Rostral Anterior Cingulate Activations Inversely Relate To Reward Payoff Maximation & Predict Depressed Mood

Pragathi Priyadharsini Balasubramani (pbalasubramani@health.ucsd.edu)
University of California, San Diego
Juan Diaz-Delgado
University of California, San Diego
Gillian Grennan
University of California, San Diego
Fahad Alim
University of California, San Diego
Mariam Zafar-Khan
University of California, San Diego
Vojislav Maric
University of California, San Diego
Dhakshin Ramanathan
University of California, San Diego
Jyoti Mishra
University of California, San Diego

Research Article

Keywords: reward prediction, risk sensitivity, reinforcement learning, anterior cingulate cortex, depression

Posted Date: October 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-996915/v1

Rostral Anterior Cingulate Activations inversely relate to Reward Payoff Maximation & predict Depressed Mood

Pragathi Priyadharsini Balasubramani[1*], Juan Diaz-Delgado[1], Gillian Grennan[1], Fahad Alim[1], Mariam Zafar-Khan[1], Vojislav Maric[1], Dhakshin Ramanathan[1,2], Jyoti Mishra[1*]

[1]Neural Engineering and Translation Labs, Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

[2]Department of Mental Health, VA San Diego Medical Center, San Diego, CA

[*]Correspondence should be addressed to:
Pragathi Priyadharsini Balasubramani & Jyoti Mishra
University of California, San Diego
Neural Engineering & Translation Labs (NEAT Labs)
9500 Gilman Drive Mail Code 0875
La Jolla, CA 92037
pbalasubramani@health.ucsd.edu & jymishra@health.ucsd.edu

Abstract

Choice selection strategies and decision making are typically investigated using multiple-choice gambling paradigms that require participants to maximize reward payoff. However, research shows that performance in such paradigms suffers from individual biases towards the frequency of gains to choose smaller local gains over larger longer term gain, also referred to as melioration. Here, we developed a simple two-choice reward task, implemented in 186 healthy human adult subjects across the adult lifespan to understand the behavioral, computational, and neural bases of payoff maximization versus melioration. The observed reward choice behavior on this task was best explained by a reinforcement learning model of differential future reward prediction. Simultaneously recorded and source-localized electroencephalography (EEG) showed that diminished theta-band activations in the right rostral anterior cingulate cortex (rACC) correspond to greater reward payoff maximization, specifically during the presentation of cumulative reward information at the end of each task trial. Notably, these activations (greater rACC theta) predicted self-reported depressed mood symptoms, thereby showcasing a reward processing marker of potential clinical utility.

Keywords. reward prediction; risk sensitivity; reinforcement learning; anterior cingulate cortex; depression

Introduction

Cognitive and neural responses to reward and risk are quintessential to understanding human behavior. Gambling tasks predominantly form the experimental test beds for measuring reward and risk processing abilities in humans [1,2]. However, it is widely debated how well these tasks can separate decision-making based on frequency of gains/losses versus expected value of payoff for different choice-sets [1,3,3–12].

For instance, studies that have controlled for gain frequency in the Iowa Gambling Task (IGT) show that subject choices reflect their inherent gain frequency preferences in the task, which portray relatively immediate reinforcement based choice behavior rather than a behavior that maximizes expected value or long-term payoff [13–17]. In contrast, some researchers suggest such local minima choices also referred to as melioration may be the rational solution in the face of uncertainty [18]. Additionally, task related performance measures for long-term payoff versus immediate gain frequency bias may reflect distinct reinforcement learning and decision making strategies, e.g. sensitivity to rewards and risks, learning rate, and other behavioral execution strategies [19–23], which prior neural studies of reward payoff maximization have not accounted for.

