

Effects of methylphenidate on reinforcement learning depend on working memory capacity

Rostami Kandroodi, Mojtaba; Cook, Jennifer L.; Swart, Jennifer C.; Froböse, Monja I. ; Geurts, Dirk E. M. ; Vahabie, Abdol-Hossein; Nili Ahmadabadi, Majid; Cools, Roshan; den Ouden, Hanneke E. M.

DOI:

[10.1007/s00213-021-05974-w](https://doi.org/10.1007/s00213-021-05974-w)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Rostami Kandroodi, M, Cook, JL, Swart, JC, Froböse, MI, Geurts, DEM, Vahabie, A-H, Nili Ahmadabadi, M, Cools, R & den Ouden, HEM 2021, 'Effects of methylphenidate on reinforcement learning depend on working memory capacity', *Psychopharmacology*, vol. 238, no. 12, pp. 3569-3584. <https://doi.org/10.1007/s00213-021-05974-w>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.



Effects of methylphenidate on reinforcement learning depend on working memory capacity

Mojtaba Rostami Kandroodi^{1,2} · Jennifer L. Cook³ · Jennifer C. Swart² · Monja I. Froböse² · Dirk E. M. Geurts^{2,4} · Abdol-Hossein Vahabie^{1,5} · Majid Nili Ahmadabadi¹ · Roshan Cools^{2,4} · Hanneke E. M. den Ouden²

Received: 12 April 2021 / Accepted: 25 August 2021
© The Author(s) 2021

Abstract

Rationale Brain catecholamines have long been implicated in reinforcement learning, exemplified by catecholamine drug and genetic effects on probabilistic reversal learning. However, the mechanisms underlying such effects are unclear.

Objectives and methods Here we investigated effects of an acute catecholamine challenge with methylphenidate (20 mg, oral) on a novel probabilistic reversal learning paradigm in a within-subject, double-blind randomised design. The paradigm was designed to disentangle effects on punishment avoidance from effects on reward perseveration. Given the known large individual variability in methylphenidate's effects, we stratified our effects by working memory capacity and trait impulsivity, putatively modulating the effects of methylphenidate, in a large sample ($n = 102$) of healthy volunteers.

Results Contrary to our prediction, methylphenidate did not alter performance in the reversal phase of the task. Our key finding is that methylphenidate altered learning of choice-outcome contingencies in a manner that depended on individual variability in working memory span. Specifically, methylphenidate *improved* performance by adaptively *reducing* the effective learning rate in participants with higher working memory capacity.

Conclusions This finding emphasises the important role of working memory in reinforcement learning, as reported in influential recent computational modelling and behavioural work, and highlights the dependence of this interplay on catecholaminergic function.

Keywords Reversal learning · Methylphenidate · Catecholamines · Dopamine · Working memory · Computational modelling of behaviour

These authors contributed equally: Roshan Cools, Hanneke E.M. den Ouden

✉ Mojtaba Rostami Kandroodi
mojtaba.rostami@ut.ac.ir

✉ Hanneke E. M. den Ouden
h.denouden@donders.ru.nl

¹ School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

² Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

³ School of Psychology, University of Birmingham, Birmingham, United Kingdom

⁴ Department of Psychiatry, Radboud University Medical Centre, Nijmegen, The Netherlands

⁵ School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Introduction

Brain catecholamines (dopamine and noradrenaline) are well known to play a fundamental role in reinforcement learning and decision-making. Most notably, in the last 2 decades, a wealth of studies have shown that dopaminergic midbrain firing increases when experience exceeds expectations (Montague et al. 1996; Fiorillo et al. 2003; Schultz 2016). This dopaminergic signalling is widely accepted to function as a teaching signal driving reinforcement learning (Niv and Montague 2009). Dopamine has also been implicated in the ability to flexibly adjust behaviour to changing environments (Swainson et al. 2000; Cools et al. 2001; Chudasama and Robbins 2006; Dodds et al. 2008; Boulougouris et al. 2009; Clatworthy et al. 2009; Clarke et al. 2011; Cools and D'Esposito 2011; Groman et al. 2012; den Ouden et al. 2013). For example, selective lesioning of striatal dopamine in marmoset monkeys impaired the ability to reverse learnt

stimulus-reward associations (Clarke et al. 2011), in line with classic findings from studies in rodents showing that reversal learning is altered by dopaminergic modulation of the (ventral) striatum (Taghzouti et al. 1985; Smith et al. 1999; Goto and Grace 2005). In patients with Parkinson's disease, dopaminergic medication has been shown to impair performance selectively on the reversal learning stage of a probabilistic reversal learning task while leaving learning during an initial acquisition phase unaltered (Cools et al. 2001). In line with the proposal that this impairment reflects detrimental overdosing of relatively intact dopamine levels in the ventral striatum, dopaminergic medication in Parkinson's disease was shown to attenuate reversal-related BOLD signal in the ventral (but not dorsal) striatum (Cools et al. 2007a). Subsequent studies in young healthy volunteers have shown that administration of the dopamine (and noradrenaline) transporter blocker methylphenidate to healthy volunteers modulates reversal-related BOLD signal in the striatum (Dodds et al. 2008) and impaired reversal learning in proportion to the degree that methylphenidate increased striatal dopamine release (Clatworthy et al. 2009). While these studies establish a causal role for striatal dopamine specifically in reversal learning, the mechanism by which dopamine alters the ability to reverse responding remains unclear.

In the present study, we aimed to elucidate the nature of the catecholaminergic effects on perseverative behaviour during reversal learning. Specifically, perseveration may result from either a 'stamping in' of rewarded behaviour, leading to an inability to 'let go' of responding to a previously rewarded stimulus, or from a changed ability to approach a previously punished stimulus (i.e. punishment avoidance). While striatal Go/NoGo pathway models, including the Opponent Actor Learning model, posit that perseveration might follow from both increasing the impact of reward and reducing the impact of punishment (Frank 2005; Collins and Frank 2014), we hypothesised that reversal deficits are more likely to follow from disproportionate stamping in of rewarded behaviour, rather than from diminished punishment avoidance. This hypothesis is grounded in seminal work with experimental rodents, showing that injection of D-amphetamine in the nucleus accumbens of rats potentiates behavioural control by stimuli formerly associated with reward (i.e. conditioned reinforcement) in a DA-dependent way (Robbins et al. 1989; Parkinson et al. 1999). Moreover it follows directly from our prior genetic study of probabilistic reversal learning (den Ouden et al. 2013). Specifically, we have shown that perseveration elicited by genetic variability in the dopamine transporter *DAT1* can be accounted for by progressively increased reliance on prior reinforcement, captured by an increase in an experience-weight parameter in an augmented reinforcement learning model (den Ouden et al.

2013). Thus, natural genetic *DAT1* variation was associated with a stronger correlation between reinforcement history and perseveration (den Ouden et al. 2013). Accordingly, we hypothesised that increasing catecholamine signalling would alter perseverative behaviour, specifically by inducing an inability to 'let go' (i.e. stop choosing) a previously rewarded stimulus rather than by impairing the ability to approach a previously punished stimulus (Frank et al. 2004; Cools et al. 2006).

To dissociate these two alternative mechanisms of perseverative behaviour arising from punishment avoidance or excessive adherence to previously rewarded stimuli, we introduce a novel reversal learning paradigm that included a 'neutral' choice option. To assess whether and which of these mechanisms are affected by catecholamine signalling, we combined this novel paradigm with administration of catecholamine transporter blocker methylphenidate, which acts by blocking dopamine and noradrenaline transporters (DAT/NAT). This blockade increases extracellular catecholamine availability in the synaptic cleft, without stimulating release or acting as a receptor (ant)agonist (Volkow et al. 2002). It is thought to prolong the effect of both dopamine and noradrenaline release, as reuptake is slowed (Madras et al. 2005; Berridge et al. 2006).

Finally, previous studies have shown that there is large inter-task and inter-individual variability in catecholaminergic drug effects on cognitive task performance (Kimberg et al. 1997; Cools et al. 2007b; Van Der Schaaf et al. 2013; Linssen et al. 2014; Swart et al. 2017; Froböse et al. 2018; Cook et al. 2019), including probabilistic reversal learning (Clatworthy et al. 2009). Given that methylphenidate prolongs the effects of catecholamine release by blocking the reuptake of catecholamines, it is likely that the effect of methylphenidate on catecholamine-dependent function is a function of dopamine synthesis capacity and release. Simply put, if there is no release, there is no reuptake to block. To take into account the established large individual variability in methylphenidate effects, we collected a large sample ($n = 102$) to expose individual differences and stratified methylphenidate effects by two measures that have been previously demonstrated to relate to baseline dopamine function: working memory (WM) span for its relation to striatal dopamine synthesis capacity (Cools et al. 2008; Landau et al. 2009) and trait impulsivity for its relation to dopamine (auto)receptor availability (Lee et al. 2009; Buckholtz et al. 2010; Reeves et al. 2012; Kim et al. 2014). Based on previous studies where we used methylphenidate in combination with various reinforcement learning tasks (Van Der Schaaf et al. 2013; Swart et al. 2017; Cook et al. 2019), we hypothesised that WM span specifically would predict the inter-individual differences in reversal learning.

