ORIGINAL ARTICLE   OPEN

Impaired arbitration between reward-related decision-making strategies in Alcohol Users compared to Alcohol Non-Users: a computational modeling study

Srinivasan A. Ramakrishnan[1,2,9], Riaz B. Shaik[2,9], Tamizharasan Kanagamani[3], Gopi Neppala[2], Jeffrey Chen [4], Vincenzo G. Fiore[2], Christopher J. Hammond[5], Shankar Srinivasan[1], Iliyan Ivanov[2], V. Srinivasa Chakravarthy[3], Wouter Kool[6] and Muhammad A. Parvaz[2,7,8 ✉]

Reinforcement learning studies propose that decision-making is guided by a tradeoff between computationally cheaper model-free (habitual) control and costly model-based (goal-directed) control. Greater model-based control is typically used under highly rewarding conditions to minimize risk and maximize gain. Although prior studies have shown impairments in sensitivity to reward value in individuals with frequent alcohol use, it is unclear how these individuals arbitrate between model-free and model-based control based on the magnitude of reward incentives. In this study, 81 individuals (47 frequent Alcohol Users and 34 Alcohol Non-Users) performed a modified 2-step learning task where stakes were sometimes high, and other times they were low. Maximum *a posteriori* fitting of a dual-system reinforcement-learning model was used to assess the degree of model-based control, and a utility model was used to assess risk sensitivity for the low- and high-stakes trials separately. As expected, Alcohol Non-Users showed significantly higher model-based control in higher compared to lower reward conditions, whereas no such difference between the two conditions was observed for the Alcohol Users. Additionally, both groups were significantly less risk-averse in higher compared to lower reward conditions. However, Alcohol Users were significantly less risk-averse compared to Alcohol Non-Users in the higher reward condition. Lastly, greater model-based control was associated with a less risk-sensitive approach in Alcohol Users. Taken together, these results suggest that frequent Alcohol Users may have impaired metacontrol, making them less flexible to varying monetary rewards and more prone to risky decision-making, especially when the stakes are high.

LAY SUMMARY

This study explored decision-making in frequent alcohol users versus alcohol non-users. Normally, people use more thoughtful strategies when rewards are high. The study shows alcohol users are not doing so and remain less cautious and take more risks when stakes are high. This suggests that frequent alcohol users may struggle to adapt their decision-making based on reward size, leading to riskier choices especially when potential gains are significant.

INTRODUCTION

Excessive alcohol consumption and alcohol use disorders (AUD) are growing public health concerns, costing $249 billion annually in the United States alone [1, 2]. During the COVID-19 pandemic, deaths from excessive alcohol and AUD increased, especially among women [3–5]. Therefore, it is vital to increase our understanding of neurobehavioral factors linked to alcohol consumption and its consequences. Impaired decision-making is one such behavioral aspect that is found to be associated with alcohol and substance use disorders. Research in this domain has painted the relatively straightforward picture that substance abuse reduces reward sensitivity [6–11], and thereby reward-related learning [12, 13].

However, decision-making is not a unitary construct. Humans are equipped with a range of choice strategies, varying in accuracy and demand [14, 15]. Therefore, efficient decision-making requires us to allocate cognitive resources between these strategies [16, 17]. Even though such arbitration is critical for efficiently navigating the world [18, 19], it is unknown whether and how alcohol and substance abuse also impair this form of higher-order decision-making.

In this paper, we address this question through the lens of reinforcement learning (RL). Recent advances in RL have provided

[1]Department of Health Informatics, Rutgers - School of Health Professions, Piscataway, NJ, USA. [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Biotechnology, Indian Institute of Technology, Madras, Chennai, India. [4]University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [5]Department of Psychiatry & Behavioral Sciences, Division of Child & Adolescent Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [6]Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA. [7]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]These authors contributed equally: Srinivasan A. Ramakrishnan, Riaz B. Shaik. ✉email: Muhammad.parvaz@mssm.edu

formalizations for two systems that control choice: a fast, automatic system and a slow deliberative system. In the language of machine learning, these systems are mapped to "model-free" and "model-based" RL [20–22]. Model-free strategies learn through trial-and-error, which is computationally cheap but inflexible. Model-based strategies are computationally demanding but more accurate because they plan through an internal model of the environment towards goals [23–26].

The arbitration between these systems follows a cost-benefit tradeoff. A greater influence of model-based decision-making is observed for high reward stakes to increase reward rate, but only when it is more accurate than model-free control [27, 28]. People rely more on model-free control when cognitive resources are taxed [29] or when planning complexity increases [30]. Moreover, the ability to arbitrate between these systems varies between populations: it emerges throughout adolescence [31] and declines during aging [32]. In non-human studies, it is observed that substance use alters the balance between model-free and model-based control in reward-related learning [33, 34]. Moreover, some studies in humans have shown that substance use results in dysfunctional and in some cases risk insensitive decision-making [35–37].

Here, we use this framework to study impairments in decision-making in individuals with long-term and frequent alcohol use [38, 39]. These impairments may be driven by an increased reliance on less taxing model-free control [7] or by reduced risk sensitivity (i.e., the trade-off between the value preference and the risk preference) [40]. However, others have reported that alcohol [41] and substance use [42] may be associated with reduced model-free decision-making, whereas some have reported comparable use of model-free and model-based control between Alcohol Users and Alcohol Non-Users [43]. However, it remains unknown whether alcohol use affects the arbitration between systems. That is, whether it is associated with a reduced ability to shift between strategies based contextual factors such as reward magnitude.

To examine these questions, we used a modified version of the "two-step task" [20, 24], a sequential decision-making paradigm that dissociates between model-based and model-free contributions to a choice. This task included a reward magnitude manipulation to test participants' ability to increase model-based control when there is a heightened opportunity for reward [23].

We hypothesized that, compared to Alcohol Non-Users, (1) Alcohol Users would show less model-based decision-making in overall task performance, and (2) Alcohol Users would be less sensitive to reward amplification, showing less model-based decision-making in the high-stakes condition compared to the low-stakes condition. In addition, we also explored whether the hypothesized difference in arbitration is associated with different levels of risk sensitivity in Alcohol Users.

METHODS
Participants
Eighty-one individuals (58% female) participated in an online study that was conducted during the COVID-19 epidemic (July 2020 – August 2022). The study was advertised on social media platforms and participants throughout the continental United States were eligible to participate in this study. All participants provided informed consent. The study was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

All participants completed the 'Coronavirus Health Impact Survey' (CRISIS) questionnaire [44]. Data on alcohol use frequency from the CRISIS survey item 146 ("For the 3 months DURING LOCKDOWN, how frequently did you use alcohol ?") was used to stratify participants into different groups. Participants who reported no alcohol use and those who reported consuming alcohol less than once a month were grouped as Alcohol Non-Users ($n = 34$), and those who reported consuming alcohol multiple times a month to more than once a day were grouped as Alcohol Users ($n = 47$). The Alcohol Use Disorders Identification Test (AUDIT) [45, 46], Patient Health Questionnaire PHQ-9) [47] and Generalized Anxiety Disorder

Questionnaire (GAD-7) [48] were also used to assess the severity of alcohol use, depressive and anxiety symptoms, respectively.