In this study, we uniquely separate immediate gain frequency bias driven decision-making from advantageous longer term payoff-based decision making. Specifically, we designed a two-choice paradigm with two distinct blocks – a $\Delta_0$payoff (baseline) block where two reward choice options have equal payoffs and reward variance suitable for measuring the immediate gain frequency bias, and a Δpayoff (difference) block where the two-choice options have unequal payoffs suitable for measuring payoff influences. We thereby, tease apart measurements of immediate gain frequency biased response from

expected value or long-term payoff based response, to understand the distinct cognitive and neural mechanisms underlying payoff decisions.

Second, we capitalize on computational reinforcement learning (RL) models to understand the basis of individual differences in reward and risk based learning across subjects [24–28]. The RL framework provides the ability to simulate the observed behavior and estimate the hidden parameters forming the basis for individual differences in learning and performance. Specifically in this study, we checked whether the observed subject behavior for reward payoff maximization versus gain frequency bias melioration can be explained by extent of future reward prediction [29], or differential risk seeking towards gains versus losses [25].

Neurally, earlier studies have suggested the significant role of frontal executive regions, particularly the medial prefrontal cortex (mPFC) in reward and risk based performance [1–3,30–32]. Uniquely, in this study, we estimate the neural correlates for payoff relevant decisions while accounting for significant RL parameters and individual differences in immediate gain frequency biases. Finally, we understand the relationship between the identified neural correlates for reward maximization to individual's self-reported mood symptom variations. We show that reward payoff maximization correlates within mPFC are sensitive to depressed mood, and hence our experimental and computational framework for identifying the regions of interest in mPFC may serve future clinical utility as a biomarker for depression.

Results

Options associated with larger payoff show win-stay behavior and their choice fraction is proportional to individual gain frequency bias.

Healthy adult subjects (N = 186, ages 18-80 years, 115 females) performed a two-choice gambling task, *Lucky Door*, which implemented two distinct blocks of choices; the $\Delta_0$payoff block delivered choice-sets with different gain frequencies but no differences in payoff, while the Δpayoff block delivered choice-sets with same gain frequencies as the $\Delta_0$payoff block yet with long-term payoff (i.e. expected value) differences (**Figure 1A**; **Supplementary Table 1**). Specifically, the $\Delta_0$payoff block only varied the gain frequency associated with the choice doors, with one door leading to 70% positive reward outcomes (Rare Loss or RareL door) while the other resulting in 70% negative reward outcomes (Rare Gain or RareG door), yet maintaining the same reward average or long-term expected value/payoff. The Δpayoff block, presented in a random sequence order relative to the $\Delta_0$payoff block across subjects, had the same gain and loss frequency setup as the $\Delta_0$payoff block, but the rewards associated with the RareG door resulted in a larger long-term payoff than the RareL door. Participants executed 40 trials per block. Thus, we calculated the payoff related performance measure, *Perf* as the difference in probability of RareG selections on Δpayoff vs. $\Delta_0$payoff blocks. Gain frequency *Bias* was measured as the difference in proportion of RareL vs. RareG choices on the $\Delta_0$payoff block.

In terms of subject behavior, the actual proportion of RareG choices between the two blocks did not differ (paired Wilcoxon signed-rank test: N_{RareG} Δpayoff:17.74±0.39; $\Delta_0$payoff:18.47±0.40; p=0.07). But, we found that Win-Stay behavior for the RareG choices distinguished performance on the two blocks (mean±SEM Win-Stay RareG,

Δpayoff:0.61±0.02; $\Delta_0$payoff:0.55±0.02; block difference: 0.06± 0.03, p=0.02, **Figure 1B**). Corresponding lose-shift behavior did not differ between blocks (p=0.33).

Additionally, we found that the payoff-based responses, *Perf* were significantly correlated to individual gain frequency *Bias* (r=0.65, p<0.0001, **Figure 1C**).