Methods

General procedure and pharmacological intervention

Data was collected April 15–September 1, 2014, and took place at the Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging. The study consisted of two test sessions with an interval of 1 week to 2 months. The first test day started with informed consent, followed by a medical screening. Participation was discontinued if participants met any of the exclusion criteria (supplemental methods 3). On both test days, participants first completed baseline measures, as well as the Instrumental and Pavlovian phases of the Pavlovian-Instrumental transfer (PIT) task (Geurts et al. 2021). Participants received a capsule containing either 20 mg of catecholamine transporter blocker methylphenidate (Ritalin®, Novartis) or placebo, in a double-blind, placebo-controlled, cross-over design. This relatively low dose was selected because (i) this minimises any potential risks, (ii) it has been found sufficient to affect cognitive performance and to do so in a manner that is indistinguishable from administration of a higher (40 mg) dose (e.g. Elliott et al. 1997), and (iii) dopamine microdialysis in macaque monkeys has shown that low dose of MPH leads to relatively preferential effects on striatal (relative to prefrontal) DA release (Kodama et al. 2017). When administered orally, methylphenidate has a maximal plasma concentration after 2 h and a plasma half-life of 2–3 h (Kimko et al. 1999). Below we denote capsule intake as $t = 0$.

The probabilistic reversal learning task was the last task participants completed following capsule intake, at $t = 186.1$ (7.9) min, mean (st.d.). This task was preceded by 5 other tasks published elsewhere (Swart et al. 2017; Froböse et al. 2018; Cook et al. 2019) (Fig. 1). Both test days lasted approximately 4.5 h, which participants started at the same time of day (maximum difference of 45 min). Blood pressure, mood, and potential medical symptoms were monitored three times daily: before capsule intake ($t = -5.3$ (1.7) min), directly prior to start of the task battery ($t = 47.4$ (7.6)), and after finishing the task battery ($t = 190.9$ (7.9)). Mood and medical symptom ratings are described in supplemental methods 4. Participants were instructed to abstain from alcohol and recreational drugs 24 h prior to testing and from smoking and drinking coffee on testing days. Participants completed self-report questionnaires at home between test days. Upon completion of the study, participants received a monetary reimbursement or study credits for participation. The study was in line with the local ethical guidelines approved by the local ethics committee (CMO/METC Arnhem Nijmegen: protocol

NL47166.091.13), preregistered (trial register NTR4653, <http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4653>), and in accordance with the Helsinki Declaration of 1975.

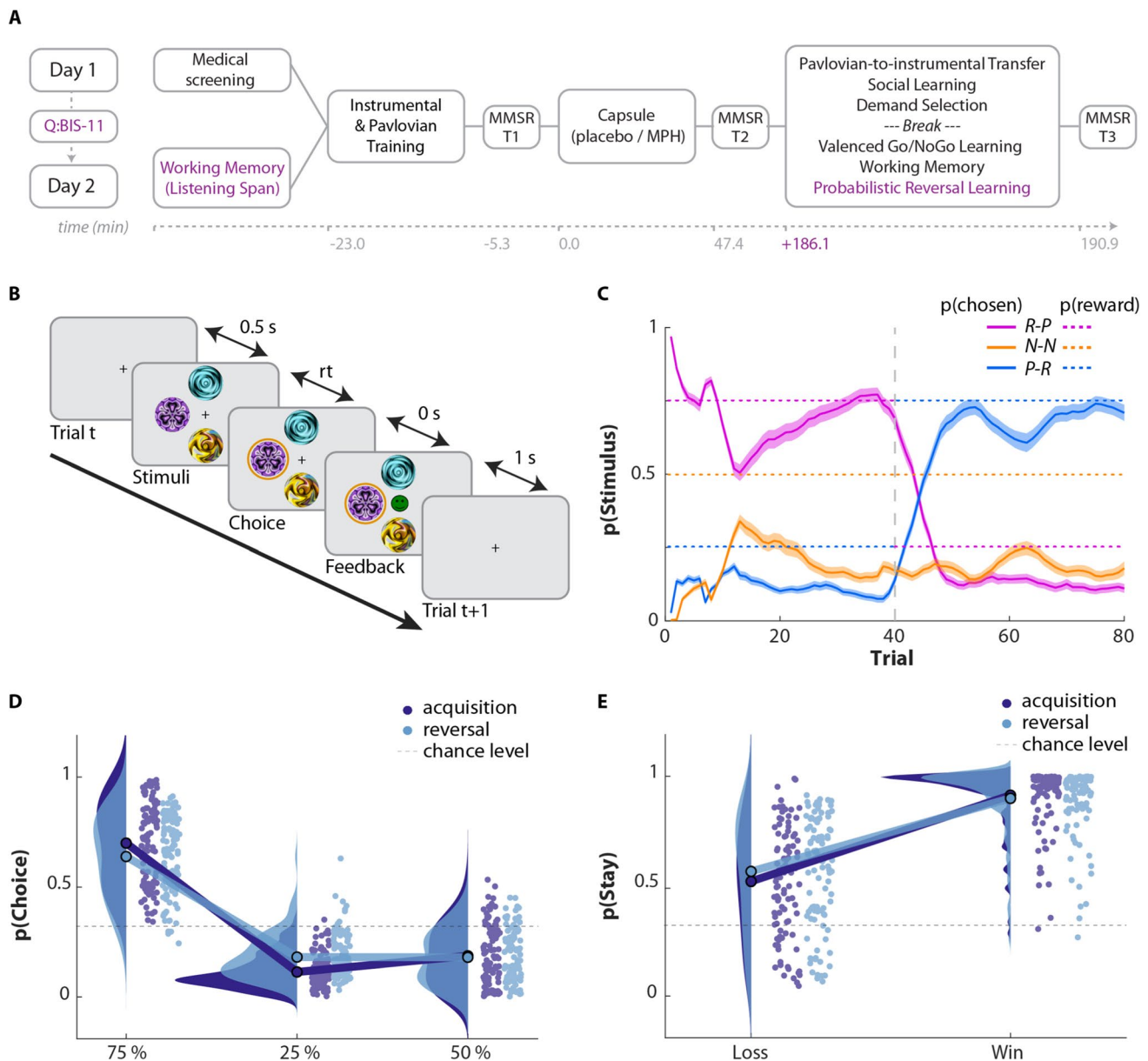
Participants were allocated pseudo-randomly to the intervention order (placebo vs methylphenidate first) in a double-blind, cross-over design. For details on the randomisation procedure, see supplemental methods 5.

Participants

As specified in the preregistration, we planned to test a sample of 100 participants. To this end, 176 healthy, young adults were recruited via flyers around the campus and the digital participant pool of the Radboud University, Nijmegen. All participants were native Dutch speakers and provided written informed consent to participate. Exclusion criteria comprised a history of psychiatric, neurological, or endocrine disorders (see supplemental methods 3 for a complete overview of the exclusion criteria). One hundred six participants who met the inclusion criteria were included in this study to reach the planned 100 participants: Data from four participants could not be collected on day 2 due to medical reasons (mild arrhythmia: $n = 1$, elevated heart rate and nausea: $n = 1$) and drop-out ($n = 2$). These were replaced. A further 2 participants were replaced, who had difficulty swallowing the capsules and for whom the capsule content was suspended in water on both testing days. These 2 participants were included in the final analyses, as we verified that their inclusion/exclusion did not affect the results. Thus, the final analyses include 102 adult participants (aged 18–28 years, mean = 21.5, st.d. = 2.3, 51 women, 81 right-handed), where 50 participants received methylphenidate on the first testing session. Additional demographic information and results from baseline neuropsychological assessment and self-report questionnaires of included participants are reported in supplemental methods 6.

Probabilistic reversal learning task

Participants performed a probabilistic reversal learning (PRL) task with three choice options in each trial. This task is an adjusted version of den Ouden et al. (2013), which consisted of 2 cues that were either predominantly rewarded during acquisition and predominantly punished during reversal (stimulus ‘ $R-P$ ’) or the reverse (stimulus ‘ $P-R$ ’). We here added a neutral stimulus with 50–50% contingencies throughout the task (stimulus ‘ $N-N$ ’). The rationale for the addition of this neutral stimulus was to allow us to dissociate between two putative causes of perseverative behaviour. On the one hand, perseveration may result an inability to stop choosing a previously rewarded stimulus, so true ‘perseveration’ of the previously rewarded response. On the



other hand, it may result from punishment avoidance, i.e. an inability to approach a previously punished stimulus. Introducing a neutral stimulus makes differential predictions for these two scenarios. In case of the former, perseverative behaviour would manifest as continued selection of the previously rewarded response. However, in case of punishment avoidance, following reversal, the participant is able to unlearn the previously rewarded response but would fail to learn to select the previously punished response and will now preferentially select the neutral outcome.

On each of 80 trials, three visual stimuli $\{R-P, P-R, N-N\}$ were represented in three out of four pseudo-randomly selected locations (left, right, top, or bottom; Fig. 1B). Participants chose one of the stimuli with a mouse click and subsequently received feedback. There

was no time limit for responses. Feedback was either a reward (green, happy emotion) or punishment (red, sad emotion). To maximise reward, participants had to learn by trial-and-error to choose the mostly rewarded stimulus. During the acquisition phase, stimulus $R-P$ (defined as the first stimulus that was chosen by the participant) gives reward/punishment with contingencies of 75:25%, while $P-R$ results in reward/punishment with the opposite (25:75%) contingency ratio. After 40 trials, these reinforcement contingencies reversed. Thus, the task consisted of an acquisition and a reversal phase. The third stimulus $N-N$ is 'neutral', as its selection results in 50:50 ratio reward/punishment throughout the task. For information on the generation of the feedback sequence and full task instructions, see supplemental methods 1 and 2.