Modified Two-Step Task
The task used in this study is a modified version of a previously developed two-step learning task [24, 49], aimed to test whether choice behavior shows increased model-based control in the face of increased reward magnitude [27]. A detailed description of the task has been published previously [27].

Each trial starts randomly in one of two first-stage states, each with a choice between two spaceships that lead to a red or purple planet in the second stage (Fig. 1). These planets provide the opportunity to earn rewards. These independently drift over time for each planet. The task dissociates model-free from model-based decision making because each first-stage state offers an identical choice between transitioning to the red and the purple planet. A model-based decision-maker uses this equivalence to transfer experiences between states, while a model-free decision-maker relies on action-reward contingencies without considering the task's transition structure [23, 27, 49].

Our version of this task included a reward magnitude manipulation [23]. Each trial started with a randomly picked cue that indicated whether it involved low or a high stakes. In high-stakes trials, the rewards earned in the second-stage state were multiplied by 5, whereas in low-stakes trials they were unaltered (Fig. 1). For each trial, there was a 50% probability it would involve low stakes, and 50% probability it would involve high stakes. This ensured there was no bias in the number of high- and low-stakes trials between the two groups of participants.

Each subject completed 200 trials. The task was programmed in jsPsych [50] and was hosted on Pavlovia platform (https://pavlovia.org). In addition to receiving compensation for participating in the study, participants were told that they could earn up to $5 based on their performance. However, in the end, all participants were paid $5 for completing this task.

Computational modeling
Dual-system RL model. We used a dual-system RL model [24, 51, 52] to describe behavior in terms of both model-based and model-free control. The model-free system learns state-action values for all first and second-stage states through a simple temporal difference-learning algorithm [21], whereas the model-based system uses the transition structure of the task to plan towards goals. The relative tradeoff between these systems is modeled as a mixture between action values computed by these two systems.

The model consists of a function Q(s,a) that maps each state-action pair to estimates of future rewards. The task consists of four available states across two stages, two available actions at the first-stage states (a_A and a_B) and one action at the second-stage states (a_C).

The model-free system uses the SARSA(λ) temporal difference learning algorithm [53] to calculate values of each state-action pair. This algorithm updates the value of the state-action pair (s,a) at each stage 'i' and trial 't', according to:

$$Q_{MF}(s_{i,t}a_{i,t}) = Q_{MF}(s_{i,t}a_{i,t}) + a\delta_{i,t}\varepsilon_{i,t}(s, a)$$

where,

$$\delta_{i,t} = r_{i,t} + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t})$$

is the reward prediction error (RPE), α is the learning rate (determining the degree to which new information is incorporated), and $\varepsilon_{(i,t)}$ (s, a) is the eligibility trace. The eligibility trace of each action is set to zero at the start of each trial and updated according to

$$\varepsilon_{i,t}(s_{i,t}, a_{i,t}) = \varepsilon_{i-1,t}(s_{i,t}, a_{i,t}) + 1$$

before the Q-value update. After each update, the eligibilities of all state-action pairs are then decayed by λ. It is important to note that λ is a free parameter, but that ε is not. The former dictates the degree to which eligibilities are decayed, the latter simply counts which actions have been chosen.

Since there is no reward in the first state, the RPE (δ) for the first stage is only driven by the value of the second-stage action Q_{MF} ($s_{2,t}$, $a_{2,t}$):

$$\delta_{1,t} = Q_{MF}(s_{2,t}, a_{2,t}) - Q_{MF}(s_{1,t}, a_{1,t})$$

Only the first-stage action is updated by this prediction error, and its eligibility for future updates on the current trial is only decayed by λ after this update.
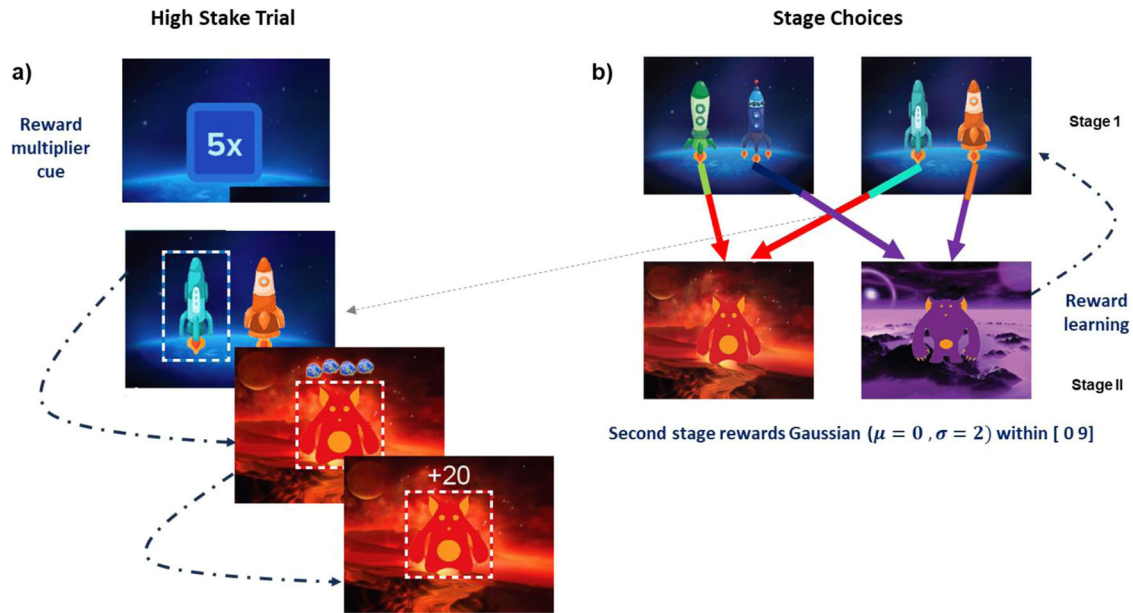
Fig. 1 Modified 2-step learning task. a Cue indicating the nature of the trial by highlighting the stake multiplier 5x, indicating a highstakes. trial **b** Modified 2-step task: Participants choose the first-stage option between two spaceships, followed by a probabilistic transition to the second stage on either the red or purple planet [56].