Demographics	**Median ± MAD**
Age	25.00 ± 14.87
Gender n (%)	
Male	71 (38.17)
Female	115 (61.83)
Ethnicity n (%)	
Caucasian	116 (62.37)
Black or African American	4 (2.15)
Native Hawaiian, Pacific Islander	0 (0)
Asian	37 (19.89)
American Indian, Alaska Native	4 (2.15)
Multi-racial	12 (6.45)
Others	12 (6.45)
Race n (%)	
Hispanic or Latino	25 (13.51)
Not Hispanic or Latino	155 (83.78)
Unknown	5 (2.70)
SES	5.00 ± 1.34
Mental Health	**Median ± MAD**
Anxiety	3 ± 2.88
Depression	3 ± 2.76
Inattention	4 ± 3.99
Hyperactivity	3 ± 2.96
Behavior	**Median ± MAD**
Perf	-0.02±0.13
Bias	0.10±0.21
Model γ	0.99± 0.07

Table 1. Subject characteristics. Median ± MAD for subjects demographics variables, mental health self-report scores, and parameter γ from the preferred reinforcement learning model. MAD: median absolute deviation, SES: socioeconomic status score.

Next, we implemented multivariate regression to model the payoff-related performance, *Perf*, based on all self-reported demographic (age, gender, race, ethnicity, socio-

economic status SES) and mental health (anxiety, depression, inattention and hyperactivity) predictors as per **Table 1**. This model also included individual gain frequency *Bias* and order of block presentation. The overall *Perf* model was significant (R^2=0.43, p<0.0001). Interestingly the only variable that significantly predicted *Perf* was Gain frequency *Bias* (β=0.37 ±0.04, t(151)=9.18, p<0.0001). Thus, hereafter, we control for *Bias* in all payoff-relevant analyses below.

