



## ARTICLE

# Attention-deficit/hyperactivity disorder and the explore/exploit trade-off

Merideth A. Addicott<sup>1</sup>, John M. Pearson<sup>2</sup>, Julia C. Schechter<sup>2</sup>, Jeffrey J. Sapyta<sup>2</sup>, Margaret D. Weiss<sup>1</sup> and Scott H. Kollins<sup>2</sup>

The ability to maximize rewards and minimize the costs of obtaining them is vital to making advantageous explore/exploit decisions. Exploratory decisions are theorized to be greater among individuals with attention-deficit/hyperactivity disorder (ADHD), potentially due to deficient catecholamine transmission. Here, we examined the effects of ADHD status and methylphenidate, a common ADHD medication, on explore/exploit decisions using a 6-armed bandit task. We hypothesized that ADHD participants would make more exploratory decisions than controls, and that MPH would reduce group differences. On separate study days, adults with ( $n = 26$ ) and without ( $n = 23$ ) ADHD completed the bandit task at baseline, and after methylphenidate or placebo in counter-balanced order. Explore/exploit decisions were modeled using reinforcement learning algorithms. ADHD participants made more exploratory decisions (i.e., chose options without the highest expected reward value) and earned fewer points than controls in all three study days, and methylphenidate did not affect these outcomes. Baseline exploratory choices were positively associated with hyperactive ADHD symptoms across all participants. These results support several theoretical models of increased exploratory choices in ADHD and suggest the unexplained variance in ADHD decisions may be due to less value tracking. The inability to suppress actions with little to no reward value may be a key feature of hyperactive ADHD symptoms.

*Neuropsychopharmacology* (2021) 46:614–621; <https://doi.org/10.1038/s41386-020-00881-8>

## INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is characterized by symptoms of inattention, hyperactivity, and impulsivity that negatively affect psychosocial functioning, education, and self-esteem [1–3]. While ADHD is typically considered to be a childhood-onset disorder that can persist into adulthood [4], ADHD may also have an adult onset [5] and ultimately affects ~2.5% of the U.S. adult population [6]. More than 90% of adults with ADHD report moderate to severe lifetime impairment related to their ADHD symptoms [7].

ADHD is known to cause cognitive problems in measures of response inhibition, vigilance, and working memory [8–10]. For example, individuals with ADHD tend to show more reaction time variability and make more omission errors on sustained attention tests, suggesting impaired vigilance [9]. ADHD has also been associated with altered motivation and reward sensitivity, such as a preference for immediate over delayed rewards [11–14]. In particular, reinforcement contingencies have a stronger effect on improving attentional performance among individuals with ADHD, suggesting low levels of intrinsic motivation and/or elevated reward thresholds [15].

Both cognitive performance and motivated behavior have been linked to catecholamine signaling, primarily dopamine (DA), in the mesocorticolimbic brain pathway [16]. DA plays a role in responding to cues that predict reward, which motivates exploration or exploitation in search of that reward; DA also modulates flexible cognitive control processes that are sensitive to changes in motivation [16]. ADHD is associated with lower DA transporter availability and lower D2/D3 receptor densities in the

mesocorticolimbic pathway [17–19]. Methylphenidate (MPH), a common psychostimulant prescribed for ADHD, blocks the reuptake of DA and norepinephrine (NE) [20] and improves ADHD symptoms and cognitive deficits [21–24].

DA and NE are also believed to modulate the explore/exploit trade-off, which is the decision between choosing a familiar option with the highest expected reward value or choosing an unfamiliar option with an unknown or uncertain reward value [25]. Exploration provides useful information about the environment, but exploitation maximizes rewards, thus a flexible balance between exploration and exploitation is needed for advantageous performance [26–28]. This trade-off is particularly relevant to ADHD, since several neurobiological theories of ADHD predict alterations in explore/exploit decision making [29]. In general, they predict that ADHD would be associated with less reward-driven and more exploratory decisions, usually as a result of faster reaction times and more impulsive/random selections [12, 29–32]. Explore/exploit decisions can be measured with an  $n$ -armed bandit task. Choices on the bandit task are often modeled using reinforcement learning, which summarizes players' strategies in a small number of parameters such as a learning rate. Learning rates represent how quickly expected reward value is updated, and these rates can vary depending on environmental stability. In a rapidly changing environment, reward values should be updated quickly resulting in higher learning rates and more exploration. Such models are thus useful tools for understanding mechanisms that underlie normal and abnormal decision making.

Despite an abundance of theoretical models proposing abnormal reinforcement learning as part of the etiology of ADHD

<sup>1</sup>University of Arkansas for Medical Sciences, Little Rock, AR, USA and <sup>2</sup>Duke University School of Medicine, Durham, NC, USA  
Correspondence: Merideth A. Addicott (maddicott@uams.edu)

Received: 19 May 2020 Revised: 22 September 2020 Accepted: 29 September 2020  
Published online: 11 October 2020

[12, 13, 15, 30–33], relatively little empirical research has been conducted. One study reported that adolescents with ADHD made more exploratory choices, which was not related to differences in learning rates or random selections [34]. Another study reported that adults with ADHD made more novel choices (i.e., preferred previously unseen options) and had lower learning rates than controls, and DAergic medication improved ADHD performance and increased their learning rates [35]. Here, we tested the DA/NE modulation of explore/exploit decisions in relation to ADHD status. Medication-free adults with and without ADHD completed a 6-armed bandit task (6ABT) at baseline and after a single dose of MPH (40 mg) or placebo in counter-balanced order. We hypothesized that ADHD participants would make more exploratory decisions than controls, and that MPH would reduce group differences.

## METHODS

### Participants

Participants were recruited from the Durham, North Carolina ( $n = 12$  ADHD, 9 control) and Little Rock, Arkansas ( $n = 14$  ADHD, 14 control) communities via social media, flyers, and word-of-mouth. Participants completed a phone interview and in-person screening session to determine eligibility. Eligible participants were between the ages of 18–45 years and were not currently taking stimulant medications. To be eligible, ADHD participants had to have T-scores  $\geq 65$  for inattentive and/or hyperactive-impulsive symptoms on the Conners' Adult ADHD Rating Scale (CAARS) [36], and were evaluated to meet criteria for a primary diagnosis of ADHD based on the Conners' Adult ADHD Diagnostic Interview for DSM-IV [37]. Controls had to have CAARS T-scores  $< 55$  for inattentive, hyperactive-impulsive, and total symptoms.