Since there is no further stage beyond the second, the second-stage prediction error is only driven by the second-stage reward $r_{2,t}$:

$$\delta_{2,t} = r_{2,t} - Q_{MF}(s_{2,t}, a_{2,t})$$

Both the first- and second-stage values are updated at the second stage using this RPE. As mentioned above, the first-stage value is updated with the second-stage RPE down-weighted by the eligibility trace decay, λ. This means that, when $\lambda = 0$, only the value of the current second-stage action value is updated. When $\lambda = 1$, the values of the chosen first-stage and second-stage action values are updated by the same amount.

The model-based agent computes first-stage action values by combining a transition function, which maps the first-stage state-action pairs to a probability distribution over the subsequent states, with the second-stage (model-free) values. For each action a_j in first-stage state $s_{1,I}$, these model-based values are defined in terms of the expected values of each first-stage action using the transition structure P:

$$Q_{MB}(s_{1,i}, a_j) = P(s_{2,1}|s_{1,i}, a_j)Q_{MF}(s_{2,1}, a_c) + P(s_{2,2}|s_{1,i}, a_j)Q_{MF}(s_{2,2}, a_c)$$

At the second stage, the model-based and model-free coincide, and so $Q_{MB} = Q_{MF}$.

To connect the values to choices, the Q-values of both systems are mixed according to a weighting parameter ω:

$$Q_{net}(s_{1,i}, a_j) = \omega Q_{MB}(s_{1,i}, a_j) + (1 - \omega)Q_{MF}(s_{1,i}, a_j)$$

Thus, higher values of this parameter (closer to 1) reflect increased model-based control, whereas lower values (closer to 0) suggest stronger model-free control. To accommodate our stake manipulation, we defined two different weights for the different trial types. Specifically, we set $\omega = \omega_{low}$ for low-stake trials and $\omega = \omega_{high}$ for high-stake trials.

We used the soft-max rule to translate these Q-values to actions. This rule computes the probability for an action reflecting the mixture of action values weighted by an inverse temperature parameter. At both states, the probability of choosing action a on trial t is computed as:

$$p(a_{i,t} = a|s_{i,t}) = \frac{\exp[\beta(Q_{net}(s_{i,t}, a) + \pi \cdot rep(a) + \rho \cdot resp(a)]}{\sum_{a'} \exp[\beta(Q_{net}(s_{i,t}, a'') + \pi \cdot rep(a') + \rho \cdot resp(a')]}$$

where the inverse temperature β determines the randomness of the choice (with values close to zero reflecting fully random choice, and large positive values reflecting exploitation). The indicator variables *rep(a)* and *resp(a)* are set to 1 for actions or key presses that the participants chose on the previous trial and are zero otherwise. Multiplied with the 'stickiness' parameter π and 'response stickiness' parameter ρ, respectively, these capture the degree to which people show choice perseveration or switching at the first stage state.

We used a maximum *a posteriori* estimation with empirical priors based on prior work [32], implemented using the mfit toolbox [54] (https://github.com/sjgershm/mfit) to fit the free parameters in the computational models to observed data. For all parameters bounded between 0 and 1 ($a, \omega_{low}, \omega_{high}, \lambda, \rho$) we used a Beta (2,2) prior. For the inverse temperature β, we used a Gamma (3, 0.2) prior and for the stickiness parameters (π, ρ), we used a $\mathcal{N}(0,1)$ prior. To avoid local optima in estimation, the optimization was run 100 times for each participant with random initializations for each parameter. The final estimations for all parameters were extracted from the run with the maximal posterior probability. (Supplementary table S1) reports the estimated parameters.

We also fit an "exhaustive" model that varied all parameters between the high- and low-stake trials. We used the same maximum a posteriori estimation with the same empirical priors as described above to obtain estimates for these 12 parameters.

Utility function. The utility function U [55] is a modified version of the utility formulation [56, 57] that estimates the risk sensitivity with varying reward magnitude and the trade-off between value preference and risk preference for a given state 's' and action 'a'. It is computed as:

$$U_t(s, a) = R_t(s, a) - \mu.sign(R_t(s, a)) \cdot \sqrt{h_t}(s, a)$$

Here, the risk sensitivity parameter μ reflects risk preference, the term $sign(R_t(s,a))$ reduces the magnitude of estimation function for positive values of R and increases it for negative values of R, and the return variance or risk function ($\sqrt{h_t}$). Thus, this function naturally incorporates the notion of increased risk-seeking behavior for gains and increased risk-aversive behavior for losses.

This function estimates five parameters μ (risk sensitivity), β (explore-exploit tradeoff), η (learning rate), Γ (discount factor for long-term rewards), and δ_{limit} (maximum error magnitude associated with reward). Higher μ implies risk aversion and lower value implies risk seeking, higher β favors exploitation, η is the learning rate, Γ weighs long-term over immediate rewards, and δ_{limit} is the maximum error magnitude.

For risk-value estimation, the model only considers first stage states since the second stage action does not impact the risk trade-off. We fit parameters separately for low- and high-stake trials, using the genetic algorithm [58] to identify the best fit considering response changes (mutations), learning adjustments (crossovers) and choices. The action selection is based on soft-max probabilities, where the expected return is updated using a learning rate (η) and a temporal difference error (δ). The

genetic algorithm evolves mutations and the crossovers over the initial population (size = 1000, generations = 100) to optimally fit parameter values. The free parameters are estimated by maximizing the fitness function iterated over 200 trials.

Of the five free parameters, the risk sensitivity parameter (μ) was of most interest for statistical analysis. Therefore, the **Supplementary Section** provides additional details of the utility function. Here, we have used two different models (1) to estimate the degree of model-based vs model-free control, and (2) to estimate risk sensitivity. Given the relatively small sample size of Alcohol Users ($n = 47$) and Alcohol Non-users ($n = 34$), to ensure the reliability of results, we did not split the data further into test and training sets to cross-validate the model accuracy.

Statistical approach. Parameter fits were analyzed using a 2×2 analysis of variance (ANOVA) using stakes (High vs. Low) and group (Alcohol users vs. Alcohol Non-Users) as between-subject factors. Since the parameter fits of both models were not normally distributed, we used the Mann-Whitney U test for hypothesis testing. Furthermore, we used Spearman Rank correlation to assess the associations between model-based and model-free weighting (ω) and risk sensitivity (μ) parameters (calculated for high-stakes trials, low-stakes trials, and the difference between high- and low-stakes trials) with each other and with validated measures of alcohol use severity (i.e., AUDIT scores) and depression symptom severity (i.e., PHQ-9 scores). Finally, we performed a correlation analysis to investigate the relationship between the weighting parameter (ω) and risk sensitivity (μ) obtained from two models. The analyses were conducted first for the entire sample and then separately for Alcohol Users and Alcohol Non-Users. Bonferroni corrections were applied to adjust for multiple comparisons. To assess the robustness of these correlations, as both parameters are derived from the same dataset, we employed a cross-validation approach by splitting each subject's dataset into two halves. We conducted two types of correlation analyses: one by splitting the data into the first 100 trials and the second 100 trials for preserving temporality, and the other by separating all odd-numbered trials from even-numbered trials, correlating ω and μ derived from these distinct datasets. Results from these cross-validations, along with the sensitivity analysis are presented in the Supplemental Section.