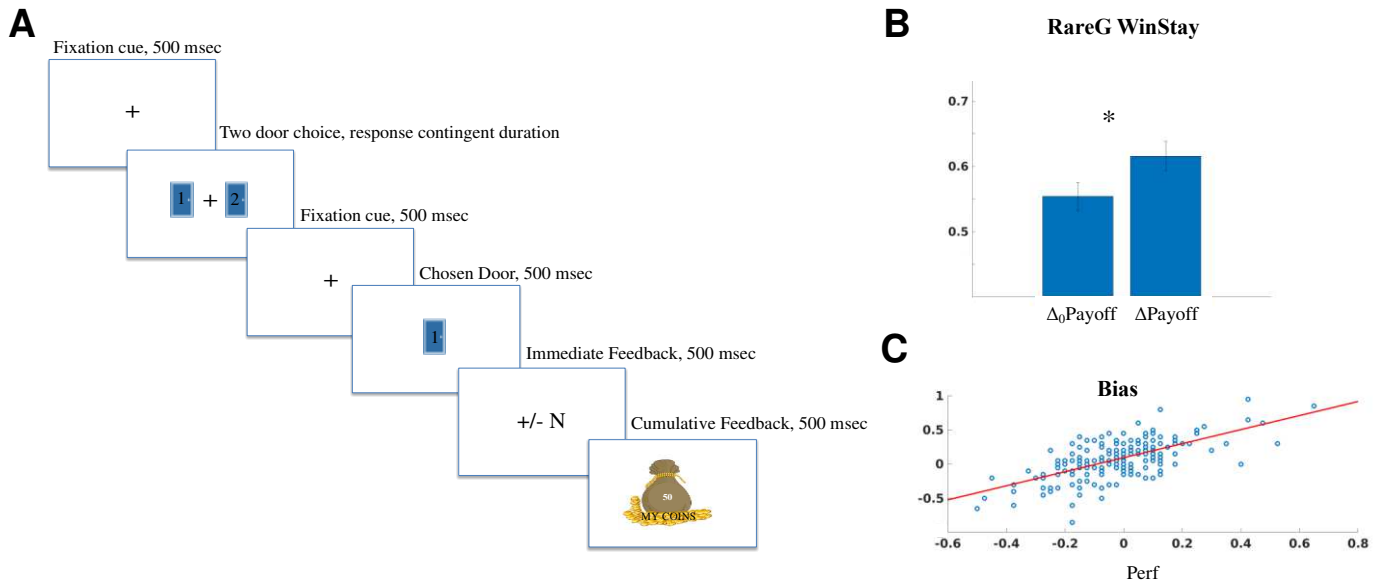

Figure 1. Reward task and associated behavior A) As per the task schematic, participants fixated for 0.5 sec, then chose from one of two choice doors. Post-response, fixation reappeared for 0.5 sec, followed by presentation of the chosen door for 0.5 sec, then immediate gain or loss feedback provided for 0.5 sec, and finally, cumulative feedback of all gains/losses up to the present trial shown for 0.5 sec. Reward distributions for the door choices are presented in Supplementary Table 1. **B)** Win-Stay behavior for the rare gain RareG door is significantly greater on the ΔPayoff versus Δ_0Payoff block. *:p<.05. **C)** Gain frequency *Bias* significantly predicts payoff performance, *Perf* (r=0.65, p<0.0001).

Individual differences in reward prediction explain task behavior.

Prior RL modeling studies have suggested that task behavior can be explained with greater accuracy when both the long term expected value or payoff and the expected risk are accounted for during decision-making [25,33]. The major advantages of building these RL models are that the models can characterize the individual differences in each participant in terms of computational parameters that manifest in learning and executive control. Secondly, they can explain the converged behavioral dynamics in each participant sans experimental trial limitations.

Using this RL framework, we were interested in investigating how are choice decisions affected by 1) extent of integration of rewards over time i.e. time scale of reward prediction, and 2) differential risk sensitivity to gains and losses affecting the choices, where risk is the variance in reward outcomes. In order to find whether the observed subjective behavioral differences are driven by one of the two or both of the above decision making measures, we built three separate reinforcement learning models (RL, **Figure 2A**) that simulate behavior by optimizing time scale of reward prediction (γ: model A), risk sensitivity (α: model B) or both γ and α parameters (model C). In model A, higher γ parameter values represent long-sightedness; risk sensitivity α is set to 0 in this model. In model B, higher α parameter values indicate risk aversiveness and lower values indicate risk seeking; the reward prediction γ parameter is set to 0 in this model. All models were optimized using robust solvers in MATLAB (see Methods to find global minima with multiple random initial conditions) to match to the observed choice selection behavior observed in each block.