Participants were excluded if they reported serious health problems (e.g., uncontrolled cardiovascular disease) or neurological problems (e.g., seizure disorder or traumatic brain injury), met criteria for a psychiatric disorder other than ADHD (except for symptoms of depression or anxiety co-morbid with ADHD) based on the MINI International Neuropsychiatric Interview [38], reported drug or alcohol dependence in the past 12 months (other than tobacco), reported daily use of medication for ADHD in the past 6 months, had hypertension (i.e., blood pressure  $> 140/90$  mmHg), or had contraindications for MPH (e.g., motor tics). Participants were also excluded if they tested positive for drugs (iCup, Alere Toxicology Services Portsmouth, VA), alcohol (Alco-Sensor III, Intoximeters Inc St. Louis, MO), or pregnancy (QuickVue+, Quidel Corporation San Diego, CA).

Seventy-nine individuals were consented and screened to participate in the study, and 28 participants were ineligible because they did not meet ADHD/control criteria ( $n = 11$ ), had hypertension ( $n = 6$ ), had a positive drug screen ( $n = 4$ ) had another Axis I diagnosis ( $n = 3$ ), or withdrew before the study day ( $n = 4$ ). Of the 51 participants that met eligibility criteria and began the study, 49 participants completed the baseline 6ABT and were included in the data analysis. Participants provided written informed consent and this protocol was approved by Duke University's and University of Arkansas for Medical Sciences' Institutional Review Boards.

### 6-Armed bandit task (6ABT)

This version of the "restless bandit" task was adapted from previous studies [39–45] and has been published previously [46]. On each trial, six bandit options were depicted on a computer screen and participants selected one to play by pressing a corresponding number on the keypad. Following the selection, the number of points awarded was displayed on the screen for 500 ms. The number of points paid off by each option gradually changed from trial to trial, independently of other bandit options. See Fig. 1. The point values were calculated as follows: bandit options began with an initial point-value of 50 on the first trial and subsequent values were drawn from a Gaussian distribution with a

standard deviation ( $\sigma$ ) = 2.8 around a moving mean and rounded to the nearest integer. Point values were randomly adjusted according to a biased random walk,

$$r_{i,t+1} = \lambda(r_{i,t} - \theta) + \theta + \eta, \quad (1)$$

where  $r_{i,t}$  is the reward value of the  $i$ th target on trial  $t$ ,  $\theta$  is the asymptotic mean reward value (equal to 50), and  $\lambda$  is a central tendency parameter that represents the tendency of  $r$  to drift back toward  $\theta$ .  $\eta$  is a Gaussian random variable with mean zero and standard deviation  $\sigma$ . We used parameter values of  $\lambda = 0.015$  and  $\sigma = 2.8$  since this yielded payouts variable enough to encourage exploration and a low likelihood that a single option would remain most profitable for the entire task. The number of points awarded by option  $i$  on trial  $t$  was allowed to range between 0 and 100 (the resulting range was  $-4$  to 105). A single version of the task was administered to all participants, and all participants received the same pattern of point values.

Participants were told the goal of this task was to earn as many points as possible. Each time they played the task, participants could earn up to an additional \$5 based on the ratio of the number of points they earned to the total number of points possible (up to \$15 total). Each task lasted approximately 15 min and consisted of 900 trials. The bandit task was programmed in Matlab (MathWorks, Inc. Natick, MA) using the Psychophysics Toolbox [47].

### Procedure

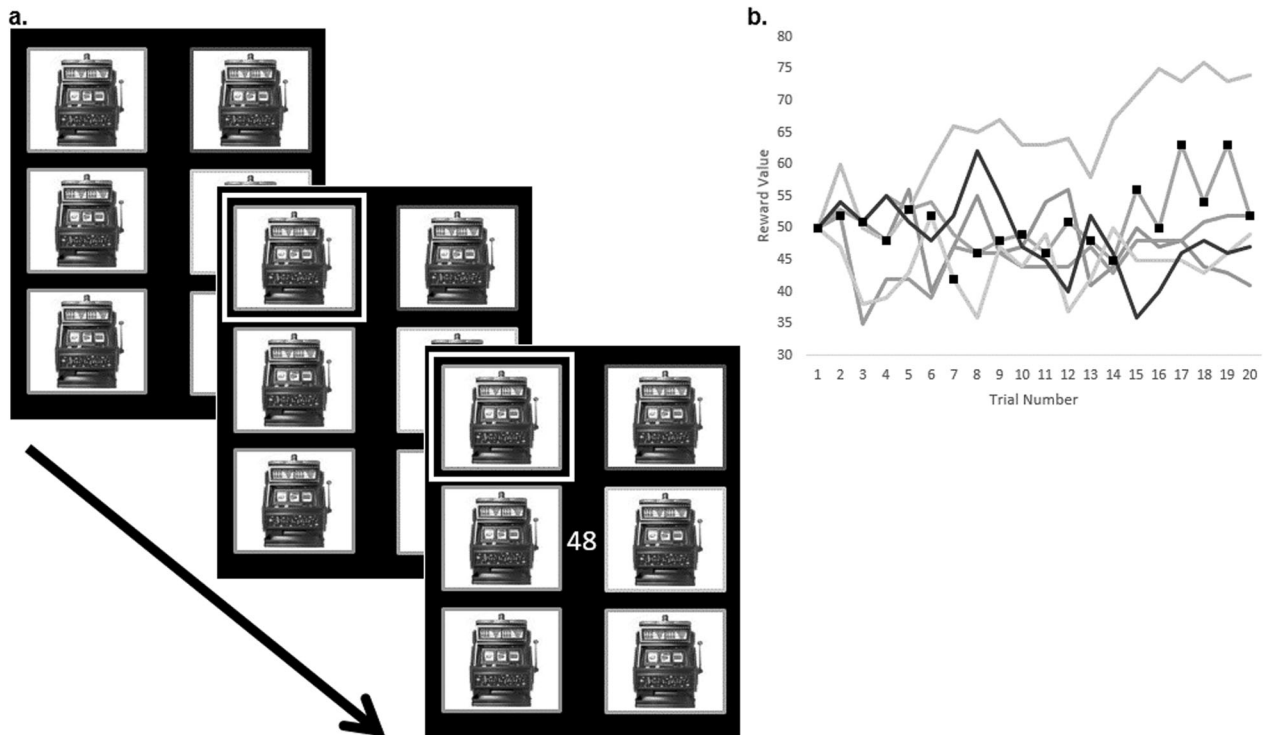
After consenting and eligibility evaluation, participants completed a baseline 6ABT. Then, participants were scheduled for 2 more study visits. These visits occurred within 2 weeks of each other but were at least 48 h apart. For each participant, both study visits occurred either in the afternoon or the morning. Participants were instructed to skip the meal prior to the study visit (i.e., either breakfast or lunch). Participants were administered either immediate-release methylphenidate (MPH: 40 mg) or a matching placebo (PLA) under double-blind conditions and in counter-balanced order. Drugs were ordered and compounded through a pharmacy, and the placebo consisted of lactose. After drug administration, participants were given two cereal bars, a fruit cup, and 8 oz of water and rested for 1 h to allow for drug absorption. The study visit lasted for a total of 3 h, and the 6ABT was completed approximately 2 h after drug administration. At the end of the visit, participants rated to what extent they felt a drug effect on a scale from 1 (not at all) to 10 (extremely). The protocol included other tasks and questionnaires, which have been described previously [48].