Fig. 1 Experimental design and basic results. **A** Study timeline. Participants took part in 2 sessions, where on day 1, they started with a 20-min medical screening and on day 2 completed a working memory test (listening span). Neuropsychological questionnaires (Quest) were completed at home between the two sessions. Mood and medical symptom ratings (MMSR) were acquired at 3 time points. A battery of 6 tasks (Swart et al. 2017; Froböse et al. 2018; Cook et al. 2019) was performed in a fixed order, where the probabilistic reversal learning (PRL) task was always performed last. Average timings are indicated, with timings most relevant for the current study in purple. **B** Reversal learning design. On each trial, three visual stimuli were represented in three out of four randomly selected locations. Participants had to choose one of the stimuli with a mouse click and subsequently received feedback. The feedback would either be a reward (green, happy emotion) and punishment (red, sad emotion). During acquisition, the rewarded stimulus (here purple) resulted in a 75:25 ratio of reward/punishment. Selecting the neutral stimulus (orange) and punished stimulus (blue) led to 50:50 and 25:75 ratio of reward/punishment, respectively. After 40 trials, the reversal phase started, and rewarded and punished stimulus contingencies reversed. The participant now had to learn to select the blue stimulus. **C** Trial-by-trial choice. Trial-by-trial averaged probability of selection of each stimulus. A sliding window with 5-trial width is used for smoothing. Overall, participants learned to make the correct selection for each of the phases and showed relatively rapid reversal. **D** Average choice probability. Distribution of choice probability averaged within acquisition (dark blue) and reversal (light blue) phases. In both phases, participants learnt to select 75% rewarded option but do significantly less well during the reversal phase. **E** Feedback sensitivity. The degree to which people repeated a choice was modulated by the valence of the previous outcome for that choice: People were more likely to reselect a stimulus ('stay') after it had been rewarded than after it was punished. This effect was weaker during the reversal phase (less stay after a win, more stay after a loss), in line with slower learning during reversal. Note that the intercept in this plot was chance level stay (1/3), even after a loss people were more likely to stay than chance (simple effect of pStay following loss: $F(1,99)=77.1$; $p<0.0001$, following win $F(1,99)=1603.4$; $p<0.0001$), reflecting a relatively small impact of a single (negative) outcome and thus slow learning

Behavioural data analysis

Data quality assessment

A priori we decided to exclude any participant who selected the same stimulus the entire experiment as they likely did not understand the task (cf. den Ouden et al. 2013). However, this did not happen in the current sample. Trials with RT faster than 200 ms likely reflect responses that were not based on a deliberate choice between the stimuli. We confirm that results do not change whether these trials are included (main article) or excluded (supplemental results & discussion 3).

Choice accuracy

The statistics software SPSS (version 25) was used to analyse the behavioural choices. The probability of choosing each stimulus was calculated separately for the acquisition and reversal phases. To assess the putative mechanisms

that may drive reduced performance following reversal, we contrasted two measures of performance accuracy: (i) the probability of selecting the 75% rewarded stimulus (pReward) and (ii) the probability of avoiding the 75% punished stimulus (1-pPunish or pAvoidPunish). To ensure that the intercept of this analysis was interpretable, we corrected scores for chance performance (i.e. subtracted 1/3). The basic design of this analysis was a 2×2 factorial ANOVA with factors Phase (acquisition/reversal) and Valence (pReward/pAvoidPunish). Note that these two measures are rendered (relatively) independent by inclusion of the neutral stimulus, which is the implicit baseline. The intercept of this ANOVA indexes the ability to learn to select the rewarded stimulus and avoid the punished stimulus. A main effect of Phase can capture the relative reduction of performance in the reversal phase. Most importantly, an interaction of Valence \times Phase assesses whether there is a difference in the degree to which people fail to 'let go' of the previously rewarded stimulus (reduced pAvoidPunish selection during reversal) or the degree to which they fail to approach the previously punished stimulus (reduced pReward during reversal).

To assess the effect of methylphenidate on performance, this basic design was extended with the factor Drug, resulting in a three-way repeated measure ANOVA with Valence (pReward, pAvoidPunish), Phase (acquisition, reversal), and Drug (methylphenidate, placebo) as within-subject factors. Furthermore, listening span total score and Barratt impulsiveness score were added as covariates of interest. For reporting significant results, the Huyn-Feldt correction was used when significant non-sphericity was detected.

Any significant interactions were broken down into their component simple effects to aid interpretation of the effects.

Control analyses of effects of no interest

To verify that our findings are not confounded by covariates (age and Nederlandse Leestest voor Volwassenen (NLV); Dutch adult reading test; a measure of verbal intelligence) and factors (gender, testing order) of no interest, we repeated the analyses above including these covariates and factors and confirmed that significant results remained significant and non-significant results remained non-significant. Furthermore, mood and medical symptom ratings were monitored for safety reasons. Control analyses regarding the mood and medical symptom ratings are reported in the supplemental results (supplemental results & discussion 1). Finally, for consistency with previous work (Chamberlain et al. 2006; den Ouden et al. 2013), we also assessed trial-by-trial behavioural adjustments following rewards and punishments (illustrated in Fig. 1E, details in supplemental results & discussion 2).

Computational modelling

We employed a computational modelling approach to quantify and compare latent mechanisms underlying the task behaviour and particularly the effects of methylphenidate as a function of working memory capacity. For this, we augmented a previously established model of a simpler variant of this reversal learning task (den Ouden et al. 2013) and assessed effects of methylphenidate on the various parameters in this model. We defined a family of four ‘base’ models that could capture behaviour on this task, in a 2×2 model space. Briefly, models could either contain a monotonically decreasing learning rate (Experience Weighted Attraction: EWA model (den Ouden et al. 2013)) or a learning rate that was allowed to increase or decrease as a function of surprise (RL-Pearce-Hall hybrid model (Li et al. 2011; Piray et al. 2019b)). Additionally, both EWA and hybrid models were also tested with an extension where the value of unchosen options is ‘forgotten’ with a forgetting learning rate (Ito and Doya 2009). We first fitted and compared these 4 ‘base’ models across both drug and placebo sessions, to establish the best model independent of drug (cf. model fitting and comparison, below). We then tested variations of the winning ‘base’ model to assess which parameter was affected by methylphenidate. The equations for all four base models are described in detail in the supplemental materials 8. Here we only present the winning base model and methylphenidate extensions.

The winning Experience Weighted Attraction model (EWA) (Camerer and Ho 1999) is an extended version of a standard reinforcement learning (RL) model. We have previously shown that this model can capture (variability in) perseverative behaviour in a simpler version of the current paradigm (den Ouden et al. 2013). The key feature of this model is the so-called experience-weight parameter, which models the increasing impact of past experience on subsequent decisions. With increasing exposure to each stimulus, its experience-weight increases, resulting in a reluctance to update beliefs about this stimulus. This feature makes the EWA model particularly suitable for modelling reversal learning impairment, effectively embodying a learning rate that reduces over time, thus rendering behaviour less flexible. The EWA model is described by the following equations:

$$\begin{aligned} n_{c,t+1} &= n_{c,t} \times \rho + 1 \\ V_{c,t+1} &= (V_{c,t} \times \varphi \times n_{c,t} + \lambda_t) / n_{c,t+1} \\ V_{\sim c,t+1} &= V_{\sim c,t} \end{aligned} \quad (1)$$

where $n_{c,t}$ is the experience-weight of choice c on trial t , which is updated on every trial, using the experience decay factor ρ . The expected value of choice c on trial t , $V_{c,t}$, is updated by integrating the feedback, $\lambda_t \in \{-1, 1\}$; the decay

factor for previous payoffs (inverse learning rate), φ ; and the experience-weight $n_{c,t}$. Initially, the effective learning rate for each choice is high, and by experiencing a choice, the experience-weight increases, resulting in a reluctance to update the stimulus value based on new outcomes.

In the current three choice option task, it is more likely that an option remains unchosen on consecutive trials. Thus, the value of an unchosen option may be ‘forgotten’. This is reflected by the fact that the winning model EWA + F was augmented with a forgetting rate α_f :

$$V_{\sim c,t+1} = (1 - \alpha_f)V_{\sim c,t} \quad (2)$$

For $\alpha_f = 0$, the model is equivalent to the base EWA model (Eq. 1), and for more positive values, the value of the unselected option will converge to the initial value faster.

For all models, to select an action based on the computed values, a soft-max function was employed to calculate the probability of each choice.

$$p(c_t = i) = \frac{\exp(\beta V_{c=i,t})}{\sum_j \exp(\beta V_{c=j,t})} \quad (3)$$

Here, β is the inverse temperature parameter. The $j \in \{1, 2, 3\}$ contains the possible actions.

We took this winning EWA + F base model as the basis to assess what mechanism drove the behavioural difference between placebo and methylphenidate sessions. For this, we built a collection of models where we allowed each parameter in turn to be estimated separately for the methylphenidate and placebo sessions (cf. Swart et al. 2017) and then assessed which of these models (or the winning base model) was the best explanation of our data.

Model fitting and comparison

All models were fitted to the trial-by-trial choices of each participant using Hierarchical Bayesian Inference (HBI) for concurrent model comparison, parameter estimation, and inference at the population level (Piray et al. 2019a). This approach has important advantages for both parameter estimation and model comparison, as parameters estimated by the HBI show smaller errors compared to other methods, while model comparison by HBI is robust against outliers and is not biased towards overly simplistic or complex models. The winning model was selected based on the protected exceedance probability (Rigoux et al. 2014), and we also report model frequency. As described above, we report two sets of model comparison. First, we established the winning ‘base’ model where we fitted data across both sessions (methylphenidate and placebo) to establish the overall best model to describe the data. We then extended the winning base model such that we allowed each parameter in turn to

be differentially estimated for methylphenidate and placebo. In the supplemental materials, we include two further control analyses to verify the assumption that methylphenidate affected only one parameter (c.f. supplemental results & discussion 5). For further details on model fitting and comparison, see supplemental methods 9.

