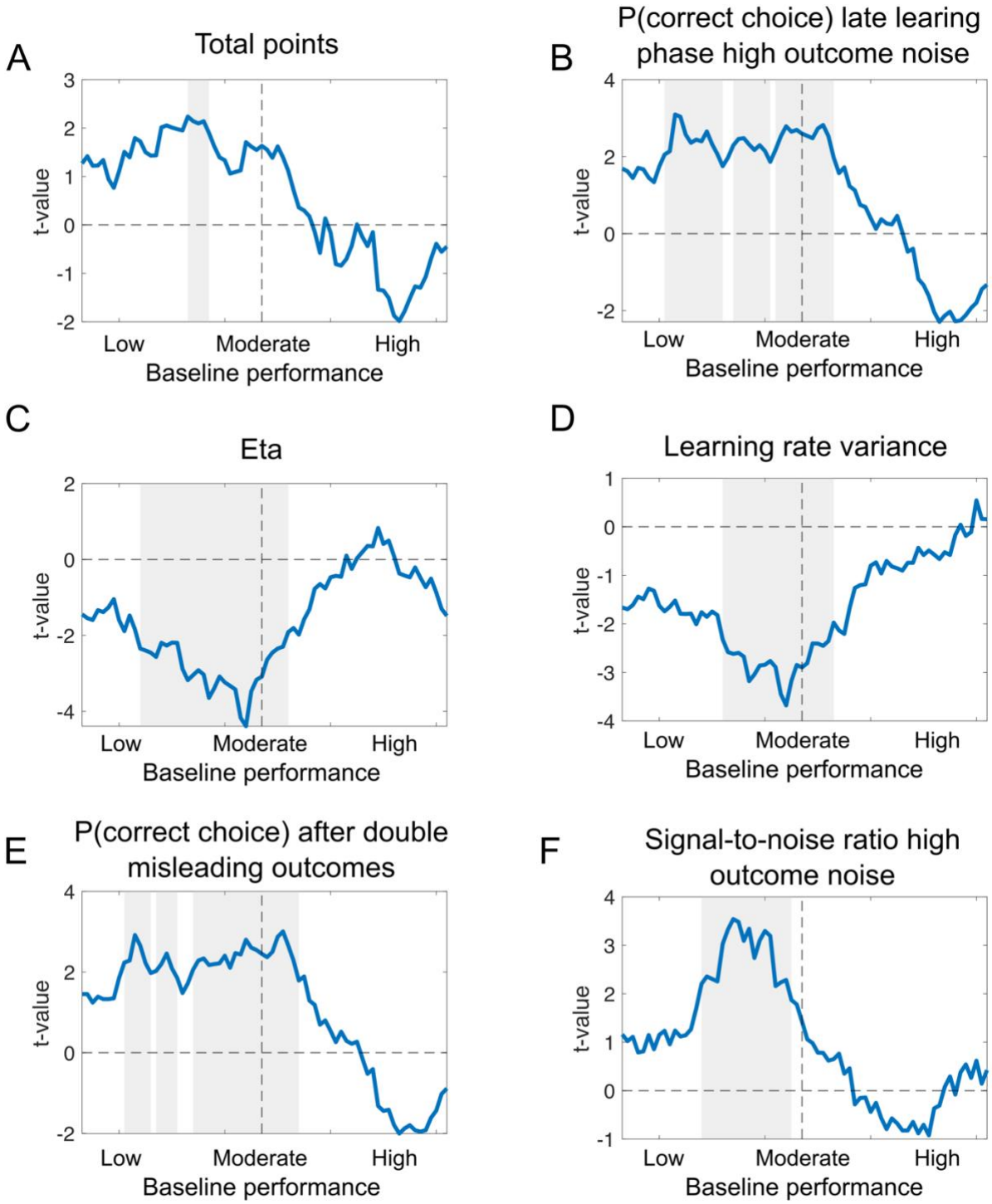
- $Outcome\ Variable_i$ is the same as described above
- $\beta_0$is the fixed intercept.
- β_1 $\beta_2$are the fixed-effect coefficients.
- Session Order_i and Drug Session_i remain as fixed effects.
- $u_{s(i)} \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept for subject $s(i)$.
- $\varepsilon_{s(i)} \sim \mathcal{N}(0, \sigma_u^2)$ is the residual error term.