### Modeling of the bandit task

Choices made in the bandit task were classified as exploratory or exploitative according to model-based account of participants' individual choices (previously described in [39, 44, 46, 49]). Four reinforcement learning models, which each calculate the estimated bandit option pay-offs differently, were initially fit to the participants' data and compared using the Bayesian Information Criterion (BIC). The BIC is a test of the efficiency of the reinforcing learning model for predicting the data (smaller values represent better fit). The results from the best fitting model are reported here; see Supplementary information for a description of the other three models. On each trial, selection of the option with highest expected value (based on previously seen options) was coded as exploitative, all other choices as exploratory.

As in previous studies, the best fitting model valued the bandit options according to a softmax rule and Kalman filter [39, 50]. The softmax rule describes how individuals select among multiple options, specifically, how individuals choose bandit options probabilistically based on their expected reward values:

$$P(i|\beta, Q_i) = \frac{e^{\beta Q_i}}{\sum_j e^{\beta Q_j}}, \quad (2)$$



**Fig. 1 The 6-Armed bandit task.** **a.** Trial structure of the 6-armed bandit task. On each trial, six bandit options were displayed. A number pad was used to select a single bandit and the selected bandit was outlined in white. Then, the reward value of the selected bandit on that trial was displayed onscreen. **b.** The hidden reward values during the first 20 trials of the 6-armed bandit task. On the y-axis are the values of each of the 6 bandit options per trial. Each bandit option began with an initial value of 50, and values for subsequent trials were randomly adjusted by a biased random walk. Only when a bandit is selected by the player is the bandit’s value revealed. The selections made by a single control participant are shown as black squares.

where  $P(i|\beta, Q_i)$  is the probability of choosing option  $i$ , and  $\beta$  is a so-called softmax decision temperature parameter. A lower value of  $\beta$  typically leads to a higher percentage of explore decisions. The Kalman filter [50] is a Bayes-optimal filtering process used to predict the values of options available for future selection based on the values of options previously chosen. Here, the posterior probability estimates for the option values took the form of normal distributions with mean and variance for all options updated each trial according to a drift rule:

$$\mu_i \leftarrow (1 - \zeta)\mu_i + \zeta\theta, \tag{3}$$

$$\sigma_i^2 \leftarrow (1 - \zeta)^2\sigma_i^2 + D^2, \tag{4}$$

where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of the previous estimate of each option’s value,  $\zeta$  is a central tendency of options to drift toward an asymptotic mean reward value,  $\theta$ , and  $D$  reflects the growing variance in an unchosen option’s value over time due to drift. Due to random changes in the option values over time, uncertainties of unchosen options grow each trial, and mean values decay slowly back toward a subject-specific asymptotic value. Note that participants did not know the true value of the central tendency, so  $\zeta \neq \lambda$  in general. In addition, for the chosen option, we calculated learning parameters as follows:

$$\delta_i = r - \mu_i, \tag{5}$$

$$\alpha_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_0^2}. \tag{6}$$

With  $r$  the outcome on the current trial,  $\mu_i$  the mean of the chosen option, and  $\sigma_0$  the previous standard deviation of the option. As usual,  $\delta$  is the reward prediction error and  $\alpha$

the learning rate, used to update the chosen target value according to

$$\mu_i \leftarrow \mu_i + \alpha_i\delta_i, \tag{7}$$

$$\sigma_i^2 \leftarrow (1 - \alpha_i)\sigma_i^2. \tag{8}$$

As a result, each trial yields a single  $\delta$  and  $\alpha$ , along with vectors  $\mu$  and  $\sigma$ . The learning rate is the rate at which values of the options are updated (i.e., the sensitivity to the most recent reward value of each bandit option). Learning rates are higher in more variable environments and typically positively correlate with exploration.

**Data analysis**

Participant demographic data and CAARS T-scores were analyzed using independent-samples t-tests and Chi-Square tests. Age tended to negatively correlate with the percentage of exploratory decisions (e.g., during the baseline performance  $r = -0.280, p = 0.051$ ). Age was included as a covariate of no interest in all subsequent analyses due to between-group differences.

The 6ABT consisted of 900 trials. The main dependent variable was the percentage of trials coded as “exploratory.” Average reaction time, within-subject reaction time variability (i.e., standard deviation), and reward points (percentage of total points possible) were measured. In addition, we explored two other trial-to-trial variables to investigate qualitative differences in bandit performance between groups, the softmax decision temperature parameter and the learning rate.

6ABT variables were natural-log (LN) transformed to adjust for non-normal distributions and analyzed using univariate analysis of covariance (ANCOVA) (controlling for age) and 2 (drug)  $\times$  2 (group) repeated-measures ANCOVA (controlling for age). Associations between 6ABT exploratory choices and CAARS scores were

**Table 1.** Participant demographics for ADHD and control groups, mean  $\pm$  standard deviation.

	ADHD (n = 26)	Control (n = 23)	p value
Sex (M/F)	14/12	10/13	$p > 0.4$
Age (years)	32.4 $\pm$ 8.3	27.7 $\pm$ 6.6	$p = 0.034$
Years of education	16.1 $\pm$ 3.0	15.8 $\pm$ 2.8	$p > 0.7$
Race (White/Black/other)	20/2/4	12/7/4	$p = 0.051$
CAARS DSM Inattentive T-score	83.2 $\pm$ 8.0	41.8 $\pm$ 5.4	$p < 0.001$
CAARS DSM Hyperactivity T-score	72.2 $\pm$ 9.2	39.6 $\pm$ 5.4	$p < 0.001$
CAARS DSM ADHD T-score	82.7 $\pm$ 7.9	39.6 $\pm$ 5.9	$p < 0.001$

performed using multiple regression, controlling for age. 6ABT data were missing from one ADHD and one control participant during drug administration study days. Drug effect self-report data was missing from one non-ADHD participant. All analyses were performed with SPSS (Chicago, IL) with alpha set to 0.05.