Model validation

Model comparison evaluates whether a winning model is better than other models using an estimation of model evidence, which evaluates goodness of fit relative to model complexity. However, a winning model is not necessarily a good model. A good model should be able to regenerate key features of the original data (Wilson and Collins 2019), which we assessed through simulations. Using each participant's estimated parameters, 100 artificial agents are simulated playing the task, and their choices averaged to represent each individual participant's behaviour. Trial-by-trial simulated data were re-analysed in order to assess whether they captured the key effects of interest.

Parameter inference

Finally, we assessed the nature of the effect of methylphenidate in the winning model. We used a *t*-test to establish whether there was a significant difference in parameter estimates under methylphenidate vs placebo and correlated the drug-induced change in parameter estimates to the covariates of interest, working memory span, and trait impulsivity (see below), using a Spearman correlation. Last, we assessed whether methylphenidate-induced changes in parameter estimates predicted the methylphenidate-induced change in raw behaviour. For all parameter analyses, we extracted subject level parameters from the first, non-hierarchical, estimation step. This is to prevent bias (specifically one magnifying the difference between drug conditions) that could result from the hierarchical model fitting procedure (Piray et al. 2019a).

Covariate analyses: working memory capacity and trait impulsivity

Two covariates, the listening span test total score (Dane-man and Carpenter 1980; Salthouse and Babcock 1991) and Barratt impulsiveness scale (BIS-11; (Patton et al. 1995)) were included in the main analyses, as (preregistered) putative proxies of inter-individual variability in baseline dopamine function. These measures have been shown with PET to correlate positively to dopamine function (Cools et al. 2008; Landau et al. 2009) and have been shown to predict dopaminergic drug effects (Kimberg et al. 1997; Kimberg and D'Esposito 2003; Frank and Claus 2006; Cools and

D'Esposito 2011; van der Schaaf et al. 2014). For detailed descriptions of these measures, see supplemental methods 7.

Results

Data quality assessment

All 102 healthy participants completed the probabilistic reversal learning task in two sessions and were included in the analysis. No participants chose the same response option throughout. Participants made very few responses with RT < 200 ms (mean (st.d.) 0.2 (0.6)% of trials, range 0–4%), and these trials were excluded for the basic analyses. For computational modelling, analyses were repeated with and without these trials.

Behavioural analyses

Choice accuracy

Participants successfully learned the three-option PRL task. People overall learnt to select the rewarded and avoid the punished stimulus (Intercept: $F(1,99) = 1515.1$, $p < 0.001$, $\eta^2 = 0.94$; Fig. 1C). However, choice accuracy was lower during the reversal phase than during the acquisition phase (Phase: $F(1,99) = 19.7$, $p < 0.001$, $\eta^2 = 0.17$, c.f. Figure 1D). There was no evidence of a differential preference to either fail to select the previously punished stimulus or fail to 'let go' of the previously rewarded stimulus (Valence \times Phase: $F(1,99) = 0.3$, $p = 0.6$, $\eta^2 = 0.003$), indicating that there was no difference in the reversal phase in terms of the degree to which people stuck to the previously rewarded stimulus relative to the degree to which they avoided the previously punished stimulus.

Methylphenidate did not consistently affect either overall performance (Drug: $F(1,99) = 3.3$, $p = 0.074$, $\eta^2 = 0.032$), differential learning during acquisition and reversal (Phase \times Drug: $F(1,99) = 0.6$, $p = 0.43$, $\eta^2 = 0.006$), nor, importantly, reward-based perseveration relative to punishment-based avoidance (Phase \times Valence \times Drug: $F(1,99) < 0.01$, $p = 0.96$, $\eta^2 < 0.001$). However, methylphenidate affected performance differentially during acquisition versus reversal as a function of WM span (Phase \times Drug \times WM span: $F(1,99) = 7.1$, $p = 0.009$, $\eta^2 = 0.067$, Fig. 2C), which even weakly interacted with Valence (Phase \times Valence \times Drug \times WM span: $F(1,99) = 5.4$, $p = 0.022$, $\eta^2 = 0.052$). We followed up this interaction in two ways: First, to understand which factor drove this four-way interaction, we broke it down into simple effects for each of the factors (reported in detail in Supplemental results & discussion 3). The first key

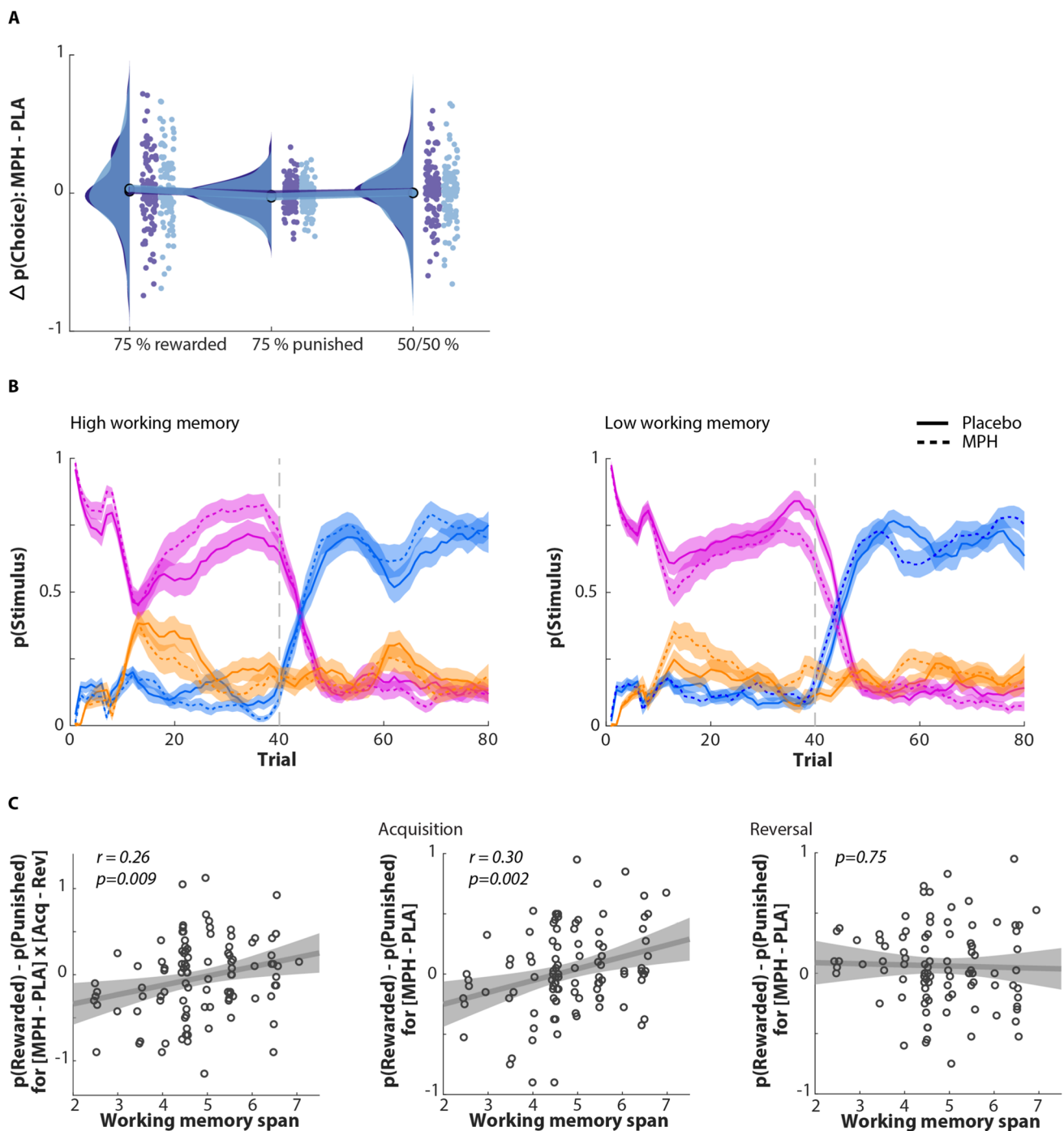


Fig. 2 Effects of methylphenidate on reversal learning task performance. **A** No significant main effect of methylphenidate on average probability of stimulus selection. Distribution of difference of stimulus selection probability between two sessions (MPH-placebo) for acquisition and reversal phase is demonstrated in dark and light blue, respectively. Methylphenidate did not consistently affect either overall learning or differential learning during acquisition and reversal. **B** Trial-by-trial averaged probability of selection of each stimulus (median split based on WM span). Left panel: high WM group ($n=48$), probability of ‘rewarded’ stimulus selection increased under methylphenidate (dash line) in comparison to placebo (solid line) during the acquisition phase.

Right panel: low WM group ($n=54$), probability of ‘rewarded’ stimulus selection decreased under methylphenidate in comparison to placebo during the acquisition phase. A sliding window with 5-trial width is used for smoothing. **C** Methylphenidate effects predicted by WM span. Methylphenidate increased the accuracy of selecting ‘rewarded’ option vs ‘punished’ option in acquisition phase more than reversal phase for high WM span participants yet decreased it for low WM span participants ($r=0.26$, $p=0.009$). By splitting up the MPH effect by phase factor, during the acquisition, middle panel, participants with high WM improved under methylphenidate ($r=0.30$, $p=0.002$), but during the reversal, right panel, there is no significant interaction

observation was that methylphenidate affected learning during the acquisition ($F(1,99) = 9.8, p = 0.002$), but surprisingly not during the reversal phase of the task ($F(1,99) = 0.1, p = 0.75$). During the acquisition phase, people with high working memory improved under methylphenidate, while people with low working memory performed more poorly (Fig. 2B). In contrast, during reversal, there was no interaction of performance with drug and working memory span. The second key observation, from post-hoc simple effects as a function of valence, was that methylphenidate strongly affected the ability to learn to select the rewarded stimulus ($F(1,99) = 7.6, p = 0.007$), and while effects of methylphenidate were in the same direction for the ability to learn to avoid the punished stimulus, these were significantly weaker and by themselves only a trend ($F(1,99) = 3.7, p = 0.057$). Interestingly, there was a significant effect for the probability to select the neutral stimulus ($F(1,99) = 5.4, p = 0.022$). Summarising, these results show that under methylphenidate, during the acquisition phase, people with high working memory were significantly likely to select the mostly rewarded stimulus while significantly less likely to select the neutral stimulus, leaving the punished stimulus not significantly affected. This suggests that in high working memory participants,

methylphenidate aids to dissociate between the best and second best option.