The results of these models (see Supplementary Figure 5) indicate that indeed the drug effect is strongest in moderately low baseline performer. A key thing to notice in the sliding regression results is that, while each regression reveals that drug effects depend on baseline performance, they do so non-linearly, with most variables of interest showing a saturating effect at low baseline performance levels and the strongest slope (dependence on baseline) at or near the median level of baseline performance, explaining why our median splits were able to successfully pick up on these baseline-dependent effects.

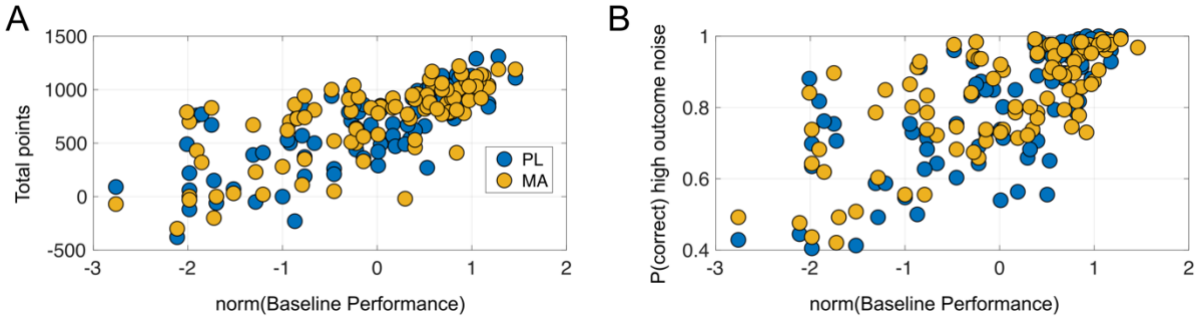
Together, these results suggest that methamphetamine primarily affects individuals with medium-to-low baseline performance.