## RESULTS

### Participants

A total of 26 ADHD (14 men) and 23 controls (10 men) were included in the analysis. Participant demographics are shown in Table 1. Groups did not differ in sex ratio or years of education. ADHD participants were older (independent-samples  $t$ -test  $t(47) = 2.2$ ,  $p = 0.034$ ). As expected, ADHD participants had greater CAARS T-scores for inattentive symptoms ( $t(47) = 20.8$ ,  $p < 0.001$ ), hyperactivity symptoms ( $t(47) = 14.8$ ,  $p < 0.001$ ), and DSM ADHD score ( $t(47) = 21.3$ ,  $p < 0.001$ ).

There was a trend towards participants feeling a greater drug effect after MPH (mean  $\pm$  standard deviation: 4.8  $\pm$  2.9) compared to PLA (1.8  $\pm$  1.5) ( $F(1,45) = 3.8$ ,  $p = 0.057$ ). There were no significant group or drug  $\times$  group interaction effects for self-reported drug effect.

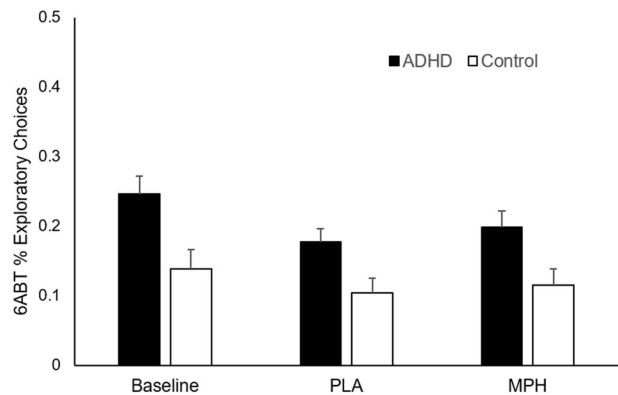
### 6ABT performance

**Baseline performance.** BIC values for the best fitting reinforcement learning model were larger for ADHD (estimated marginal mean  $\pm$  standard error: 1390  $\pm$  119) than for controls (905  $\pm$  127), indicating a better model fit for controls (between-group effect:  $F(1,46) = 7.4$ ,  $p = 0.009$ , partial  $\eta^2 = 0.139$ ) and greater unexplained variance among ADHD.

ADHD participants made more exploratory choices than controls across all task blocks (between-group effect:  $F(1,46) = 4.8$ ,  $p = 0.034$ , partial  $\eta^2 = 0.094$ ). ADHD earned fewer points ( $F(1,46) = 7.8$ ,  $p = 0.008$ , partial  $\eta^2 = 0.145$ ) and had lower learning rates ( $F(1,46) = 7.8$ ,  $p = 0.008$ , partial  $\eta^2 = 0.145$ ) compared to controls. There were no other significant differences in performance measures. See Fig. 2 for illustration, Table 2a for means and standard errors, and Supplementary Fig. S2 for box plots of baseline 6ABT performance data. A summary of the parameters from the best fitting softmax rule and Kalman filter model is shown in Supplementary Table S1.

**Methylphenidate versus placebo.** Across the two study days, ADHD made more exploratory choices than controls across both drug conditions (between-group effect:  $F(1,44) = 4.8$ ,  $p = 0.034$ , partial  $\eta^2 = 0.098$ ). See Fig. 2 and Table 2b. There were no significant drug or interaction effects on exploratory choices.

Reaction times were faster after MPH than PLA (drug effect:  $F(1,44) = 4.8$ ,  $p = 0.034$ , partial  $\eta^2 = 0.098$ ) and showed a drug  $\times$  group interaction ( $F(1,44) = 4.2$ ,  $p = 0.046$ , partial  $\eta^2 = 0.087$ ). Follow-up analyses showed a trend towards controls having slower



**Fig. 2 Overall percent exploratory choices (estimated marginal means) for ADHD and controls across baseline, placebo (PLA) and methylphenidate (MPH) administration.** At baseline and across the two drug administration conditions, ADHD made more exploratory choices than controls ( $p$ 's  $< 0.05$ ). Error bars are standard error of the mean.

reaction times after MPH (drug effect:  $F(1,20) = 3.8$ ,  $p = 0.064$ ), but no significant effect among ADHD. There was no significant between-group effect on reaction time.

Across both study days, ADHD had greater reaction time variability compared to controls (between-group effect:  $F(1,44) = 4.7$ ,  $p = 0.036$ , partial  $\eta^2 = 0.097$ ). ADHD also earned fewer points compared to controls (between-group effect:  $F(1,44) = 8.6$ ,  $p = 0.005$ , partial  $\eta^2 = 0.163$ ). There were no other significant differences in performance measures.

**Associations between ADHD symptoms and 6ABT performance**  
In a multiple regression model including CAARS hyperactive T-scores, inattentive T-scores, and age as predictor variables, baseline 6ABT percent exploratory choices (LN transformed data) positively associated with hyperactive T-scores from both ADHD and control participants ( $\beta = 0.031$ , standard error = 0.013,  $p = 0.019$ ). The association between 6ABT percent exploratory choices and inattentive T-scores was not significant ( $\beta = -0.015$ , standard error = 0.011,  $p > 0.1$ ), see Fig. 3. Within each group, the beta coefficients between exploratory choices and hyperactive T-scores were relatively greater than with inattentive T-scores, although associations were not significant within the smaller samples (ADHD hyperactive T-scores  $\beta = 0.030$ , standard error = 0.016,  $p = 0.074$ , inattentive T-scores  $\beta = -0.015$ , standard error = 0.018,  $p > 0.4$ ; control hyperactive T-scores  $\beta = 0.048$ , standard error = 0.031,  $p > 0.1$ , inattentive T-scores  $\beta = -0.031$ , standard error = 0.029,  $p > 0.2$ ).