Finally, there were some weak (trend) lower level interaction effects that were however qualified by the key higher-order interaction described above (Drug \times WM span: $F(1,99) = 3.6, p = 0.06, \eta^2 = 0.036$). There was no significant effect of BIS on performance, nor in interaction with methylphenidate (of interest: Phase \times Drug \times BIS: $F(1,99) = 0.07, p = 0.8, \eta^2 = 0.001$; Phase \times Valence \times Drug \times BIS: $F(1,99) < 0.01, p = 0.9, \eta^2 < 0.001$; all other $p > 0.1$).

To control for potential confounding factors, we repeated this ANOVA by including the confound variables gender and test order (between subject factors) and age (covariate). Including these confound variables did not alter the significance of the observed effects although there was an effect of age (but this did not interact with the effect of interest; for details, see the supplemental results & discussion 1).

Model comparison

We compared our previously established EWA model to an extended version that included a forgetting factor, following from the extension of stimulus space (3-option design) along with the RL-Pearce-Hall hybrid models that allowed

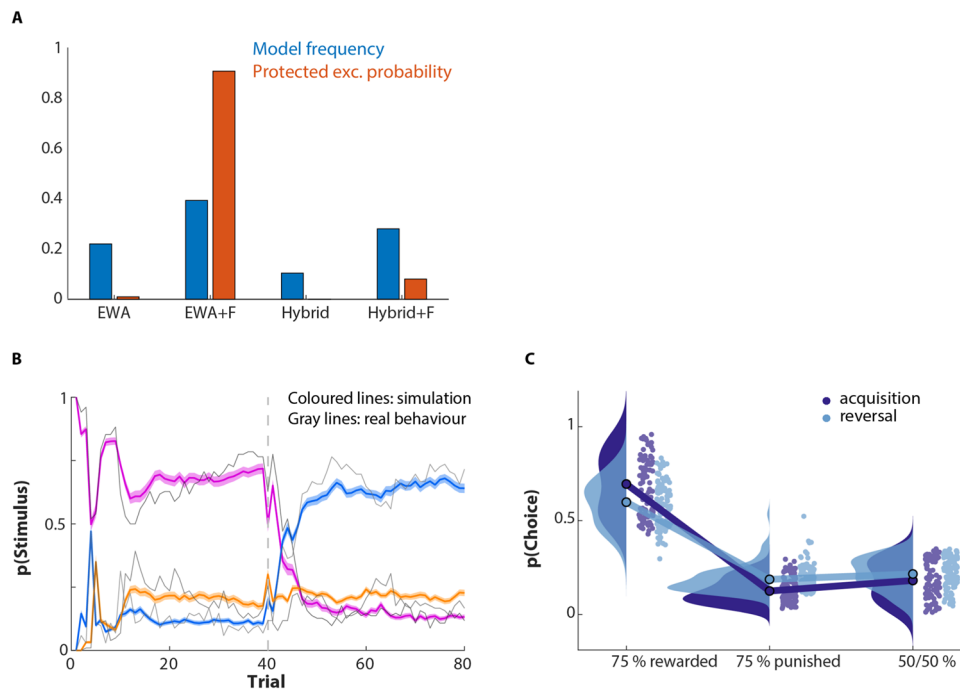


Fig. 3 Model fitting and simulation result for PRL task modelling. **A** Model comparison on base models. Model frequency and protected exceedance probability indicate that the EWA + F model (EWA with forgetting rate for unchosen options) provides the best description of the data (PXP = 0.91). **B** Trial-by-Trial simulated choice. Model simulations of the winning base model verify that the EWA + F model captures the behavioural data (grey lines indicate average raw data).

C Simulated average choice probability. Distribution of stimulus selection probability for acquisition and reversal phase is demonstrated in dark and light blue, respectively (compare with Fig. 1D). The simulated data for EWA + F model qualitatively and quantitatively replicate the participants' behaviour and regenerate key features of the data

Table 1 Model evidence for base model and methylphenidate model families and parameter estimates for winning models in each family (see supplementary table 1 for full details for all models)

Base model family		Param.*	Constraint	Median	Range (25–75%)	pxp _i	p (m _i data, m)
Model							
	EWA					0.01	0.22
	EWA + F					0.91	0.40
		φ	$0 \leq \varphi \leq 1$	0.77	0.29–0.87		
		ρ	$0 \leq \rho \leq 1$	0.63	0.27–0.83		
		β	$0 \leq \beta \leq +\infty$	4.23	3.11–7.88		
		α_f	$0 \leq \alpha_f \leq 1$	0.35	0.02–0.68		
	Hybrid					0.0	0.10
	Hybrid + F					0.08	0.28
Methylphenidate model family							
	EWA + F					0.02	0.25
	EWA + F + Δ_ρ					0	0.13
	EWA + F + Δ_φ					0.98	0.42
		$\varphi_{Placebo}$	$0 \leq \varphi_{Placebo} \leq 1$	0.70	0.26–0.86		
		φ_{MPH}	$0 \leq \varphi_{MPH} \leq 1$	0.70	0.33–0.84		
		ρ	$0 \leq \rho \leq 1$	0.56	0.21–0.79		
		β	$0 \leq \beta \leq +\infty$	4.46	3.13–7.94		
		α_f	$0 \leq \alpha_f \leq 1$	0.22	0.01–0.62		
	EWA + F + Δ_{α_f}					0	0.12
	EWA + F + Δ_β					0	0.08

*A weakly informative Gaussian prior was used for all parameters ($x \sim N(\mu, \sigma^2)$ where the mean value $\mu = 0$ and the variance $\sigma^2 = 10$). According to theoretical constraints of parameters, sigmoid or exponential transformations are applied.

for adaptive, prediction error–based changes in learning rate. Model comparison showed convincing evidence in favour of the EWA + F model (protected exceedance probability: 0.91, Fig. 3A). See Table 1 and supplemental results & discussion 4 for parameter estimates for all four base models.

To capture the effects of methylphenidate, we allowed each of the 4 parameters of this winning model to be affected by methylphenidate in turn, by fitting that parameter separately for methylphenidate and placebo sessions. We then compared all 4 ‘drug’ models plus the ‘baseline’ model without an effect of methylphenidate. Model comparison showed that the winning model allows methylphenidate to affect the inverse learning rate parameter φ (protected exceedance probability: 0.98, see Fig. 4B). Parameter estimates for the methylphenidate models and model comparison statistics are reported in Table 1 and supplemental results & discussion 4. As described above, we assumed that methylphenidate affected only a single parameter (i.e. computational mechanism). We validate this assumption in the supplemental materials (see supplemental results & discussion 5).

Model validation

Model simulation is essential to evaluate the model’s ability to regenerate key features of the data. In order to examine the reproducibility of the winning model, we generated

data using the winning model. Our simulated data qualitatively and quantitatively replicate the participants’ behaviour (Figs. 3B–C and 4C–D). Again, in the simulated data, there is a significant interaction of Phase \times Drug \times WM span ($F(1,99) = 5.1, p = 0.026, \eta^2 = 0.049$) (Fig. 4D). Furthermore, breaking down this interaction into simple effects by phase, our observations from raw data were replicated. Methylphenidate improved initial learning in high-WM span participants while reducing performance in low-WM span participants ($F(1,99) = 5.1, p = 0.027, \eta^2 = 0.049$). Again, there was no significant Drug \times WM span interaction during the reversal phase ($F(1,99) = 1.7, p = 0.3, \eta^2 = 0.011$).

Parameter inference

Model EWA + F, which allowed for a different value for the inverse learning rate φ on vs off methylphenidate, was convincingly the best model. However, there was no significant difference between $\varphi_{Placebo}$ and φ_{MPH} when assessed across all individuals ($t = -0.4, p = 0.7$). This indicates strong evidence for a change in φ following drug intake, but the sign of this change was variable over participants. Indeed, mirroring the raw behavioural results, the difference between $\varphi_{Placebo}$ and φ_{MPH} correlated with WM span ($r = 0.20, p = 0.043$, Fig. 4E). There was a strong correlation between methylphenidate-induced changes in inverse learning rate (Δ_φ) and the methylphenidate-induced change in raw performance