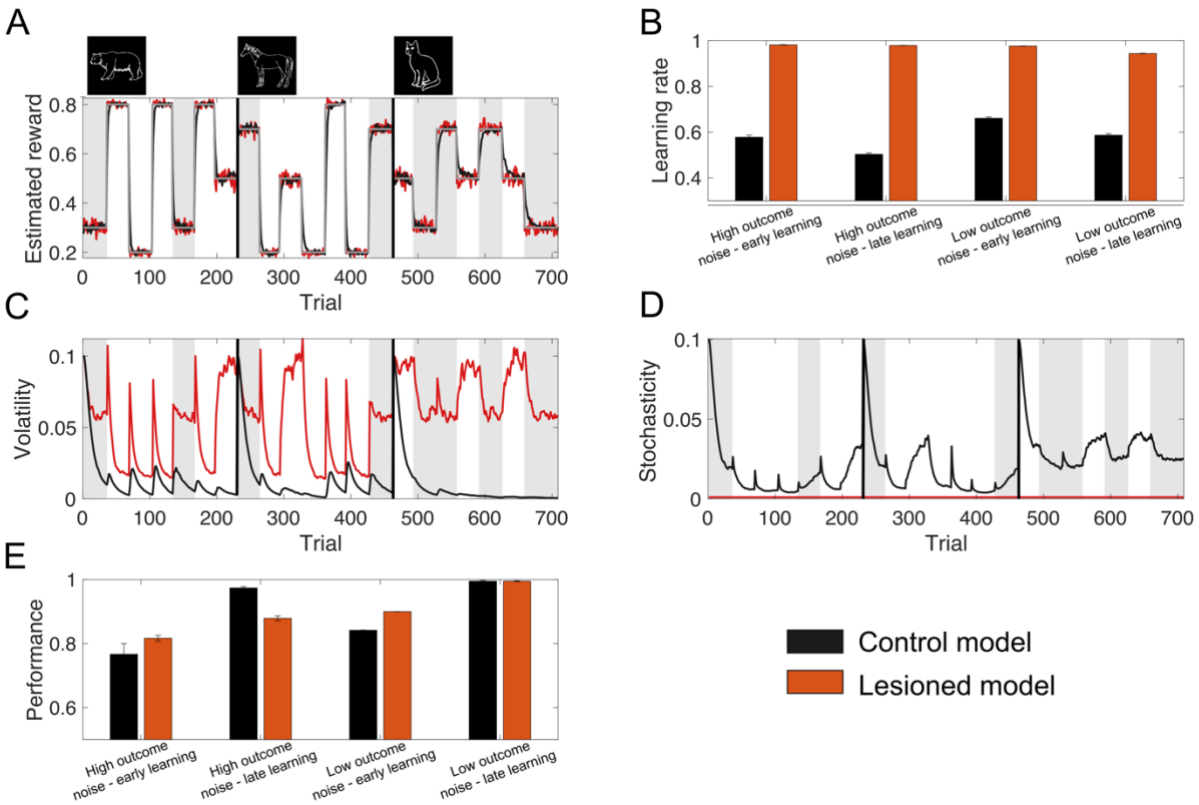
Supplementary Figure S4. The effect of methamphetamine depends on baseline performance. To examine the relationship between methamphetamine effects and baseline performance, we plotted difference scores (i.e., the outcome variable under the drug condition minus placebo condition) against baseline performance. **(A-F)** show nonlinear relationships between baseline performance and outcome variables (**(A)** total points, **(B)** probability of a correct choice (P(correct choice)) in the late learning phase with high outcome noise, **(C)** parameter estimate for eta, **(D)** learning rate variance, **(E)** probability of a correct choice (P(correct choice)), and **(F)** signal-to-noise ratio in high outcome noise). Specifically, we show that drug effects were maximal in moderately low baseline performer. It is noteworthy that these subjects had particularly high Eta's on placebo (inset in **C**) which may have allowed the drug effects to have a larger impact on their performance. This in line with our key finding, that methamphetamine brings eta (parameter controlling dynamic adjustments of learning rate according to recent prediction errors) closer to optimal levels. Vertical dashed line indicates medium baseline performance. MFB = misleading feedback.



Supplementary Figure S5. The effect of methamphetamine is strongest in moderately low baseline performers. This figure shows the drug effect from the sliding window linear mixed-effects model analysis plotted against baseline performance for a set of dependent variables: **(A)** total points, **(B)** probability of a correct choice (P(correct choice)) in the late learning phase with high outcome noise, **(C)** parameter estimate for eta, **(D)** learning rate variance, **(E)** probability of a correct choice (P(correct choice)), and **(F)** signal-to-noise ratio in high outcome noise. Grey areas indicate clusters with significant drug effects after correcting for multiple comparisons using a permutation test for cluster mass (Maris & Oostenveld, 2007).