## DISCUSSION

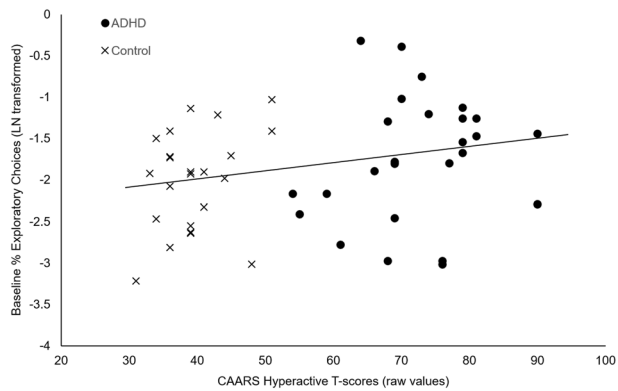
In summary, non-medicated adults with and without ADHD completed the 6ABT, a computerized measure of explore/exploit decision making, at baseline and then after methylphenidate (MPH, 40 mg) and PLA on separate occasions. In support of our first hypothesis, ADHD participants made more exploratory choices and earned fewer points. Across all participants, the number of exploratory choices positively associated with hyperactivity symptoms. These results are consistent with theoretical models of increased exploratory decisions in ADHD [12, 29–33]. Contrary to our second hypothesis, MPH did not affect exploratory choices. ADHD participants continued to make more exploratory choices and earned fewer points than controls in both drug administration sessions. The results of the present study show that individuals with ADHD consistently explore low-value options at the expense of maximizing their rewards. The inability to suppress actions with little to no reward value may be a key feature of

**Table 2.** a. 6-Armed Bandit Task performance data at baseline for ADHD and control participants. ANCOVA analyses were performed on natural-log transformed data controlling for age. Shown in the table are antilog values of the estimated marginal means  $\pm$  standard error (lower to upper 95% confidence intervals). b. 6-Armed Bandit Task performance data across placebo (PLA) and methylphenidate (MPH) administration for ADHD and control participants. ANCOVA analyses were performed on natural-log transformed data controlling for age. Shown in the table are antilog values of the estimated marginal means  $\pm$  standard error (lower to upper 95% confidence intervals).

	ADHD (n = 26)		Control (n = 23)		Between-group p value
	mean $\pm$ s.e.	95% CI	mean $\pm$ s.e.	95% CI	
% Exploratory choices	0.189 $\pm$ 0.029	(0.145–0.246)	0.123 $\pm$ 0.020	(0.093–0.163)	p = 0.034
Reaction time (s)	0.338 $\pm$ 0.041	(0.272–0.419)	0.287 $\pm$ 0.038	(0.228–0.361)	p > 0.3
Reaction time variability (s)	0.548 $\pm$ 0.066	(0.443–0.677)	0.446 $\pm$ 0.058	(0.356–0.559)	p > 0.1
% Reward points	0.890 $\pm$ 0.006	(0.877–0.903)	0.916 $\pm$ 0.007	(0.903–0.932)	p = 0.008
Learning rate	0.174 $\pm$ 0.063	(0.102–0.297)	0.528 $\pm$ 0.208	(0.298–0.936)	p = 0.008
Decision temperature	1.05 $\pm$ 0.43	(0.58–1.90)	0.95 $\pm$ 0.43	(0.50–1.79)	p > 0.8

	ADHD (n = 25)		Control (n = 22)		Between-group p value
	mean $\pm$ s.e.	95% CI	mean $\pm$ s.e.	95% CI	
% Exploratory choices	PLA		PLA		p = 0.034
	0.142 $\pm$ 0.018	(0.111–0.183)	0.089 $\pm$ 0.012	(0.068–0.116)	
Reaction time (s)	MPH		MPH		p > 0.2
	0.246 $\pm$ 0.029	(0.195–0.310)	0.177 $\pm$ 0.023	(0.138–0.227)	
Reaction time variability (s)	PLA		PLA		p = 0.036
	0.306 $\pm$ 0.043	(0.233–0.403)	0.193 $\pm$ 0.029	(0.144–0.258)	
% Reward points	MPH		MPH		p = 0.005
	0.901 $\pm$ 0.005	(0.891–0.912)	0.919 $\pm$ 0.006	(0.907–0.930)	
Learning rate	PLA		PLA		p > 0.7
	0.268 $\pm$ 0.060	(0.176–0.409)	0.274 $\pm$ 0.065	(0.175–0.431)	
Decision temperature	MPH		MPH		p > 0.1
	1.25 $\pm$ 0.38	(0.71–2.19)	0.91 $\pm$ 0.30	(0.49–1.67)	



**Fig. 3** Scatterplot between CAARS hyperactive T-score (raw data values) and baseline 6ABT percent exploratory choices (LN transformed data). Multiple regression analysis indicated a significant association between hyperactive T-scores and exploratory choices, controlling for age and inattentive T-scores ( $\beta = 0.031$ , standard error = 0.013,  $p = 0.019$ ).

hyperactive ADHD symptoms. The lack of an MPH effect is consistent with other studies showing no effects of MPH on some higher-order cognitive processes [51, 52]. For example, MPH reliably improves task performance on measures of eye movement control, attention/vigilance, and inhibitory control, but MPH is less effective on measures of working memory/divided attention, potentially because multiple cognitive processes are engaged and MPH-modulation of DA/NE signaling has less direct influence over these processes [51].

While many theoretical models have proposed increased exploratory decisions in ADHD, potential explanations have varied. These results challenge some straightforward explanations. For example, ADHD participants may make faster, more impulsive decisions and have more lapses in attention, resulting in shorter reaction times and more reaction time variability [53, 54]. Ultimately, this can manifest as less reward-driven/more exploratory decisions [12, 29, 31]. However, we report similar reaction times and reaction time variability between groups at baseline, in spite of differences in exploratory choices. This suggests that cognitive processing times and attentional performance were similar between groups and did not contribute to differences in explore/exploit decisions.

The advantage of reinforcement learning models is that they can provide insight into decision-making sub-functions and reveal impairments not always evident in gross behavioral measures, such as reaction time. The reinforcement learning model used here consists of two complementary components, a Kalman filter that describes how the expected values of bandit options are updated based on the experienced reward history, and a softmax rule that describes how options are selected. The equations for these components provide sub-function values that can inform how differences in explore/exploit decision making may occur.