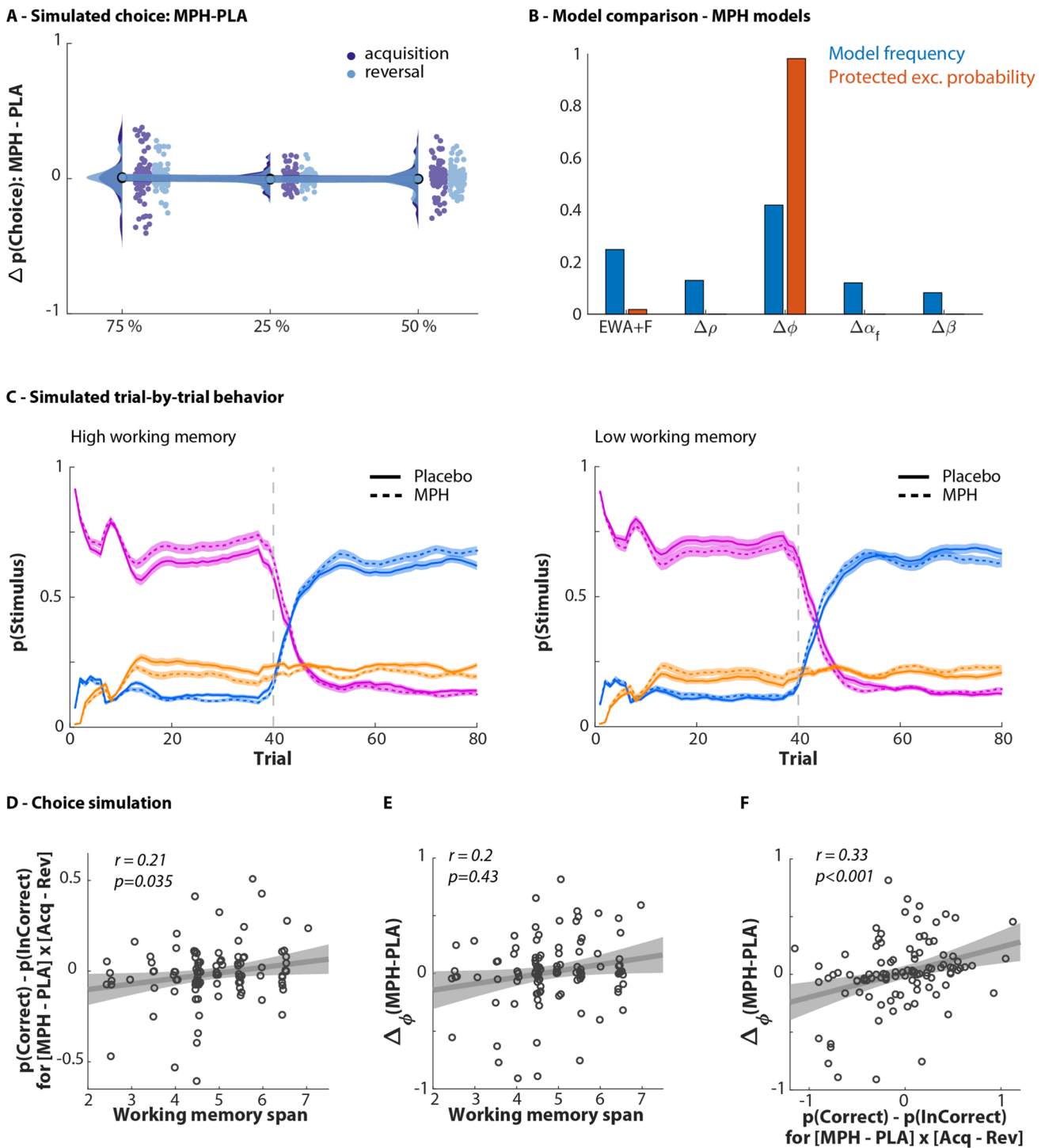


Fig. 4 Modelling the effects of methylphenidate. **A** Simulated choice. Simulated data replicated observed behaviour for no main effect of methylphenidate (see Fig. 2A). **B** Model comparison on methylphenidate models. Model frequency and protected exceedance probability (PXP=0.98) indicate that the EWA+F that allows for differential learning rates under methylphenidate and placebo best captures the data. Model validation: C-F. **C** Simulated trial-by-trial behaviour. Simulated data replicate observed behaviour for high WM and low WM participants (see Fig. 2B). **D** Choice simulation. The simulated data using winning model regenerate quantitative characteristics of

data, particularly a positive effect of methylphenidate on performance for high WM participants (see Fig. 2C). **E** Methylphenidate changes inverse learning rate as a function of working memory span. The difference in inverse learning rates under methylphenidate vs placebo ($\Delta\phi$) covaries with WM span ($r=0.2$; $p=0.043$). Methylphenidate increases ϕ in high WM participants, where decreases it in low WM participants. **F** Methylphenidate-induced effect on raw performance scores. The methylphenidate-induced change in inverse learning rate ($\Delta\phi$) is correlated with methylphenidate-induced change in raw performance ($r=0.33$; $p<0.001$)

metrics ($r=0.32$, $p<0.001$, Fig. 4F). Thus, methylphenidate administration increased performance in high working memory participants, which was captured by an increase in the inverse learning rate φ , thus effectively a decrease in learning rate. In contrast, methylphenidate administration decreased performance in low working memory participants, which was captured by a decrease in inverse learning rate φ . There was no correlation between the methylphenidate-induced changes in inverse learning rate ($\Delta\varphi$) and trait impulsivity ($r=-0.02$, $p=0.8$).

Optimal learning rate analysis

Compared with previous studies in which tasks were employed with two response options, we observed that performance in the acquisition phase was substantially worse, while performance during the reversal phase was much better in our current 3-option version (cf. den Ouden et al. 2013). Concomitant with this, values of φ were much higher, i.e. learning rate was lower, in the current dataset than in our previous dataset. We therefore performed a supplemental analysis to compare the optimal learning rates across the two different versions of the task to assess whether this changed learning rate across paradigms was adaptive. In short, the 3-option version of the task had a lower optimal learning rate. This observation can be understood when realising that optimal performance on this probabilistic task required participants to dissociate a 50% reward option from a 75% reward. Integrating information over too short a time window (i.e. a high learning rate) would have made it more difficult to correctly dissociate between these two options. In line with this, the observed decrease in learning rate in high WM span participants under methylphenidate was adaptive, as the effective learning rates moved closer to the optimal learning rate. For details, see supplemental results & discussion 6.

Discussion

This study aimed to uncover the causal role of dopamine in perseverative behaviour during probabilistic reversal learning. To this end, a large sample of participants ($n=102$) performed a novel 3-choice PRL task both on and off catecholamine transporter blocker methylphenidate. Contrary to predictions, methylphenidate did not consistently affect perseverative behaviour in this novel task, as indexed by an absence of change in reversal performance. In contrast, methylphenidate altered performance in the acquisition phase, in a manner that depended on individual variability in working memory span. Specifically, methylphenidate increased the inverse learning rates in participants with a higher working memory span. In other words, in high WM individuals, methylphenidate reduced the degree to which

values were updated following any single outcome, which made learning more robust in the probabilistic context of the task, thereby improving initial learning.

Methylphenidate effects on value learning versus perseveration

We set out in this study to assess the computational mechanism by which catecholamine blockade affects reversal performance. Specifically, we asked whether reversal deficits were due to a failure to learn to approach a previously punished stimulus or to ‘let go’ of a previously rewarded stimulus. Perhaps a surprising finding in the current study is that acute catecholamine reuptake blockade did not affect reversal performance and perseveration at all. This is particularly surprising given the considerable literature concerning the role of dopamine in habitual actions (Daw et al. 2005; Everitt and Robbins 2005; Balleine and O’Doherty 2010) and perseverative behaviour (Cools et al. 2001; Rutledge et al. 2009), particularly direct findings of genetic variability in the *DAT1* genotype (den Ouden et al. 2013) and methylphenidate-induced reversal impairments (Clatworthy et al. 2009). While it is possible that this observed discrepancy between the previous literature and current study reflects a true non-replication in this large sample ($n=102$, vs most previous studies $n=20-40$), a perhaps more likely possibility is that, by introducing a neutral choice option, we have significantly altered the nature of the paradigm.

Indeed, as presented in the supplemental analyses, across participants, we observed that compared to the previously used 2-choice PRL paradigms, the learning rate was significantly lower. This reduction in effective learning rate provides an important clue as to why the observed effects in the current study are particularly obvious in the acquisition phase. By including the 50/50 choice option, the task was rendered much more difficult as optimal performance now required dissociation of two choice options (50/50 vs 75/25) that were much closer in terms of feedback than the previous (70/30 vs 30/70) dissociation that had to be learnt. To be able to make this dissociation, one needs to integrate information over a longer window, which is exactly what a reduced learning reflects. Indeed, this adjustment is adaptive, as optimal learning rates in the current paradigm are much lower than for the previous version of the task. In contrast, performance during the reversal phase was much more robust in the current paradigm.

This increased task difficulty may have had the unanticipated effect that this (difficult) initial learning was less robust across participants and became more sensitive to (e.g. drug) manipulations. We propose that dopamine affects both initial value learning, likely through ventral striatal prediction error like RL mechanisms that affect learning rate, but also affects longer-term ‘stamping in’ of responses and

habit formation, through dorsal-striatal habit systems. Discrepancies in findings across studies, then, might reflect the relative sensitivity of various tasks tapping into these mechanisms. Thus, the current paradigm increased difficulty of initial learning but was also associated with less vulnerable reversal performance, while the reverse may be true for the 2-option task. It is unclear why performance during the reversal phase was so much better than in the 2-option task—perhaps the task structure was more obvious to participants. Regardless, the absence of methylphenidate effects on reversal performance therefore unfortunately did not allow us to further disentangle putative dopaminergic mechanisms of reward-based perseveration versus a failure to overcome learned avoidance.

Mechanisms of baseline-dependent effects of methylphenidate

The finding that effects of methylphenidate on behaviour vary as a function of working memory capacity was consistent with our preregistered hypothesis. We posit two possible explanations for this effect. First, working memory span has been shown to correlate with striatal dopamine synthesis capacity (Cools et al. 2008; Landau et al. 2009). Given that methylphenidate acts by blocking the dopamine (and noradrenaline) transporter, it is likely that the effect of methylphenidate on catecholamine-dependent function is a function of dopamine synthesis capacity and subsequent release. While under placebo conditions, release and reuptake are in balance in both high and low synthesis capacity subjects, administration of methylphenidate could disturb this balance differentially. Specifically, if individuals with higher working memory capacity have higher release of dopamine (Cools et al. 2008; Landau et al. 2009), then methylphenidate might increase tonic levels of dopamine, paradoxically leading to reduced sensitivity to individual bursts and thus a reduced learning rate. In contrast, low working memory participants with low synthesis capacity may have very sensitive post-synaptic dopamine function, and blockade of transporters may increase the duration of post-synaptic impact of dopamine bursts, thereby effectively increasing the learning rate. A recent dopamine PET study indeed demonstrated disproportionate sensitivity of participants with low dopamine synthesis capacity to methylphenidate-related increases in reward impact on choice (Westbrook et al. 2020).