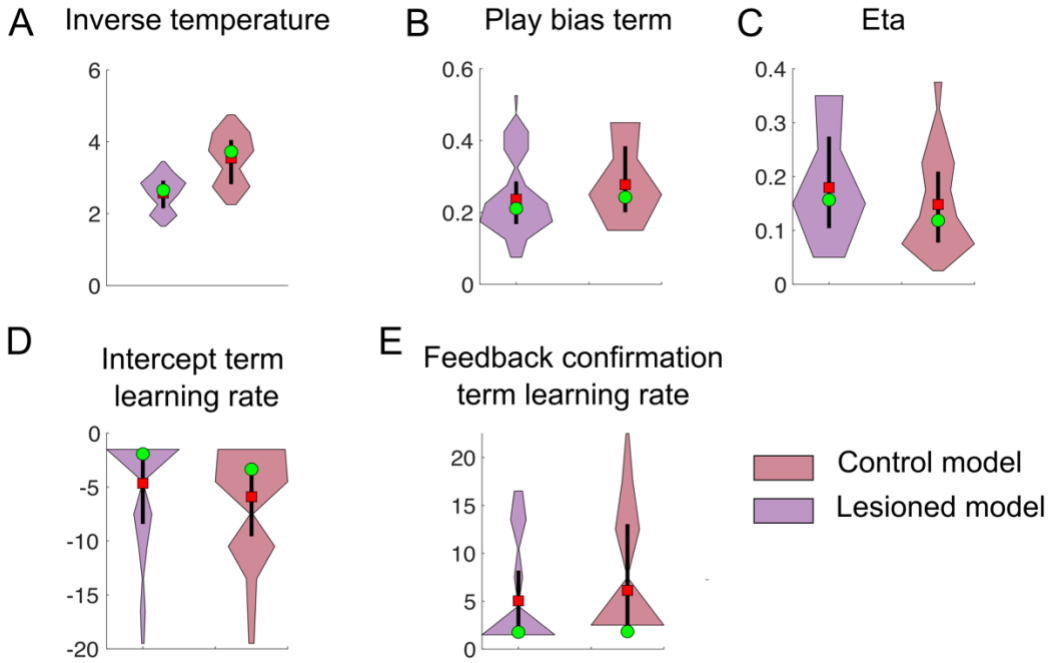


Supplementary Figure S6. High baseline performers reach near-optimal levels in bost test sessions, limiting potential drug-induced performance enhancement. (A) Overall task performance and **(B)** the probability of correct choices in the high outcome noise condition—the condition with the strongest observed drug effect—are plotted against normalized baseline performance. In both testing sessions, high baseline performers cluster around optimal performance. Furthermore, performance simulations using (a) optimal eta values and (b) observed eta values from the high baseline performance group reveal only a small, non-significant performance difference (points optimal eta: 701.91 (21.66) vs. points high performer: 694.47 (21.71); t(46)=2.84,p=0.07, d=0.059 t(46)=2.84, p=0.07, d=0.059). These results suggest that high baseline performers are already at or near their performance ceiling, limiting the potential for further drug-induced improvements.



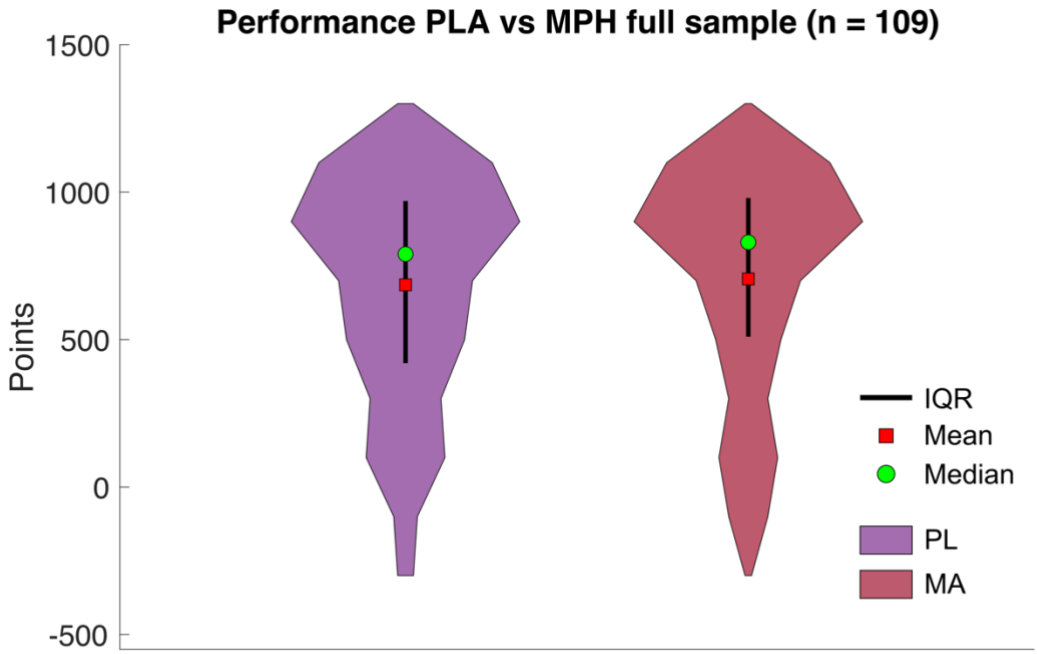
Supplementary Figure S7. The stochasticity lesion model shows a pattern of learning deficits associated low performer in our task. Behaviour of the lesioned model, in which stochasticity is assumed to be small and constant, is shown along the control model that jointly estimates stochasticity and volatility. An example of estimated reward by the models shows that the stochasticity lesion model is more sensitive to noisy outcomes **(A)**. This reduces sensitivity of the learning rate to volatility (i.e., first ten trials after a reversal **(B)**. This, however, is primarily related to inability to make inference about stochasticity, which leads to misestimation of volatility **(C–D)**. Simulations reveal that this is