RESULTS
Sample characteristics
As shown in Table 1, groups did not differ significantly in gender ($p = 0.282$), ethnicity ($p = .529$), and education ($p = 0.943$). However, they significantly differed in age ($p < 0.001$), such that a greater proportion of participants (Alcohol Non-Users: 73%, Alcohol Users: 61%) were between the ages 18 and 40. As expected, AUDIT scores were significantly higher in Alcohol Users, compared to Alcohol Non-Users ($p < 0.001$), showing more severe alcohol use in Alcohol Users (Mean and standard deviation are reported in Table 1). Age did not correlate significantly with computational variables, and therefore, was not used as a covariate. We report the results with age as a covariate in the **Supplementary Section**.

Dual system RL model outcomes
Here, we report the statistical tests on variables that relate most closely to our hypotheses, namely the average reward in the task, and the degree of model-based control (ω). The analyses on the other RL parameters are presented in the **Supplementary Section**.

A between-group Mann-Whitney U test showed no significant group differences in mean reward rate ($Z = 1.024$, $p = 0.306$) or the average number of points collected per trial. This suggests that, on average, groups performed equally well.

However, a 2 (Stakes: Low, High) × 2 (Groups: Alcohol Non-Users, Alcohol Users) ANOVA showed a significant main effect of Stakes ($F_{79,1} = 5.469$, $p = 0.022$, partial $\eta^2 = 0.065$) and a Stakes × Groups interaction ($F_{79,1} = 6.463$, $p = 0.013$, partial $\eta^2 = 0.076$) on model-based control. The main effect of Group ($F_{79,1} = 0.504$, $p = 0.480$, partial $\eta^2 = 0.006$) was not statistically significant. Follow-up within-subject Wilcoxon tests revealed that the difference between the High and Low model-based weighting parameters was statistically significant in Alcohol Non-Users

($Z = -2.402$, $p = 0.016$), but not in Alcohol Users ($Z = -0.508$, $p = 0.611$). Mann-Whitney U tests revealed no significant between-group difference in mixing weight for Low ($Z = -1.876$, $p = 0.061$) nor High ($Z = -0.612$, $p = 0.540$) stakes (Fig. 2a). This result suggests that, even though performance on the task was comparable, Alcohol Users were less likely to adopt different strategies based on the reward incentive.

Utility Function
Next, we turned our attention to our measure of risk sensitivity from the utility function. A 2 x 2 ANOVA of showed a significant main effect of Stakes ($F_{76,1} = 36.08$, $p < 0.001$, partial $\eta^2 = .322$) and a Stakes × Groups interaction ($F_{76,1} = 4.524$, $p = 0.037$, partial $\eta^2 = 0.056$). There was no significant main effect of Group ($F_{76,1} = .992$, $p = 0.322$, partial $\eta^2 = 0.013$). Post-hoc analyses using independent *t*-tests found that risk sensitivity did not differ between groups on low stakes trials ($t = -1.461$, $p = 0.152$), but that on high stakes trials risk sensitivity was significantly higher in Alcohol Non-Users compared to Alcohol Users ($t = 2.280$, $p = 0.030$). Follow-up within-subject Wilcoxon tests revealed that the risk sensitivity difference between the High and Low stake trail was statistically significant both in Alcohol Non-users ($Z = -3.702$, $p < 0.001$), and in Alcohol Users ($Z = -2.737$, $p = 0.006$). Mann-Whitney U tests revealed a significant between-group difference in risk sensitivity for High ($Z = -2.016$, $p = 0.044$) and a significant trend for Low ($Z = 1.950$, $p = .051$) stakes. This result suggests that both Alcohol Users and Alcohol Non-users were sensitive to risk and reward magnitude (Fig. 2b).

Correlation analyses
Finally, we explored the relation between the parameters from the two computational models. Correlation analyses revealed that for High stakes trials in Alcohol Users, the weighting parameter (ω) was negatively correlated with risk sensitivity (μ) ($r = -0.349$, $p = 0.018$), such that in Alcohol Users, less model-free control was associated with greater risk aversive behavior. However, this relationship was not statistically significant in Alcohol Non-Users ($r = -0.229$, $p = 0.207$) (Fig. 3a). Further, Fisher's Z transformation indicated that these correlations were not significantly different between Alcohol Users and Alcohol Non-users ($z = -0.55$, $p = 0.5823$). Moreover, in Alcohol Non-Users, the greater difference between mixing weight for High relative to Low stakes ($\Delta\omega$) was associated with lower depressive symptoms (assessed via the PHQ-8; $r = -0.360$, $p = 0.040$), such that Alcohol Non-Users with lower depressive symptoms showed greater increase model-based decision making in high- relative to low-stakes trials (Fig. 3b). (Supplemental table S2 has additional analysis)

DISCUSSION
This study used a modified two-step task to examine the arbitration between model-based and model-free RL systems in Alcohol Users compared to Alcohol Non-Users. We found that Alcohol Non-Users and Alcohol Users did not differ in the overall reliance on model-based control during the task. However, unlike Alcohol Non-Users, Alcohol Users did not increase model-based control based on reward incentives. Additionally, even though both groups exhibited increased risk aversion in high- compared to low-stakes trials, this effect was pronounced for Alcohol Users. Lastly, in Alcohol Users greater model-based control was associated with less risk aversiveness in high reward stakes, and in Alcohol Non-Users greater increase in model-based control in high relative to low reward conditions was associated with lower depressive symptoms.