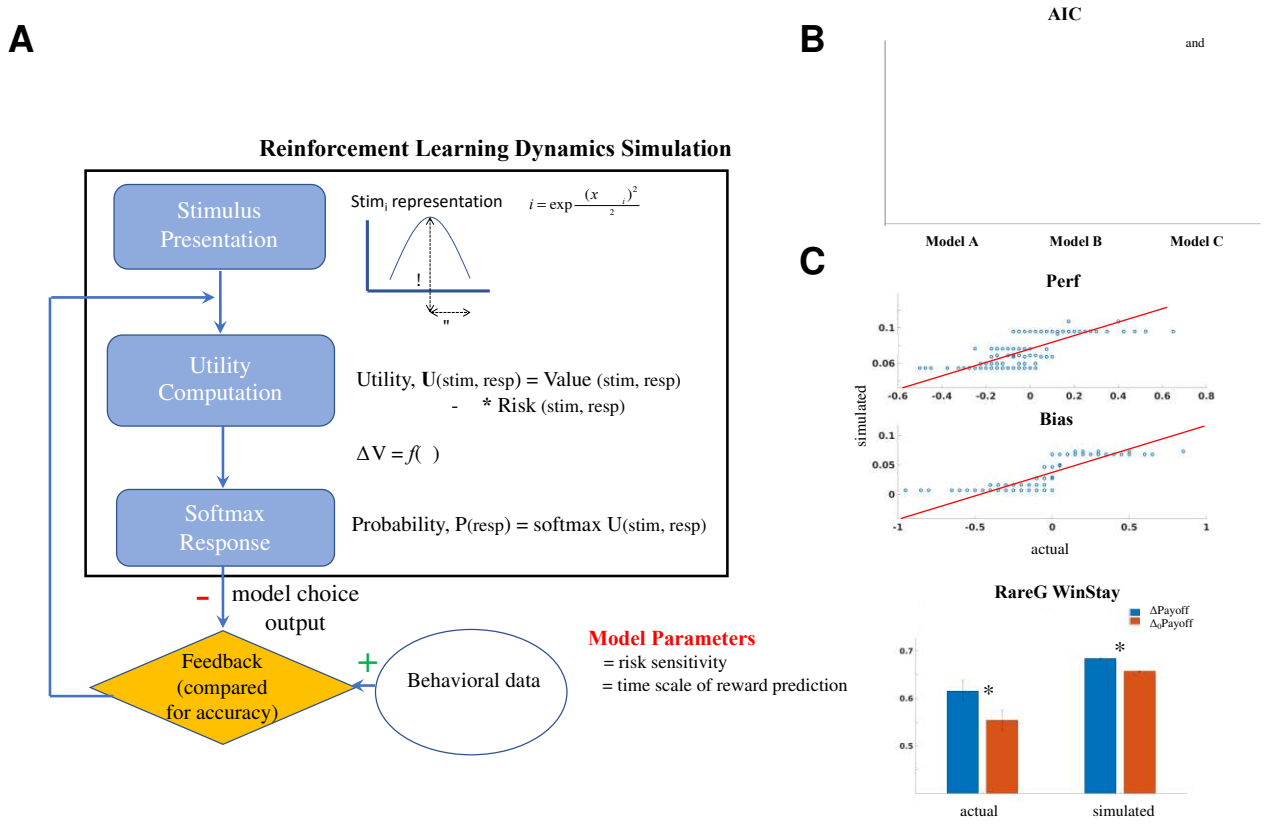

Figure 2. Reinforcement Learning Model (A) Schematic representing the stimulus, value function and choice selection modules. The model results for number of selections associated with each of the choice door stimuli in each task block are compared against the actual selections made by each subject, for purposes of model optimization. The model uses the utility, U, associated with each choice response for making the decision, where the utility is a function of reward average and reward variance associated with choices. The decision in the model is taken using the SoftMax probability, P, of making the choices. Model parameters are highlighted as α (model agent's differential risk sensitivity to gain and loss outcome uncertainties), and γ (time scale of reward prediction). **(B)** AIC values for three models, Model A (γ optimized), B (α optimized), C (γ and α optimized), show that the γ model was best performing. **(C)** The γ optimized model showed strong correlations between simulated and actual *Perf* (ρ(185)=0.81, p<0.0001), *Bias* (ρ(185)=0.89, p<0.0001) and RareG Win-Stay difference between blocks (ρ(185)=0.17 p=0.004).

Next, we computed the likelihood of any Model A, B, C to explain the significantly observed behavioral measures: which are *Perf, Bias* and the block differences in proportion of *Win-Stay* to the RareG door. The likelihood, taken as correlations between the simulated and actually observed, was used to construct model AIC (Akaike

Information Criterion). The AIC values for Models A, B and C were 2.94, 3.70 and 5.45, respectively (**Figure 2B**), suggesting that Model A can preferably explain the observed behavior (**Figure 2C**). This means the individual differences in the time scale of reward prediction can preferably explain the reward maximization (*Perf*) as well as melioration (*Bias*) behavior. On parameter recovery analysis for this model, the simulated model solutions were significantly correlated with the recovered solutions, implying that our model can reliably explain the subjective behaviors [34].

Right rostral anterior cingulate cortex encodes reward payoff maximization.

Participants performed the reward task with simultaneous EEG, which we analyzed in the theta (3-7 Hz), alpha (8-12 Hz), and beta (13-30 Hz) frequency bands in cortical source space parcellated as per the Desikan-Killiany regions of interest [35]. To identify the neural correlates underlying reward payoff maximization (*Perf*), we modeled the neural variables as predictors of *Perf* using robust multivariate linear regression, while accounting for gain frequency *Bias* that was significantly related to *Perf* (**Figure 1**), and the optimal RL model parameter (γ).