The learning rate is a sub-function of the Kalman filter and is the rate at which values of the bandit options are updated. Higher learning rates result in fast learning from recent experience, but also fast forgetting. Lower learning rates lead to slow adaptation, but also less influence from random variations in feedback. According to Ziegler et al., several neurobiological models of ADHD predict lower learning rates for rewards [29–31]. At baseline, ADHD participants had significantly lower learning rates, which is surprising since higher learning rates tend to associate with more exploratory decisions. However, if exploration occurs in a way that does not track value, the learning rate will be lower and the model will perform more smoothing of the variability in option selection. These group differences in learning rate disappeared in the subsequent drug administration sessions. Previous computational studies have shown mixed effects of ADHD status on learning rate. Sethi et al.

recently reported that, in a placebo session, adults with ADHD had lower learning rates, earned fewer points, and made more novel selections on a 3-armed bandit task compared to controls [35]. Conversely, Hauser et al. reported no differences in learning rate, although adolescents with ADHD were more exploratory during a probabilistic reversal learning task [34]. Similar learning rates indicate that ADHD participants learned the reward contingencies and provides more evidence that increased exploratory decisions is not simply more random selections [34]. Altogether, this suggests that differences in learning rates may not have caused the differences in explore/exploit decisions.

The decision temperature parameter is a sub-function of the softmax rule that influences whether a choice will be coded as explore or exploit. In the model used here, lower temperature values are expected to correspond with more exploratory decisions. Many neurobiological models predict different temperature values and more exploratory decisions among individuals with ADHD [12, 29–32]. Unexpectedly, we report similar temperature values at baseline, and nonsignificantly lower values during the drug administration sessions among ADHD. Similarly, Sethi et al. reported that ADHD participants on placebo had nonsignificantly lower temperature values [35]. In contrast, Hauser et al. reported that differences in the decision temperature parameter accounted for more exploratory decisions among adolescents with ADHD [34].

It is unclear why the expected relationships between reinforcement learning model sub-functions and explore/exploit decisions did not occur across groups. It may be that decision rules based on typical goal-directed decisions do not fully explain the atypical decisions shown in ADHD. ADHD participants had more unexplained variance in their choices, which potentially decreased the signal-to-noise ratio resulting in a lower learning rate. This may be why the ADHD participants had a significantly larger BIC, indicating the model did not fit their data as well as the control participants. One potential explanation for the unexplained variance in ADHD choices is that ADHD is akin to being in a low gain state [34, 55, 56]. Gain refers to the degree to which the salience of specific information in the environment can be enhanced and acted on immediately or be suppressed and acted on later. In a low gain state, no single bit of information dominates and less important information is monitored and acted on, which can lead to more variable, exploratory decisions. This is a recognized phenomenon in ADHD, and the CAARS queries this as “Sometimes my attention narrows so much that I’m oblivious to everything else; other times it’s so broad that everything distracts me” [36]. Anecdotally, individuals with ADHD are less able to allocate their attention appropriately and may be hyper focused on one activity and struggle to maintain attention on another activity. Not being able to suppress the salience of less valuable, alternate bandit options while exploiting the most valuable option may be a reasonable explanation for increased exploratory decisions. This may also explain the significant association between exploratory choices and hyperactive ADHD symptoms.

We had hypothesized that MPH would reduce group differences in exploratory choices based on extensive preclinical and clinical evidence that DA and NE modulate explore/exploit decisions and reinforcement learning sub-functions (reviewed in [57]). For example, elevated tonic DA levels in DA-transporter knock-down mice are associated with smaller decision temperature values, indicating more exploratory decisions [58], and reduced DA transmission in Parkinson’s disease has been associated with lower learning rates, which are increased by DAergic medications [59]. However, very few human-subject studies have tested the catecholaminergic modulation of explore/exploit decisions. A recent study administered a 4-armed bandit task to healthy men after L-dopa (a DA precursor), haloperidol (a DA antagonist), and placebo [60]. Compared to placebo, L-dopa reduced uncertainty-based exploration (i.e., trials where the option with the highest exploration

bonus was chosen); whereas haloperidol had no effect [60]. Another study administered a 3-armed bandit task with a novelty manipulation to adults with and without ADHD [35]. DAergic medication increased points earned and learning rates in ADHD compared to controls, but there was no group or drug effect on decision temperature values [35]. The NE system has also been implicated in explore/exploit decision making [49], however, administration of a NE reuptake inhibitor did not increase exploratory decisions as hypothesized [45]. More research is needed to understand these inconsistencies in the context of phasic versus tonic catecholamine signaling, and striatal versus prefrontal regulation of explore/exploit decisions.

The strengths of this study include a placebo-controlled, counterbalanced design. Several limitations include the group differences in age and the lack of a validated measure of real-world exploratory decisions or personality traits. We used a dose of MPH that our team has administered previously to adults with and without ADHD, and which has shown to produce behavioral effects [61–63]. However, 40 mg MPH is a relatively large dose among treatment-naïve individuals. MPH has different effects on behavior at small, medium and large doses and future studies should compare the effects across doses. Lastly, given the racial disparities in ADHD diagnosis and treatment, future studies should pay special attention to the recruitment of minorities.

In support of several theoretical models of ADHD [12, 29–32], these results indicate that adults with ADHD make more exploratory decisions at the expense of maximizing rewards. Future studies could investigate whether this increased exploratory decision making is related to striatal or prefrontal function using neuroimaging. These results have clinical implications. Reinforcement learning models can help elucidate higher-order cognitive impairments and provide a more nuanced explanation of symptoms. In particular, the processes that underlie exploratory decisions on the 6ABT may be driving hyperactive symptoms, and a better understanding of such processes could help guide therapy. For instance, clinicians may want to be especially attuned to the decision-making capabilities of their patients with greater levels of hyperactivity. In addition, new therapeutic methods that emphasize top-down regulation of attention and conflict detection could be useful in reducing this particular impairment [64].