accompanied by reduced performance, particularly for high outcome noise stimuli (grey patches) later in learning **(E).**

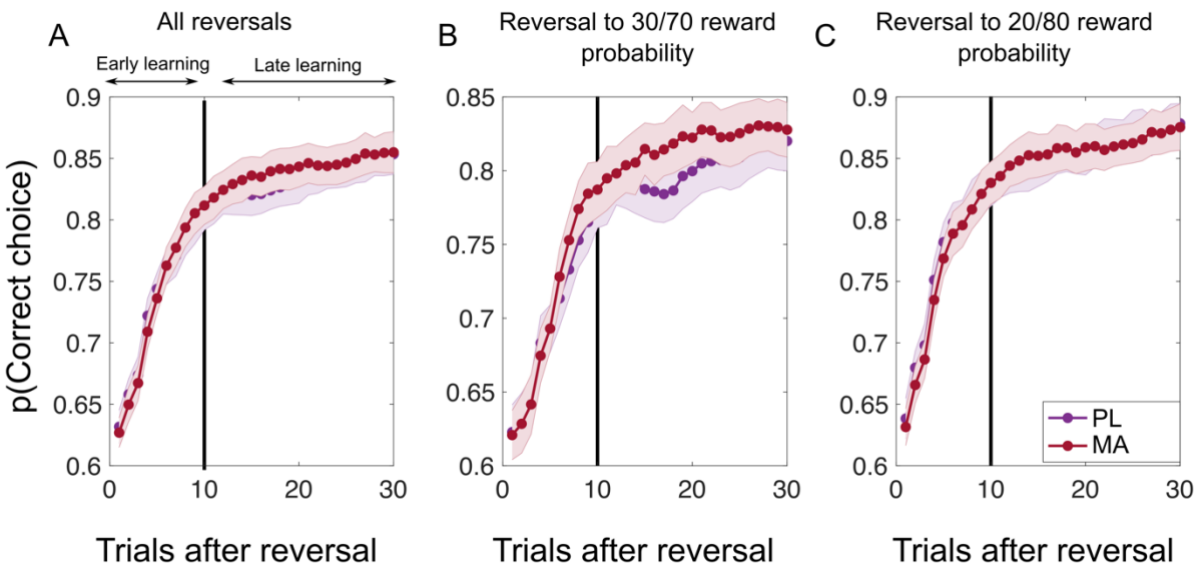


Supplementary Figure S8. Simulated data form the stochasticity lesion model shows increased eta parameters compared to the control model. Here, we fit simulated data from the control and stochasticity lesion model (100 simulations for each model) to our task model. Data from the control model reveal increased inverse temperature (**A**; 2.57 (0.04) vs. 3.54 (0.07); t(198) = -11.09, p = 0, d = 1.71); increased play bias (**B**; 0.24 (0.01) vs. 0.28 (0.01); t(198) = -2.56, p = 0.01, d = 0.39), and decreased eta's (**C**; 0.18 (0.01) vs. 0.14 (0.01); t(198) = 2.21, p = 0.02, d = 0.34). There was no difference in the intercept term of the learning rate (**D**) or the feedback confirmation term of the learning rate (**E**).