We have found evidence that alcohol use impairs the ability to arbitrate between decision making strategies. Alcohol Non-Users increased their use of model-based control on high-stakes trials, consistent with prior reports using healthy young adults [23,
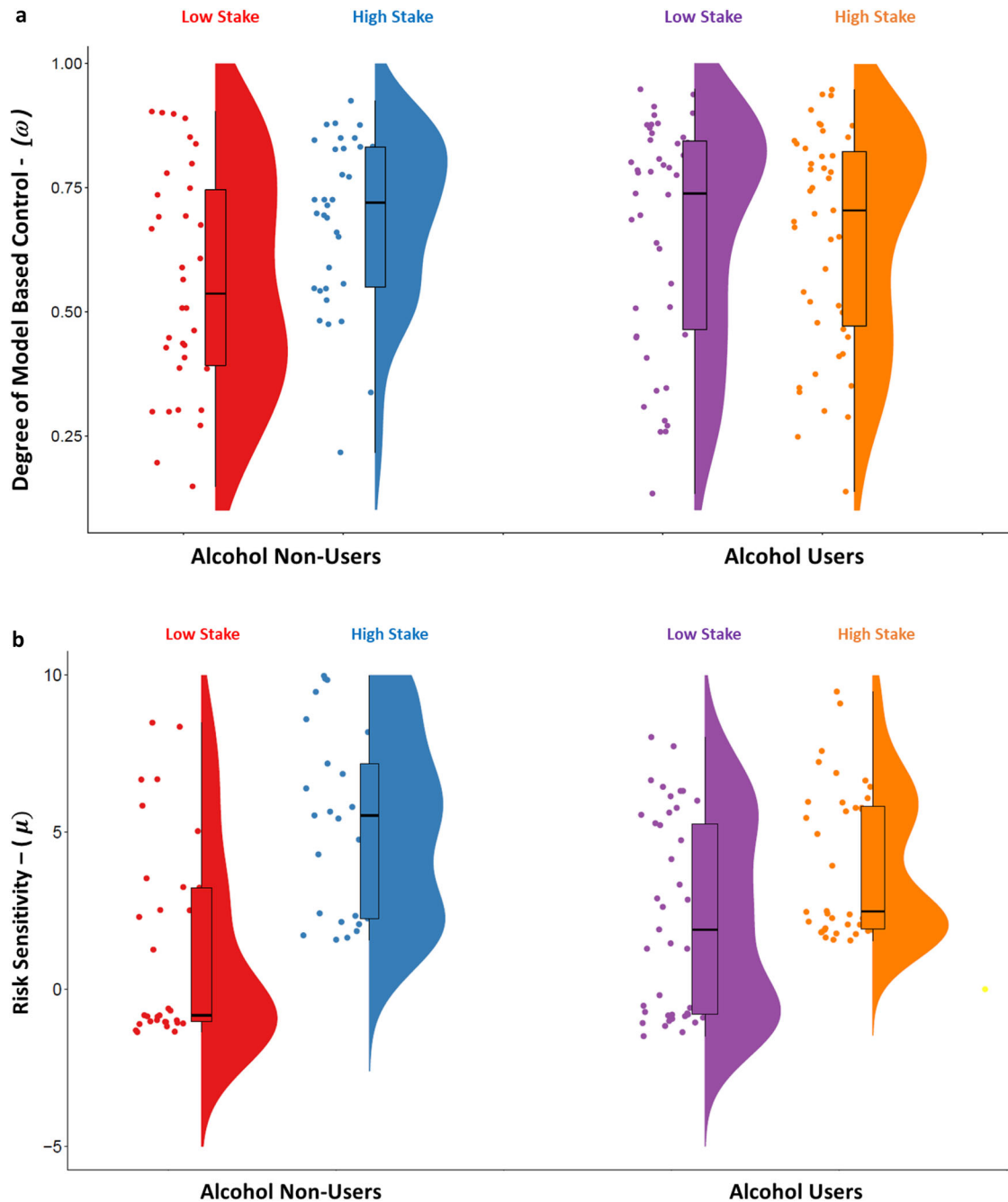
Fig. 2 Group differences in reward-related modulation in weighting parameter and risk sensitivity modulation. a Weighting parameter reflecting model-based decision control. **b** Risk sensitivity measure determining the tradeoff between risk avoidance and risk seeking.

31, 32]. However, Alcohol Users did not show such flexibility in arbitration. These findings suggest that people who use alcohol regularly struggle to shift between decision-making strategies based on contextual factors such as reward-related anticipatory cues and therefore make less optimal choices. Of course, arbitration between strategies is not only based on reward incentives. Inflexible arbitration between model-based and model-free control generalizes to other external influences. Indeed, a recent study showed that while Alcohol Non-Users employed less model-based control in high- compared to low-stress conditions, Alcohol Users did not show such a reduction in model-based control across stress levels [59]. Future work, varying other contextual variables, will broaden our understanding of how alcohol use affects controlled arbitration.

Importantly, some studies have previously shown greater reliance on model-free compared to model-based control in substance [42] and alcohol [41] users. We did not find such group differences. This can potentially be explained by differences in the task. In our modified two-step task, model-based control is more accurate than model-free control, which is a feature that is absent in prior versions of the task [27]. Moreover, model-based planning is more demanding in the older version, requiring reasoning over stochastic transitions (compared to the deterministic transitions in our version). Nevertheless, a recent study using young adults provided convergent evidence for our findings, in the form of a lack of an association between alcohol use and model-based control [43].

The reward-magnitude manipulation in our task also changed participants' explore-exploit tradeoff (results are presented in

Table 1. Demographics of Alcohol Non-Users and Alcohol Users

Demographic Details	Alcohol Non-Users n = 34	Alcohol Users n = 47	t-test statistics
Gender			
Male	11 (0.32)	6 (0.26)	t(77) = 1.083, p = 0.282
Female	22 (0.65)	17 (0.74)	
Other	1 (0.03)	4 (0.09)	
Age			
<18 Years	9 (0.26)	2 (0.04)	t(75) = −4.395, p < 0.001
18–19 Years	4 (0.12)	4 (0.08)	
20–21 Years	4 (0.12)	2 (0.04)	
22–23 Years	4 (0.12)	12 (0.26)	
24–5 Years	1 (0.03)	2 (0.04)	
25–40 Years	12 (0.35)	9 (0.19)	
>40 Years	0 (0)	12 (0.26)	
Hispanic	6 (0.18)	10 (0.21)	t(75) = −0.596, p = 0.553
Race			
White	19 (0.56)	26 (0.56)	t(73) = −0.638, p = 0.529
Black	1 (0.03)	7 (0.15)	
Other	8 (0.23)	14 (0.3)	
Education			
High School /Diploma/GED	10 (0.3)	13 (0.28)	t(75) = −0.072, p = 0.943
College degree/ associate degree	16 (0.47)	19 (0.4)	
Post Graduate degree	8 (0.24)	11 (0.23)	
Mental Health Diagnosis	12 (0.35)	16 (0.34)	t(61) = −0.930, p = 0.356
AUDIT*	19 (0.89 ± 0.21)	40 (8.5 ± 12.32)	p < 0.001

*Continuous variables are represented by mean and standard deviation, while categorical variables are represented by percentages.

Supplementary Section). On high-stakes trials, participants reduced exploration and increased exploitation, picking the high-value action in the first stage of our task, instead of choosing the alternative. This pattern, which we have also previously reported in healthy individuals [23], did not differ between the Alcohol Users and Alcohol Non-Users. This finding assuages the potential concern that Alcohol Users simply become less sensitive to rewards and were therefore less inclined to use more model-based control. Instead, these results reinforce the idea that Alcohol Users are less flexible in arbitrating between RL strategies. However, a previous study found reduced exploratory behavior in individuals with alcohol use disorder [60]. Unfortunately, differences between this task and ours limit our ability to make direct comparisons.