Alternatively, the interaction between methylphenidate effects on learning and baseline working memory capacity might reflect a modulation of interactions between working memory and reinforcement learning strategies (Collins et al. 2017). Specifically, Collins and Frank have recently established that within individuals, relative reliance on reinforcement learning versus working memory strategies varies with working memory load (Collins and Frank 2012). By

analogy, we hypothesise that this balance in any given task may vary across individuals as a function of their working memory capacity. In short, if you have a lower span, you may shift sooner to RL strategies. Neurally, methylphenidate may act on striatal levels of dopamine, as suggested above, but may also affect frontal functioning, through blockade of noradrenaline transporters in the frontal cortex (Volkow et al. 2001, 2012; Arnsten and Dudley 2005; Berridge et al. 2006; Berridge and Devilbiss 2011; Kodama et al. 2017). The relative balance of the effect of methylphenidate on either striatal (putatively RL) mechanisms versus putative direct frontal modulation on working memory functioning may differ between individuals with high versus low baseline working memory capacity, explaining the differential effects observed.

A final speculation is that methylphenidate plays a role via its action on either dopamine or noradrenaline transmission by affecting our ability to optimise the learning rate given the volatility of the environment (Nassar et al. 2010; Muller et al. 2019). This hypothesis concurs with the results of our other experiment in the same individuals in which we employed a task explicitly designed to assess effects on learning as a function of the volatility of outcome contingencies (cf. Figure 3; Cook et al. 2019). In this learning task, methylphenidate adaptively lowered the learning rate in stable versus changeable environments. In the current experiment, the learning rate decrease in high capacity participants moved it closer to the optimal learning rate for current the task and was indeed associated with an increase in initial performance, due to a better ability to distinguish between the best (rewarded) and second-best (neutral) option. We do note that this ‘meta-learning’ interpretation should be taken with caution, because the current paradigm with its single reversal was not optimised to answer this question and the models in which we allowed the learning rate to fluctuate according to the size of the prediction errors did not perform better than the winning model in which the learning rate was only allowed to be reduced over time.

Conclusion

The present study was set up to test the specific hypothesis, derived from our previous dopamine genetic study, that administration of methylphenidate would alter probabilistic reversal learning by changing the reliance on prior reward. To test this hypothesis, a novel reversal task was employed with three-choice options. Surprisingly, results revealed no effects on the reversal phase. However, an effect of methylphenidate surfaced already in the initial acquisition phase. In line with prior studies, this effect was not unidirectional across participants, but varied with individual differences in baseline working memory capacity: Methylphenidate

improved performance and reduced the learning rate to a greater degree in participants with higher working memory capacity. We hypothesise that the increased demands for learning in this 3-option task brought to the surface an effect of methylphenidate on learning rather than on flexibility.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00213-021-05974-w>.

Acknowledgements We thank Dr. Monique Timmer and Dr. Peter Mulders for the medical cover and Dr. Sean James Fallon for the advice on setting up the MPH study and all our participants for taking part in this study.

Author contribution HEMdO, RC, and JCS designed the paradigm; JLC, JCS, RC, and HEMdO set up the MPH study; JCS, MIF, JLC, and HEMdO collected the data; DEMG provided medical cover and assistance; MRK and HEMdO analysed the data; MRK, AHV, MNA, RC, and HEMdO have discussed analyses and modelling results. MRK, RC, and HEMdO wrote the manuscript; all authors edited/revised the manuscript.

Funding James S. McDonnell Foundation, James McDonnell scholar award: Roshan Cools.

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Vici Award 453–14-015: Roshan Cools.

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Vidi Grant 240–01-170: Hanneke EM den Ouden.

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Veni Grant 451–11-004: Hanneke EM den Ouden.

ZonMw, 92,003,576: Dirk EM Geurts.
H2020 European Research Council, ERC Starting Grant 757,583: Jennifer L Cook.

University of Birmingham, Birmingham Fellows Programme: Jennifer L Cook.

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Research talent grant 406-14-028: Jennifer C Swart.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. All authors express that there is no conflict of interest to disclose.

Data availability The collected data and analysis scripts of the current study are available in the Donders Institute Data repository, https://data.donders.ru.nl/collections/di/dccn/DSC_3017031.02_887.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arnsten AFT, Dudley AG (2005) Methylphenidate improves prefrontal cortical cognitive function through $\alpha 2$ adrenoceptor and dopamine D1 receptor actions: relevance to therapeutic effects in attention deficit hyperactivity disorder. *Behav Brain Funct* 1:1–9. <https://doi.org/10.1186/1744-9081-1-2>
- Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35:48–69. <https://doi.org/10.1038/npp.2009.131>
- Berridge CW, Devilbiss DM (2011) Psychostimulants as cognitive enhancers: the prefrontal cortex, catecholamines, and attention-deficit/hyperactivity disorder. *Biol Psychiatry* 69:e101–e111. <https://doi.org/10.1016/j.biopsych.2010.06.023>
- Berridge CW, Devilbiss DM, Andrzejewski ME et al (2006) Methylphenidate preferentially increases catecholamine neurotransmission within the prefrontal cortex at low doses that enhance cognitive function. *Biol Psychiatry* 60:1111–1120. <https://doi.org/10.1016/j.biopsych.2006.04.022>
- Boulougouris V, Castañé A, Robbins TW (2009) Dopamine D2/D3 receptor agonist quinpirole impairs spatial reversal learning in rats: Investigation of D3 receptor involvement in persistent behavior. *Psychopharmacology* 202:611–620. <https://doi.org/10.1007/s00213-008-1341-2>
- Buckholtz JW, Treadway MT, Cowan RL et al (2010) Dopaminergic network differences in human impulsivity. *Science* (80-) 329:532. <https://doi.org/10.1126/science.1185778>
- Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. *Econometrica* 67:827–874
- Chamberlain S, Muller U, Blackwell AD et al (2006) Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* (80-) 311:861–863
- Chudasama Y, Robbins TW (2006) Functions of frontostriatal systems in cognition: comparative neuropsychopharmacological studies in rats, monkeys and humans. *Biol Psychol* 73:19–38. <https://doi.org/10.1016/j.biopsycho.2006.01.005>
- Clarke HF, Hill GJ, Robbins TW, Roberts AC (2011) Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. *J Neurosci* 31:4290–4297. <https://doi.org/10.1523/JNEUROSCI.5066-10.2011>
- Clatworthy PL, Lewis SJG, Brichard L et al (2009) Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. *J Neurosci* 29:4690–4696. <https://doi.org/10.1523/JNEUROSCI.3266-08.2009>
- Collins AGE, Ciullo B, Frank MJ, Badre D (2017) Working memory load strengthens reward prediction errors. *J Neurosci* 37:4332–4342. <https://doi.org/10.1523/JNEUROSCI.2700-16.2017>
- Collins AGE, Frank MJ (2014) Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol Rev* 121:337–366. <https://doi.org/10.1037/a0037015>
- Collins AGE, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur J Neurosci* 35:1024–1035. <https://doi.org/10.1111/j.1460-9568.2011.07980.x>
- Cook JL, Swart JC, Froböse MI et al (2019) Catecholaminergic Modulation of Meta-Learning *Elife* 8:1–38. <https://doi.org/10.7554/eLife.51439>
- Cools R, Altamirano L, D'Esposito M (2006) Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia* 44:1663–1673. <https://doi.org/10.1016/j.neuropsychologia.2006.03.030>