Full Sample results

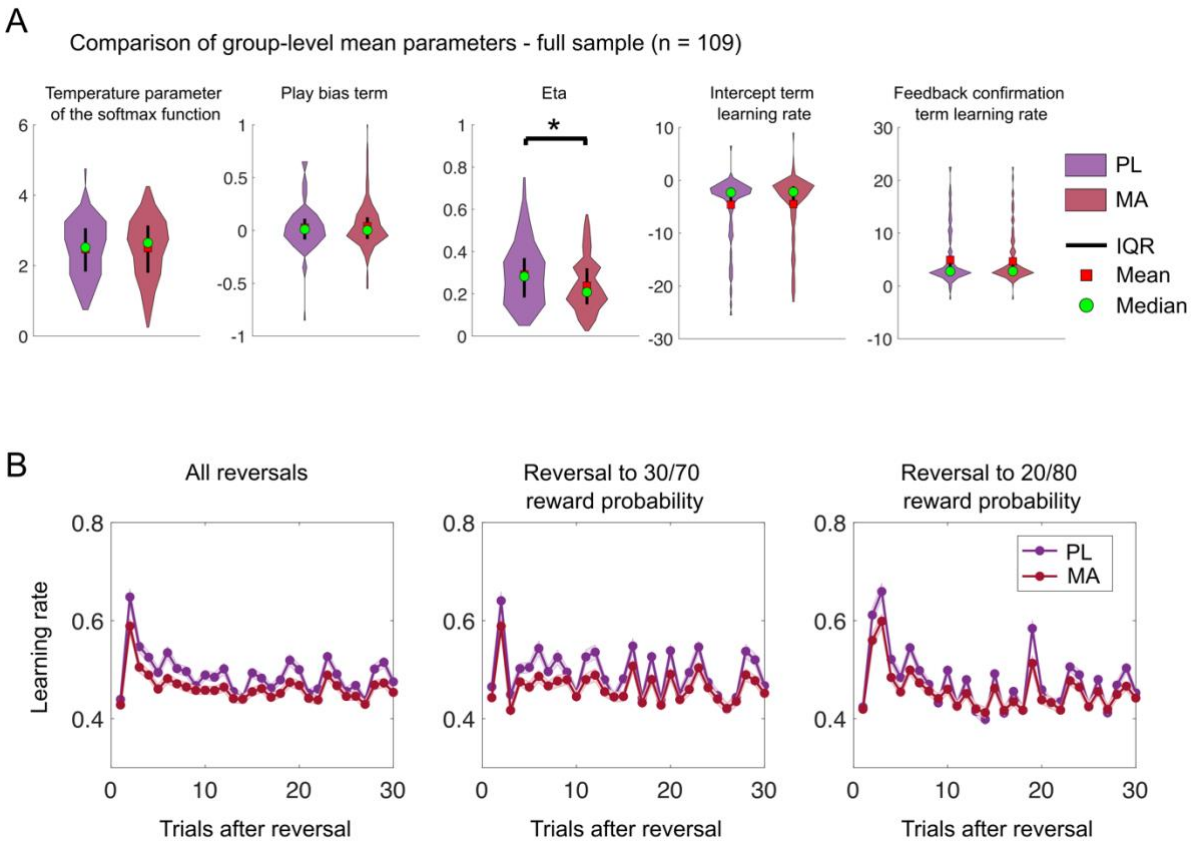


Supplementary Figure S9. Overall points full sample.
When comparing overall point in the whole sample (n = 109), we do not see a difference between MPH vs. PLA (705.68 (36.27) vs. 685.77 (35.78); $t(108) = 0.81$, $p = 0.42$, d = 0.05). Repeated mixed ANOVAs suggested, that drug effects did not depend on session order (MPH first vs. PLA first), or whether subjects performed the orientation session. Yet, participants who completed the orientation tended to performed better during the dug sessions ($F(1,107) = 3.09$, $p = 0.08$; 719.31 (26.6264) vs. 548.00 (75.09)). Note. PL = Placebo; MA = methamphetamine.