Our results also showed that Alcohol Users showed stronger transfer of reward learning from the second-stage to the first-stage of the task. In other words, the RPE in the second stage of the task had a comparatively greater influence on reward expectations in the first-stage state of the following trial for users compared to Alcohol Non-Users. This means that the model-free systems of the Alcohol Users reflect outcomes of the second stage from the previous trials more so than it does in Alcohol Non-Users. This finding is consistent with previous research indicating that individuals with a history of substance and alcohol use encounter difficulties in transferring their learning effectively [61–64].

Specifically, they tend to make random choices, relying on past experiences to judge the effectiveness of new observations and behaviors [35, 65]. The overall implication is that substance and Alcohol Users apply acquired knowledge differently in novel situations, possibly due to challenges in making adaptive decisions.

We also documented that risk sensitivity was higher for the high- relative to low-stakes conditions for both groups, which is consistent with prior literature [66]. Interestingly, however, although Alcohol Users showed greater risk aversiveness in high- compared to low-stakes trials, their risk aversiveness for high stakes was still lower than that of Alcohol Non-Users, and the increase in risk aversiveness from low- to high-stakes trials was lower than that in Alcohol Non-Users. These results are consistent with prior studies showing that Alcohol Users show more risky decision making in high-reward magnitude contexts compared to Alcohol Non-users [35]. Moreover, although model-based control is typically associated with risk aversion and loss aversion [67], our correlational results revealed that in Alcohol Users, greater model-based control was associated with lower risk aversiveness during high reward stakes, although the result did not survive validation analyses. This behavior may be influenced more by emotional instead of cognitive assessment of risk [68–70]. This highlights the notion of maladaptive decision-making strategies/models employed by Alcohol Users [59], greater use of which is associated with riskier decision-making.

The results of this study should be viewed considering some potential limitations. First, the characterization of alcohol use which led to group stratification was determined based on ordinal self-reported data from the CRISIS questionnaire, a measure designed to assess changes in substance use and psychiatric symptoms during the COVID-19 pandemic-related lockdowns and our survey data was all collected during the pandemic time window. Given our stratification measure and time window of data collection, the alcohol use group included in our study may reflect a stress-induced heavy-drinking population more so than a heavy-drinking population in the absence of external stressors. Thus, the generalizability of our sample to other populations may be another limitation. However, it is important to note that the group characterization was supported by significantly higher AUDIT scores in Alcohol Users compared to Alcohol Non-Users, validating the group stratification based on alcohol use severity. However, comparison of results with those from prior studies is limited since the clinical characteristics beyond AUDIT scores were not assessed in this study. Second, even though our results show that alcohol use diminishes the ability to shift between RL strategies, it remains unclear why this is the case. One possibility is that Alcohol Users are impaired in estimating the relative values of both habitual and goal-directed strategies, leading to reduced uncertainty about the appropriate strategy in dynamically changing contexts. Another possibility is that Alcohol Users attach a higher cost to switching between strategies, which would suggest that their lack of arbitration reflects a motivational deficit. Even though our data are unable to distinguish between these explanations, they invite a novel research program aimed at doing so. The third limitation of this survey-based study is the lack of assessment of fluid intelligence and working memory. This could have influenced the task performance and therefore limits the generalizability of these results. Nevertheless, these results provide an evidence-based foundation for future studies that should assess alcohol use and other determinants of decision making in more detail to examine the impact of frequency and recency of alcohol use on reward-related decision making. Fourth, the lack of assessment of other specific mental health disorders may have confounded these results. The CRISIS questionnaire includes data on whether participants had any history of mental health diagnoses (binary: yes or no), and the proportion of those with mental health prior diagnoses was comparable between the
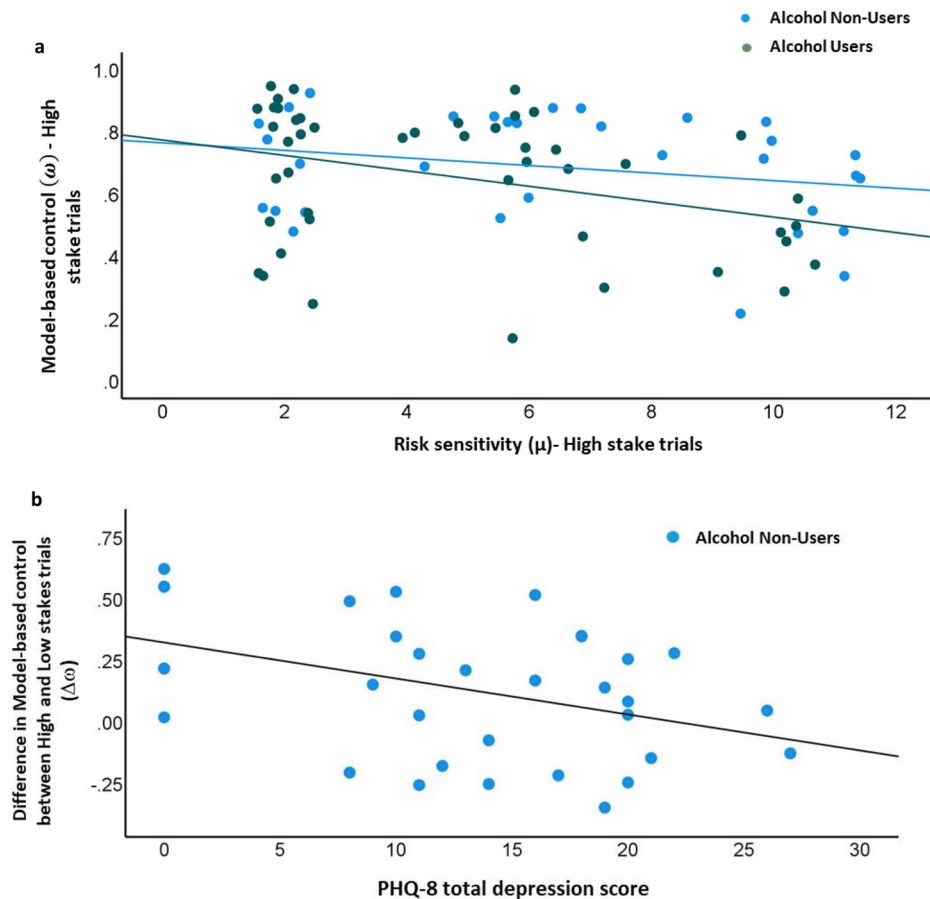
Fig. 3 Correlations between model parameters and clinical variables. a Correlation between model-based vs -free weighting parameter and Risk Sensitivity during high stakes trials in the total sample, as a function of Alcohol User and Non User group membership. **b** Correlation between model-based vs -free weighting parameter and severity of depressive symptoms in the Non User sample.

two groups. Lastly, this study could have benefited from a larger sample size with varying alcohol use severity such that computational RL models can be effectively employed to understand the relationship of reward-related decision making and alcohol use severity.