## FUNDING AND DISCLOSURE

This work was supported by the Brain and Behavior Research Foundation (Grant #23703, PI: Addicott). MDW reports nonfinancial support from Purdue Pharma CA, personal fees and nonfinancial support from Takeda, personal fees and nonfinancial support from Global Medical Education, personal fees from Huron Consulting, personal fees and nonfinancial support from Rhodes, personal fees and nonfinancial support from MHS, personal fees from Mundipharma, personal fees from Johns Hopkins University Press, nonfinancial support from Eunethydis, nonfinancial support from World Federation of ADHD, nonfinancial support from Israeli Foundation for ADHD, nonfinancial support from Canadian Attention Deficit Resource Alliance, nonfinancial support from American Professional Association for ADHD, nonfinancial support from Purdue Pharma US, personal fees and nonfinancial support from Cingulate, nonfinancial support from Children and Adults with ADD, personal fees from Boston Children's Hospital, and personal fees from Tris. Other authors declare that they have no conflict of interest pertaining to this manuscript.

## AUTHOR CONTRIBUTIONS

MAA, principal investigator, designed and conducted the study, organized and analyzed the data, and wrote the majority of the manuscript. JMP provided support for the reinforcement learning algorithms and interpretation of the results. JCS and JJS performed clinical evaluations of the ADHD participants. SHK and MDW guided

the development of the protocol and provided study oversight. All authors approved the final manuscript before submission.

## ADDITIONAL INFORMATION

**Supplementary Information** accompanies this paper at (<https://doi.org/10.1038/s41386-020-00881-8>).

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

## REFERENCES

1. Wehmeier PM, Schacht A, Barkley RA. Social and emotional impairment in children and adolescents with ADHD and the impact on quality of life. *J Adolesc Health*. 2010;46:209–17.
2. Sobanski E, Bruggemann D, Alm B, Kern S, Philipson A, Schmalzried H, et al. Subtype differences in adults with attention-deficit/hyperactivity disorder (ADHD) with regard to ADHD-symptoms, psychiatric comorbidity and psychosocial adjustment. *Eur Psychiatry*. 2008;23:142–9.
3. APA. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
4. Biederman J, Mick E, Faraone SV. Age-dependent decline of symptoms of attention deficit hyperactivity disorder: Impact of remission definition and symptom type. *Am J Psychiatry*. 2000;157:816–8.
5. Moffitt TE, Houts R, Asherson P, Belsky DW, Corcoran DL, Hammerle M, et al. Is adult ADHD a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *Am J Psychiatry*. 2015;172:967–77.
6. Simon V, Czobor P, Balint S, Meszaros A, Bitter I. Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. *Br J Psychiatry*. 2009;194:204–11.
7. Wilens TE, Biederman J, Faraone SV, Martelon M, Westerberg D, Spencer TJ. Presenting ADHD symptoms, subtypes, and comorbid disorders in clinically referred adults with ADHD. *J Clin Psychiatry*. 2009;70:1557–62.
8. Willcutt EG, Pennington BF, Olson RK, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Dev Neuropsychol*. 2005;27:35–78.
9. Boonstra AM, Oosterlaan J, Sergeant JA, Buitelaar JK. Executive functioning in adult ADHD: a meta-analytic review. *Psychol Med*. 2005;35:1097–108.
10. Hervey AS, Epstein JN, Curry JF. Neuropsychology of adults with attention-deficit/hyperactivity disorder: a meta-analytic review. *Neuropsychology*. 2004;18:485–503.
11. Sonuga-Barke EJS. The dual pathway model of AD/HD: an elaboration of neurodevelopmental characteristics. *Neurosci Biobehav Rev*. 2003;27:593–604.
12. Tripp G, Wickens JR. Neurobiology of ADHD. *Neuropharmacology*. 2009;57:579–89.
13. Luman M, Tripp G, Scheres A. Identifying the neurobiology of altered reinforcement sensitivity in ADHD: a review and research agenda. *Neurosci Biobehav Rev*. 2010;34:744–54.
14. Jackson JN, MacKillop J. Attention-deficit/hyperactivity disorder and monetary delay discounting: a meta-analysis of case-control studies. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2016;1:316–25.
15. Luman M, Oosterlaan J, Sergeant JA. The impact of reinforcement contingencies on AD/HD: a review and theoretical appraisal. *Clin Psychol Rev*. 2005;25:183–213.
16. Aarts E, van Holstein M, Cools R. Striatal dopamine and the interface between motivation and cognition. *Front Psychol*. 2011;2:163.
17. Fusar-Poli P, Rubia K, Rossi G, Sartori G, Balottin U. Striatal dopamine transporter alterations in ADHD: pathophysiology or adaptation to psychostimulants? A meta-analysis. *Am J Psychiatry*. 2012;169:264–72.
18. Volkow ND, Wang GJ, Kollins SH, Wigal TL, Newcorn JH, Telang F, et al. Evaluating dopamine reward pathway in ADHD clinical implications. *J Am Med Assoc*. 2009;302:1084–91.
19. Volkow ND, Wang GJ, Newcorn JH, Kollins SH, Wigal TL, Telang F, et al. Motivation deficit in ADHD is associated with dysfunction of the dopamine reward pathway. *Mol Psychiatry*. 2011;16:1147–54.
20. Challman TD, Lipsky JJ. Methylphenidate: Its pharmacology and uses. *Mayo Clin Proc*. 2000;75:711–21.
21. Faraone SV, Spencer T, Aleardi M, Pagano C, Biederman J. Meta-analysis of the efficacy of methylphenidate for treating adult attention-deficit/hyperactivity disorder. *J Clin Psychopharmacol*. 2004;24:24–29.
22. Castells X, Blanco-Silvente L, Cunill R. Amphetamines for attention deficit hyperactivity disorder (ADHD) in adults. *Cochrane Datab Syst Rev*. 2018:CD007813.
23. Coghill DR, Seth S, Pedroso S, Usala T, Currie J, Gagliano A. Effects of methylphenidate on cognitive functions in children and adolescents with attention-deficit/hyperactivity disorder: evidence from a systematic review and a meta-analysis. *Biol Psychiatry*. 2014;76:603–15.