Supplementary Figure S10. Learning curves after reversals full sample.

Figure shows learning curves after all reversals **(A)**, reversals to high reward probability uncertainty **(B)**, and reversals to low reward probability uncertainty **(C)** for the whole sample. Vertical black lines divide learning into early and late stages as suggested by the Bai-Perron multiple break point test. Paired-sample t-test revealed no drug related difference for all reversals during early learning (0.72 (0.01) vs. 0.72 (0.01); t(108) = -0.02, p = 0.98, d < 0.01) and late learning (0.83 (0.01) vs. 0.84 (0.01); t(108) = -0.80, p = 0.42, d = 0.04). Similarly, there was no significant differences in both learning stages for reversals to low reward probability certainty stimuli (early learning PLA vs MPH: 0.68 (0.01) vs. 0.69 (0.01); t(108) = -0.92, p = 0.35, d = 0.08; late learning PLA vs MPH: 0.80 (0.01) vs. 0.81 (0.01); t(108) = -1.48, p = 0.14, d = 0.10) or to low reward probability certainty stimuli (early learning PLA vs MPH: 0.74 (0.01) vs. 0.73 (0.01); t(108) = 0.87, p = 0.38, d = 0.06; late learning PLA vs MPH: 0.85 (0.01) vs. 0.85 (0.01); t(108) = -0.02, p = 0.97, d < 0.01). Mixed effect ANOVAs that controlled for session order effects and whether participants performed the orientation session revealed no significant effects (all p > .06). Note. PL = Placebo; MA = methamphetamine.



Supplementary Figure S11. Learning curves after reversals full sample.
(A) Here we compare MPHs effect on best-fitting parameters of the winning model in the full sample (n = 109). We found that eta (i.e., the weighting of the effect of the abs. reward prediction error on learning) was reduced under MPHs (eta MPH: 0.23 (0.01) vs. PLA 0.29 (0.01); t(108) = -3.05, *p* = 0.002, d = 0.40). Mixed effect ANOVAs that controlled for session order effects and whether participants performed the orientation session revealed this effect did not depend on these cofounds. No other condition differences emerged. **(B)** Learning rate trajectories after reversal derived from the computational model. As in the reduced sample MPH appears to be associated reduced learning rate dynamics in the full sample too. Specifically, variability in learning rate (average individual SD of learning rate) tended to be reduced in the MPH condition both during early and late stages of learning across all reversals (early PLA: 0.19 (0.01) vs. 0.18 (0.01); t(108) = 1.89, p = 0.06, d = 0.24; late PLA:

0.17 (0.01) vs. MPH: 0.16 (0.01); t(108) = 1.77, p = 0.08, d = 0.23) and reversals to high reward probability uncertainty (early PLA: 0.18 (0.01) vs. 0.16 (0.01); t(108) = 1.74, p = 0.08, d = 0.22; late PLA: 0.18 (0.01) vs. MPH: 0.16 (0.01); t(108) = 1.82, p = 0.07, d = 0.24). Condition differences became most evident in reversals to low reward probability uncertainty (early PLA: 0.19 (0.01) vs. MPH: 0.16 (0.01); t(108) = 2.18, p = 0.03, d = 0.28; late PLA: 0.18 (0.01) vs. MPH: 0.16 (0.01); t(108) = 1.93, p = 0.05, d = 0.24). Control analyses revealed that these effects were independent of session order and orientation session. Note. PL = Placebo; MA = methamphetamine.