CONCLUSIONS
We have shown that frequent alcohol use is associated with less flexible arbitration between goal-directed (i.e., model-based) and habit-based (i.e., model-free) control in situations where reward magnitude increases, and with lower risk aversiveness in high reward magnitude contexts. Although we expected greater goal-directed control to be associated with greater risk aversiveness, this relationship was inverted in Alcohol Users suggesting that the decision models developed in Alcohol Users may not only be rigid to varying reward values, but also associated with risky decision making. Results from this behavioral study paves the way for neuroimaging studies to further examine neurobiological underpinnings of maladaptive reward-related decision-making strategies in Alcohol Users and to examine these effects in other substance and/or behavioral addictions.

DATA AVAILABILITY

REFERENCES
1. CDC, C.F.D.C.A.P. *The Cost of Excessive Alcohol Use.* January 22, 2024]; Available from: https://www.cdc.gov/media/images/releases/2015/p1015-excessive-alcohol.pdf.
2. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 national and state costs of excessive alcohol consumption. Am J Prev Med. 2015;49:e73–e79.
3. Boschuetz N, Cheng S, Mei L, Loy VM. Changes in alcohol use patterns in the United States During COVID-19 pandemic. WMJ. 2020;119:171–6.
4. Karaye IM, Maleki N, Hassan N, Yunusa I. Trends in alcohol-related deaths by sex in the US, 1999-2020. JAMA Netw Open. 2023;6:e2326346.
5. Meyers JL, McCutcheon VV, Horne-Osipenko KA, Waters LR, Barr P, Chan G, et al. COVID-19 pandemic stressors are associated with reported increases in frequency of drunkenness among individuals with a history of alcohol use disorder. Transl psychiatry. 2023;13:311.
6. Tanabe J, Reynolds J, Krmpotich T, Claus E, Thompson LL, Du YP, et al. Reduced neural tracking of prediction error in substance-dependent individuals. Am J Psychiatry. 2013;170:1356–63.
7. Byrne KA, Otto AR, Pang B, Patrick CJ, Worthy DA. Substance use is associated with reduced devaluation sensitivity. Cogn Affect Behav Neurosci. 2019;19:40–55.
8. Parvaz MA, Maloney T, Moeller SJ, Woicik PA, Alia-Klein N, Telang F, et al. Sensitivity to monetary reward is most severely compromised in recently abstaining cocaine addicted individuals: a cross-sectional ERP study. Psychiatry Res. 2012;203:75–82.
9. Charles-Walsh K, Upton DJ, Hester R. Examining the interaction between cognitive control and reward sensitivity in substance use dependence. Drug Alcohol Depend. 2016;166:235–42.
10. Parvaz MA, Konova AB, Tomasi D, Volkow ND, Goldstein RZ. Structural integrity of the prefrontal cortex modulates electrocortical sensitivity to reward. J Cogn Neurosci. 2012;24:1560–70.
11. Bart CP, Nusslock R, Ng TH, Titone MK, Carroll AL, Damme K, et al. Decreased reward-related brain function prospectively predicts increased substance use. J Abnorm Psychol. 2021;130:886–98.

S.A. Ramakrishnan et al.

12. Galandra C, Basso G, Cappa S, Canessa N. The alcoholic brain: neural bases of impaired reward-based decision-making in alcohol use disorders. Neurol Sci. 2018;39:423–35.

13. Parvaz MA, Kim K, Froudist-Walsh S, Newcorn JH, Ivanov I. Reward-based learning as a function of severity of substance abuse risk in drug-naïve youth with ADHD. J Child Adolesc Psychopharmacol. 2018;28:547–53.

14. Wunderlich K, Dayan P, Dolan RJ. Mapping value based planning and extensively trained choice in the human brain. Nat Neurosci. 2012;15:786–91.

15. Wunderlich K, Rangel A, O'Doherty JP. Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci. 2009;106:17199–204.

16. Boureau YL, Sokol-Hessner P, Daw ND. Deciding how to decide: self-control and meta-decision making. Trends Cogn Sci. 2015;19:700–10.

17. Kool W, McGuire JT, Rosen ZB, Botvinick MM. Decision making and the avoidance of cognitive demand. J Exp Psychol: Gen. 2010;139:665–82.

18. Eppinger B, Goschke T, Musslick S. Meta-control: From psychology to computational neuroscience. Cogn Affect Behav Neurosci. 2021;21:447–52.

19. Smid CR, Ganesan K, Thompson A, Cañigueral R, Veselic S, Royer J, et al. Neurocognitive basis of model-based decision making and its metacontrol in childhood. Dev Cogn Neurosci. 2023;62:101269.

20. Dolan RJ, Dayan P. Goals and habits in the brain. Neuron. 2013;80:312–25.

21. Sutton RS, Barto A, Reinforcement Learning: An Introduction. 1998, 9, 1054: Cambridge: MIT Press.

22. Geffner, H, *Model-free*, model-based, and general intelligence. arXiv preprint arXiv:1806.02308, 2018.

23. Kool W, Gershman SJ, Cushman FA. Cost-benefit arbitration between multiple reinforcement-learning systems. Psychol Sci. 2017;28:1321–33.

24. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011;69:1204–15.

25. Haith, AM and JW Krakauer, Model-based and model-free mechanisms of human motor learning, In *Progress in motor control: Neural, computational and dynamic approaches*. 2013, 782, Springer. p. 1–21.

26. Filipowicz, AL, et al. *The complexity of model-free and model-based learning strategies*. BioRxiv, 2020.

27. Kool W, Cushman FA, Gershman SJ. When does model-based control pay off? PLoS Comput Biol. 2016;12:e1005090.

28. Patzelt EH, Kool W, Millner AJ, Gershman SJ. Incentives boost model-based control across a range of severity on several psychiatric constructs. Biol Psychiatry. 2019;85:425–33.

29. Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci. 2013;24:751–61.

30. Kool W, Gershman SJ, Cushman FA. Planning complexity registers as a cost in metacontrol. J Cogn Neurosci. 2018;30:1391–404.

31. Bolenz F, Eppinger B. Valence bias in metacontrol of decision making in adolescents and young adults. Child Dev. 2022;93:e103–e116.

32. Bolenz, F, Kool W, Reiter AM, Eppinger B. Metacontrol of decision-making strategies in human aging. Elife, 2019; 8:e49154.

33. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR. Model-free and model-based influences in addiction-related behaviors. Biol Psychiatry. 2019;85:936–45.

34. Fraser KM, Janak PH. How does drug use shift the balance between model-based and model-free control of decision making. Biol Psychiatry. 2019;85:886–8.