- Cools R, Barker RA, Sahakian BJ, Robbins TW (2001) Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cereb Cortex* 11:1136–1143
- Cools R, D'Esposito M (2011) Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol Psychiatry* 69:e113–e125. <https://doi.org/10.1016/j.biopsych.2011.03.028>
- Cools R, Gibbs SE, Miyakawa A et al (2008) Working memory capacity predicts dopamine synthesis capacity in the human striatum. *J Neurosci* 28:1208–1212. <https://doi.org/10.1523/JNEUROSCI.4475-07.2008>
- Cools R, Lewis SJG, Clark L et al (2007a) L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology* 32:180–189. <https://doi.org/10.1038/sj.npp.1301153>
- Cools R, Sheridan M, Jacobs E, D'Esposito M (2007b) Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. *J Neurosci* 27:5506–5514. <https://doi.org/10.1523/JNEUROSCI.0601-07.2007>
- Daneman M, Carpenter PA (1980) Individual differences in working memory and reading. *J Verbal Learning Verbal Behav* 19:450–466. [https://doi.org/10.1016/s0022-5371\(80\)90312-6](https://doi.org/10.1016/s0022-5371(80)90312-6)
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8:1704–1711. <https://doi.org/10.1038/nn1560>
- den Ouden HEM, Daw ND, Fernandez G et al (2013) Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* 80:1090–1100. <https://doi.org/10.1016/j.neuron.2013.08.030>
- Dodds CM, Muller U, Clark L et al (2008) Methylphenidate has differential effects on blood oxygenation level-dependent signal related to cognitive subprocesses of reversal learning. *J Neurosci* 28:5976–5982. <https://doi.org/10.1523/JNEUROSCI.1153-08.2008>
- Elliott R, Sahakian BJ, Matthews K et al (1997) Effects of methylphenidate on spatial working memory and planning in healthy young adults. *Psychopharmacology* 131:196–206. <https://doi.org/10.1007/s002130050284>
- Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat Neurosci* 8:1481–1489. <https://doi.org/10.1038/nn1579>
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* (80-) 299:1898–1902. <https://doi.org/10.1126/science.1077349>
- Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* 17:51–72. <https://doi.org/10.1162/0898929052880093>
- Frank MJ, Claus ED (2006) Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113:300–326. <https://doi.org/10.1037/0033-295X.113.2.300>
- Frank MJ, Seeberger LC, O'Reilly RC (2004) By Carrot Or By Stick: Cognitive Reinforcement Learning In Parkinsonism. *Science* (80-) 306:1940–1943. <https://doi.org/10.1126/science.1102941>
- Froböse MI, Swart JC, Cook JL et al (2018) Catecholaminergic modulation of the avoidance of cognitive control. *J Exp Psychol Gen* 147:1763–1781. <https://doi.org/10.1037/xge0000523>
- Geurts DEM, den Ouden HEM, Froböse MI, et al (2021) The role of catecholamines in pavlovian-instrumental transfer. *Manuscr Prep 1*
- Goto Y, Grace AA (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* 8:805–812. <https://doi.org/10.1038/nn1471>
- Groman SM, Lee B, Seu E et al (2012) Dysregulation of D2-mediated dopamine transmission in monkeys after chronic escalating methamphetamine exposure. *J Neurosci* 32:5843–5852. <https://doi.org/10.1523/JNEUROSCI.0029-12.2012>
- Ito M, Doya K (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci* 29:9861–9874. <https://doi.org/10.1523/JNEUROSCI.6157-08.2009>
- Kim JH, Son YD, Kim HK et al (2014) Dopamine D 2/3 receptor availability and human cognitive impulsivity: a high-resolution positron emission tomography imaging study with [¹¹C]raclopride. *Acta Neuropsychiatr* 26:35–42. <https://doi.org/10.1017/neu.2013.29>
- Kimberg DY, D'Esposito M (2003) Cognitive effects of the dopamine receptor agonist pergolide. *Neuropsychologia* 41:1020–1027. [https://doi.org/10.1016/S0028-3932\(02\)00317-2](https://doi.org/10.1016/S0028-3932(02)00317-2)
- Kimberg DY, D'Esposito M, Farah MJ (1997) Effects of bromocriptine on human subjects depend on working memory capacity. *NeuroReport* 8:3581–3585. <https://doi.org/10.1097/00001756-19971100-00032>
- Kimko HC, Cross JT, Abernethy DR (1999) Pharmacokinetics and clinical effectiveness of methylphenidate. *Clin Pharmacokinet* 37:457–470
- Kodama T, Kojima T, Honda Y et al (2017) Oral administration of methylphenidate (ritalin) affects dopamine release differentially between the prefrontal cortex and striatum: a microdialysis study in the monkey. *J Neurosci* 37:2387–2394. <https://doi.org/10.1523/JNEUROSCI.2155-16.2017>
- Landau SM, Lal R, O'Neil JP et al (2009) Striatal dopamine and working memory. *Cereb Cortex* 19:445–454. <https://doi.org/10.1093/cercor/bhn095>
- Lee B, London ED, Poldrack RA et al (2009) Striatal dopamine D2/D3 receptor availability is reduced in methamphetamine dependence and is linked to impulsivity. *J Neurosci* 29:14734–14740. <https://doi.org/10.1523/JNEUROSCI.3765-09.2009>
- Li J, Schiller D, Schoenbaum G et al (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci* 14:1250–1252. <https://doi.org/10.1038/nn.2904>
- Linssen AMW, Sambeth A, Vuurman EFPM, Riedel WJ (2014) Cognitive effects of methylphenidate in healthy volunteers: a review of single dose studies. *Int J Neuropsychopharmacol* 17:961–977. <https://doi.org/10.1017/S1461145713001594>
- Madras BK, Miller GM, Fischman AJ (2005) The dopamine transporter and attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57:1397–1409. <https://doi.org/10.1016/j.biopsych.2004.10.011>
- Montague P, Dayan P, Sejnowski T (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16:1936–1947. <https://doi.org/10.1523/jneurosci.16-05-01936.1996>
- Muller TH, Mars RB, Behrens TE, O'Reilly JX (2019) Control of entropy in neural models of environmental state. *Elife* 8:1–30. <https://doi.org/10.7554/eLife.39404>
- Nassar MR, Wilson RC, Heasly B, Gold JI (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30:12366–12378. <https://doi.org/10.1523/JNEUROSCI.0822-10.2010>
- Niv Y, Montague PR (2009) Theoretical and empirical studies of learning. In: Glimcher PW, Fehr E, Camerer C, Poldrack RA (eds) *Neuroeconomics*, 1st edn. Academic Press, pp 331–351
- Parkinson JA, Olmstead MC, Burns LH et al (1999) Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *J Neurosci* 19:2401–2411
- Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* 51:768–774. [https://doi.org/10.1002/1097-4679\(199511\)51:6%3c768](https://doi.org/10.1002/1097-4679(199511)51:6%3c768)
- Piray P, Dezfouli A, Heskes T et al (2019a) Hierarchical Bayesian inference for concurrent model fitting and comparison for group

- studies. *PLOS Comput Biol* 15:e1007043. <https://doi.org/10.1371/journal.pcbi.1007043>
- Piray P, Ly V, Roelofs K et al (2019b) Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals. *J Neurosci* 39:1445–1456. <https://doi.org/10.1523/JNEUROSCI.1394-18.2018>
- Reeves SJ, Polling C, Stokes PRA et al (2012) Limbic striatal dopamine D2/3 receptor availability is associated with non-planning impulsivity in healthy adults after exclusion of potential dissimulators. *Psychiatry Res-Neuroimaging* 202:60–64. <https://doi.org/10.1016/j.pscychresns.2011.09.011>
- Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies-Revisited. *Neuroimage* 84:971–985. <https://doi.org/10.1016/j.neuroimage.2013.08.065>
- Robbins TW, Cador M, Taylor JR, Everitt BJ (1989) Limbic-striatal interactions in reward-related processes. *Neurosci Biobehav Rev* 13:155–162. [https://doi.org/10.1016/S0149-7634\(89\)80025-9](https://doi.org/10.1016/S0149-7634(89)80025-9)
- Rutledge RB, Lazzaro SC, Lau B et al (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *J Neurosci* 29:15104–15114. <https://doi.org/10.1523/JNEUROSCI.3524-09.2009>
- Salthouse TA, Babcock RL (1991) Decomposing adult age differences in working memory. *Dev Psychol* 27:763–776. <https://doi.org/10.1037/0012-1649.27.5.763>
- Schultz W (2016) Dopamine reward prediction-error signalling: a two-component response. *Nat Rev Neurosci* 17:183–195. <https://doi.org/10.1038/nrn.2015.26>
- Smith AG, Neill JC, Costall B (1999) The dopamine D3/D2 receptor agonist 7-OH-DPAT induces cognitive impairment in the marmoset. *Pharmacol Biochem Behav* 63:201–211. [https://doi.org/10.1016/S0091-3057\(98\)00230-5](https://doi.org/10.1016/S0091-3057(98)00230-5)
- Swainson R, Rogers RD, Sahakian BJ et al (2000) Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. *Neuropsychologia* 38:596–612. [https://doi.org/10.1016/S0028-3932\(99\)00103-7](https://doi.org/10.1016/S0028-3932(99)00103-7)
- Swart JC, Froböse MI, Cook JL et al (2017) Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. *Elife* 6:1–36. <https://doi.org/10.7554/eLife.22169>
- Taghzouti K, Louilot A, Herman JP et al (1985) Alternation behavior, spatial discrimination, and reversal disturbances following 6-hydroxydopamine lesions in the nucleus accumbens of the rat. *Behav Neural Biol* 44:354–363. [https://doi.org/10.1016/S0163-1047\(85\)90640-5](https://doi.org/10.1016/S0163-1047(85)90640-5)
- Van Der Schaaf ME, Fallon SJ, Ter Huurne N et al (2013) Working memory capacity predicts effects of methylphenidate on reversal learning. *Neuropsychopharmacology* 38:2011–2018. <https://doi.org/10.1038/npp.2013.100>
- van der Schaaf ME, van Schouwenburg MR, Geurts DEM et al (2014) Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. *Cereb Cortex* 24:633–642. <https://doi.org/10.1093/cercor/bhs344>
- Volkow ND, Wang G-J, Fowler JS et al (2002) Relationship between blockade of dopamine transporters by oral methylphenidate and the increases in extracellular dopamine: Therapeutic implications. *Synapse* 43:181–187. <https://doi.org/10.1002/syn.10038>
- Volkow ND, Wang G-J, Fowler JS et al (2001) Therapeutic Doses of Oral Methylphenidate Significantly Increase Extracellular Dopamine in the Human Brain. *J Neurosci* 21:1–5. <https://doi.org/10.1523/JNEUROSCI.21-02-j0001.2001>
- Volkow ND, Wang GJ, Tomasi D et al (2012) Methylphenidate-elicited dopamine increases in ventral striatum are associated with long-term symptom improvement in adults with attention deficit hyperactivity disorder. *J Neurosci* 32:841–849. <https://doi.org/10.1523/JNEUROSCI.4461-11.2012>
- Westbrook A, van den Bosch R, Määttä JI et al (2020) Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science* (80-) 367:2–1366. <https://doi.org/10.1126/science.aaz5891>
- Wilson RC, Collins AGE (2019) Ten simple rules for the computational modeling of behavioral data. *Elife* 8:1–33. <https://doi.org/10.7554/eLife.49547>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.