24. Pievsky MA, McGrath RE. Neurocognitive effects of methylphenidate in adults with attention-deficit/hyperactivity disorder: a meta-analysis. *Neurosci Biobehav Rev.* 2018;90:447–55.
25. Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc B.* 2007;362:933–42.
26. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, Mass.: MIT Press. xviii; 1998. 322 p.
27. Graybiel AM. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci.* 2008;31:359–87.
28. Daw ND, Doya K. The computational neurobiology of learning and reward. *Curr Opin Neurobiol.* 2006;16:199–204.
29. Ziegler S, Pedersen ML, Mowinckel AM, Biele G. Modelling ADHD: a review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neurosci Biobehav Rev.* 2016;71:633–56.
30. Frank MJ, Santamaria A, O'Reilly RC, Willcutt E. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology.* 2007;32:1583–99.
31. Sagvolden T, Johansen EB, Aase H, Russell VA. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behav Brain Sci.* 2005;28:397–+.
32. Seeman P, Madras B. Methylphenidate elevates resting dopamine which lowers the impulse-triggered release of dopamine: a hypothesis. *Behav Brain Res.* 2002;130:79–83.
33. Hauser TU, Fiore VG, Moutoussis M, Dolan RJ. Computational psychiatry of ADHD: neural gain impairments across marrian levels of analysis. *Trends Neurosci.* 2016;39:63–73.
34. Hauser TU, Iannaccone R, Ball J, Mathys C, Brandeis D, Walitza S, et al. Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA Psychiatry.* 2014;71:1165–73.
35. Sethi A, Voon V, Critchley HD, Cercignani M, Harrison NA. A neurocomputational account of reward and novelty processing and effects of psychostimulants in attention deficit hyperactivity disorder. *Brain.* 2018;141:1545–57.
36. Conners CK, Erhardt D, Sparrow E. The Conners adult ADHD rating scale (CAARS). Multi-Health Systems, Inc: Toronto; 1998.
37. Epstein J, Johnson DE, Conners CK. Conners' adult ADHD diagnostic interview for DSM-IV (CAADID). MHS: New York; 2001.
38. Sheehan D, Janavas J, Harnett-Sheehan K, Sheehan M, Gray C. MINI International Neuropsychiatric Interview (English Version 6.0.0). Tampa, FL: University of South Florida College of Medicine; 2009.
39. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature.* 2006;441:876–9.
40. Gittins JC, Jones DH. A dynamic allocation index for the sequential design of experiments, In: Gani JM, editor. *Progress in statistics.* Amsterdam: North Holland; 1974.
41. Gittins JC. Multi-armed bandit allocation indices. Wiley-Interscience series in systems and optimization. Chichester; New York: Wiley; 1989. p. 252.
42. Berry DA, Fristedt B. Bandit problems: sequential allocation of experiments. *Monographs on statistics and applied probability.* New York: Chapman and Hall, London; 1985.
43. Whittle P. Restless bandits: activity allocation in a changing world celebration of applied probability. *J Appl Probab.* 1988;25A:287–98.
44. Pearson JM, Hayden BY, Raghavachari S, Platt ML. Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Curr Biol.* 2009;19:1532–7.
45. Jepma M, te Beek ET, Wagenmakers E-J, van Gerven JMA, Nieuwenhuis S. The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Front Hum Neurosci.* 2010;4:1–13.
46. Addicott MA, Pearson JM, Wilson J, Platt ML, McClernon FJ. Smoking and the bandit: a preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Exp Clin Psychopharmacol.* 2013;21:66–73.
47. Brainard DH. The psychophysics toolbox. *Spat Vis.* 1997;10:433–6.
48. Addicott MA, Schechter JC, Sapyta JJ, Selig JP, Kollings SH, Weiss MD. Methylphenidate increases willingness to perform effort in adults with ADHD. *Pharmacol Biochem Behav.* 2019;183:14–21.
49. Jepma M, Nieuwenhuis S. Pupil diameter predicts changes in the exploration-exploitation tradeoff: evidence for the adaptive gain theory. *J Cogn Neurosci.* 2011;23:1587–96.
50. Anderson BDO, Moore JB. *Optimal filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.
51. Pietrzak RH, Mollica CM, Maruff P, Snyder PJ. Cognitive effects of immediate-release methylphenidate in children with attention-deficit/hyperactivity disorder. *Neurosci Biobehav Rev.* 2006;30:1225–45.
52. Linssen AMW, Sambeth A, Vuurman EFP, Riedel WJ. Cognitive effects of methylphenidate in healthy volunteers: a review of single dose studies. *Int J Neuropsychopharmacol.* 2014;17:961–77.
53. Kofler MJ, Rapport MD, Sarver DE, Raiker JS, Orban SA, Friedman LM, et al. Reaction time variability in ADHD: a meta-analytic review of 319 studies. *Clin Psychol Rev.* 2013;33:795–811.
54. Tamm L, Narad ME, Antonini TN, O'Brien KM, Hawk LW, Epstein JN. Reaction time variability in ADHD: a review. *Neurotherapeutics.* 2012;9:500–8.
55. Losier BJ, McGrath PJ, Klein RM. Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: a meta-analytic review. *J Child Psychol Psychiatry Allied Discip.* 1996;37:971–87.
56. Aston-Jones G, Cohen JD. Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J Comp Neurol.* 2005;493:99–110.
57. Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML. A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology.* 2017;42:1931–9.
58. Beeler JA, Daw N, Frazier CRM, Zhuang XX. Tonic dopamine modulates exploitation of reward learning. *Front Behav Neurosci.* 2010;4:170.
59. Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *J Neurosci.* 2009;29:15104–14.
60. Chakroun K, Mathar D, Wiehler A, Ganzer F, Peters J. Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *bioRxiv.* 2019:706176.
61. Kollins SH, English J, Robinson R, Hallyburton M, Chrisman AK. Reinforcing and subjective effects of methylphenidate in adults with and without attention deficit hyperactivity disorder (ADHD). *Psychopharmacology.* 2009;204:73–83.
62. Kollins SH, Schoenfelder E, English JS, McClernon FJ, Dew RE, Lane SD. Methylphenidate does not influence smoking-reinforced responding or attentional performance in adult smokers with and without attention deficit hyperactivity disorder (ADHD). *Exp Clin Psychopharmacol.* 2013;21:375–84.
63. Sweitzer MM, Kollins SH, Kozink RV, Hallyburton M, English J, Addicott MA, et al. ADHD, smoking withdrawal, and inhibitory control: results of a neuroimaging study with methylphenidate challenge. *Neuropsychopharmacology.* 2018;43:851.
64. Mitchell JT, Zylowska L, Kollins SH. Mindfulness meditation training for attention-deficit/hyperactivity disorder in adulthood: current empirical support, treatment overview, and future directions. *Cogn Behav Pract.* 2015;22:172–91.