35. Verdejo-Garcia A, Chong TT, Stout JC, Yücel M, London ED. Stages of dysfunctional decision-making in addiction. Pharm Biochem Behav. 2018;164:99–105.

36. Voon V, Morris LS, Irvine MA, Ruck C, Worbe Y, Derbyshire K, et al. Risk-taking in disorders of natural and drug rewards: neural correlates and effects of probability, valence, and magnitude. Neuropsychopharmacology. 2015;40:804–12.

37. Lawrence AJ, Luty J, Bogdan NA, Sahakian BJ, Clark L. Problem gamblers share deficits in impulsive decision-making with alcohol-dependent individuals. Addiction. 2009;104:1006–15.

38. Campbell JA, Samartgis JR, Crowe SF. Impaired decision making on the Balloon Analogue Risk Task as a result of long-term alcohol use. J Clin Exp Neuropsychol. 2013;35:1071–81.

39. Ariesen AD, Neubert JH, Gaastra GF, Tucha O, Koerts J. Risky decision-making in adults with alcohol use disorder—a systematic and meta-analytic review. J Clin Med. 2023;12:2943.

40. Dave D, Saffer H. Alcohol demand and risk preference. J Econ Psychol. 2008;29:810–31.

41. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hägele C, et al. Model-based and model-free decisions in alcohol dependence. Neuropsychobiology. 2014;70:122–31.

42. Lucantonio F, Caprioli D, Schoenbaum G. Transition from 'model-based' to 'model-free' behavioral control in addiction: Involvement of the orbitofrontal cortex and dorsal striatum. Neuropharmacology. 2014;76:407–15. Pt B

43. Nebe S, Kroemer NB, Schad DJ, Bernhardt N, Sebold M, Müller DK, et al. No association of goal-directed and habitual control with alcohol consumption in young adults. Addict Biol. 2018;23:379–93.

44. Nikolaidis A, Paksarian D, Alexander L, Derosa J, Dunn J, Nielson DM, et al. The Coronavirus Health and Impact Survey (CRISIS) reveals reproducible correlates of pandemic-related mood states across the Atlantic. Sci Rep. 2021;11:8139.

45. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. Addiction. 1993;88:791–804.

46. Williams N. The AUDIT questionnaire. Occup Med. 2014;64:308.

47. Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry. 2010;32:345–59.

48. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care. 2008;46:266–74.

49. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND. Model-based choices involve prospective neural activity. Nat Neurosci. 2015;18:767–72.

50. de Leeuw JR. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. Behav Res Methods. 2015;47:1–12.

51. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005;8:1704–11.

52. Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 2010;66:585–95.

53. Rummery, GA, & Niranjan, M, *On-line Q-learning using connectionist systems*. Cambridge, UK: University of Cambridge, Department of Engineering, 1994. 37: p. 14.

54. Gershman SJ. Empirical priors for reinforcement learning models. J Math Psychol. 2016;71:1–6.

55. Balasubramani PP, Chakravarthy VS, Ravindran B, Moustafa AA. An extended reinforcement learning model of basal ganglia to understand the contributions of serotonin and dopamine in risk-based decision making, reward prediction, and punishment learning. Front Comput Neurosci. 2014;8:47.

56. D'acremont M, Lu ZL, Li X, Van der Linden M, Bechara A. Neural correlates of risk prediction error during reinforcement learning in humans. Neuroimage. 2009;47:1929–39.

57. Bell DE. Risk, Return, and Utility. Manag Sci. 1995;41:23–30.

58. Mitchell, M. An introduction to genetic algorithms. Complex adaptive systems. 1996, Cambridge, Mass.: MIT Press. viii, 205 p.

59. Wyckmans, F, et al. *Reinforcement Learning Strategies in Alcohol-Use Disorder: Examining the Effects of Provoked Stress*. 2023.

60. Topiwala A, Allan CL, Valkanova V, Zsoldos E, Filippini N, Sexton C, et al. Moderate alcohol consumption as risk factor for adverse brain outcomes and cognitive decline: longitudinal cohort study. BMJ. 2017;357:j2353.

61. Erdozain AM, Morentin B, Bedford L, King E, Tooth D, Brewer C, et al. Alcohol-related brain damage in humans. PLoS One. 2014;9:e93586.

62. Kanen JW, Ersche KD, Fineberg NA, Robbins TW, Cardinal RN. Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. Psychopharmacology. 2019;236:2337–58.

63. Smith R, Schwartenbeck P, Stewart JL, Kuplicki R, Ekhtiari H, Paulus MP, et al. Imprecise action selection in substance use disorder: Evidence for active learning impairments when solving the explore-exploit dilemma. Drug Alcohol Depend. 2020;215:108208.

64. Verdejo-Garcia A, Garcia-Fernandez G, Dom G. Cognition and addiction. Dial Clin Neurosci. 2019;21:281–90.

65. Petry NM, Bickel WK, Arnett M. Shortened time horizons and insensitivity to future consequences in heroin addicts. Addiction. 1998;93:729–38.

66. Hintze A, Olson RS, Adami C, Hertwig R. Risk sensitivity as an evolutionary adaptation. Sci Rep. 2015;5:8242.

67. Solway A, Lohrenz T, Montague PR. Loss aversion correlates with the propensity to deploy model-based control. Front Neurosci. 2019;13:915.

68. Wang N, Wilhite SC, Wyatt J, Young T, Bloemker G, Wilhite E. Impact of a college freshman social and emotional learning curriculum on student learning outcomes: an exploratory study. J Univ Teac Learn Pract. 2012;9. https://doi.org/10.53761/1.9.2.8.

69. Tyng CM, Amin HU, Saad MNM, Malik AS. The influences of emotion on learning and memory. Front Psychol. 2017;8:235933.

70. Decker JH, Otto AR, Daw ND, Hartley CA. From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. Psychol Sci. 2016;27:848–58.

AUTHOR CONTRIBUTIONS

performed computational modeling on the behavioral data. S.A.R. and R.B.S. led the preparation of the manuscript. V.G.F., C.J.H., S.S., and V.S.C. provided valuable direction during the manuscript preparation. All authors contributed to the preparation of the manuscript.

COMPETING INTERESTS
The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper. Dr. Hammond receives grant support from the National Institute on Drug Abuse (NIDA) (R01DA058303, Bench to Bedside Award, R33DA056230-02W1, R01DA056920), the American Academy of Child & Adolescent Psychiatry (AACAP), the Substance Abuse and Mental Health Services Administration (SAMHSA, H79 SP082126-01), the Doris Duke Charitable Foundation (Grant# 2020147), the National Network of Depression Centers (NNDC), and the Johns Hopkins University School of Medicine, and has served as a subject matter expert and consultant for SAMHSA.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44277-024-00023-8.

Correspondence and requests for materials should be addressed to Muhammad A. Parvaz.

Reprints and permission information is available at http://www.nature.com/reprints

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