We investigated neural activations from three relevant trial periods: immediately post-presentation of selected choice but prior to reward (0-500 ms selected choice period), during presentation of trial reward (0-500 ms reward period), and during presentation of the cumulative reward up to that trial in the trial sequence (0-500 ms cumulative reward period). *Perf* neural activations were the relative difference in activity on Δpayoff vs. $\Delta_0$payoff block RareG trials. Taking the relative block difference allowed non-task related individual EEG differences to cancel out. Relative responses to the RareG door were important for analysis because this door choice resulted in a larger long-

200 term payoff than the other (RareL) door in the Δpayoff block. We applied family-wise error-
201 rate (fwer) corrections to the *Perf* source-space neural correlates to account for multiple
202 comparisons across three frequency bands (theta, alpha, beta) and three time periods
203 (choice, reward, cumulative reward). **Figure 3A** shows the *Perf* source-space neural
204 correlates found to be significant in this analysis; all neural activations inversely related
205 to *Perf*.

206 We further accounted for the multiple independently significant cortical ROI
207 predictors (Figure 3A) within a unified multivariate model for *Perf* that also included the
208 significant *Bias* and RL model parameter γ as covariates. This overall multivariate model
209 for *Perf* was significant (R^2=0.69, df=126, p<0.0001). The results of this model showed
210 alpha activity in the bilateral rostral anterior cingulate cortex (rACC) during the reward
211 period (left: β=-394.32 ± 125.56, t(177)=-3.14, p=0.002; right: β=-634.89 ± 220.51,
212 t(176)=-2.88, p=0.004) and theta activity in the right rACC during the cumulative reward
213 period (β=-56.98 ± 17.93, t(177)=-3.17, p=0.002) as the most significant predictors of
214 payoff-based decisions; activity in the selected choice presentation period did not survive
215 the multivariate model. We then investigated whether these specific theta/alpha rACC
216 activations also related to gain frequency *Bias*, while controlling for the *Perf* payoff. The
217 bilateral alpha band rACC activations were significantly related to *Bias* (left: β=-
218 514.67±205.05, t(178)=-2.51, p=0.01; right: β=-886.15±348.74, t(179)=-2.54, p=0.01) but
219 not right rACC theta (p>0.05). Thus, right rACC theta activity during the cumulative reward
220 period was the distinct neural correlate of reward payoff **(Figure 3B)**; corresponding scalp
221 theta activity and topography are shown in **Figure 3C**.

Additionally, we checked whether the distinct Perf correlate of right rACC theta activation during the cumulative reward period showed any interactions with age and gender, but no significant interactions were found ($p>0.1$).

Distinct neural correlates of reward maximization in right rACC predict depressed mood.

Finally, we investigated whether the neural correlates of payoff decisions are relevant to subjective mental health by modeling anxiety, depression, inattention and hyperactivity self-reported mood symptoms scores as dependent variables in robust multivariate regression models. All demographic variables (age, gender, race, ethnicity, SES), task performance variables of *Perf* and *Bias*, the relevant RL model parameter γ, as well as the reward-processing neural correlates (bilateral rostral ACC alpha; right rostral ACC theta) were included in each model as independent predictors. Models were fwer-corrected for multiple comparisons.

Only the models for anxiety ($R^2=0.24$, $p=0.0002$) and depression ($R^2=0.21$, $p=0.002$) were significant after multiple comparisons correction. Amongst demographics, age negatively predicted both anxiety ($\beta=-0.02\pm0.004$, $t(152)=-4.95$, $p<0.0001$) and depression ($\beta=-0.01\pm0.004$, $t(152)=-3.42$, $p=0.0007$); no other demographics, RL model parameter γ, Perf or Bias were significant in these models. Notably, neural correlates of payoff performance, specifically right rACC theta during the cumulative reward period positively predicted depressed mood ($\beta=301.41\pm150.34$, $t(152)=2.00$, $p=0.046$, **Figure 3D**); payoff neural correlates did not predict anxiety.

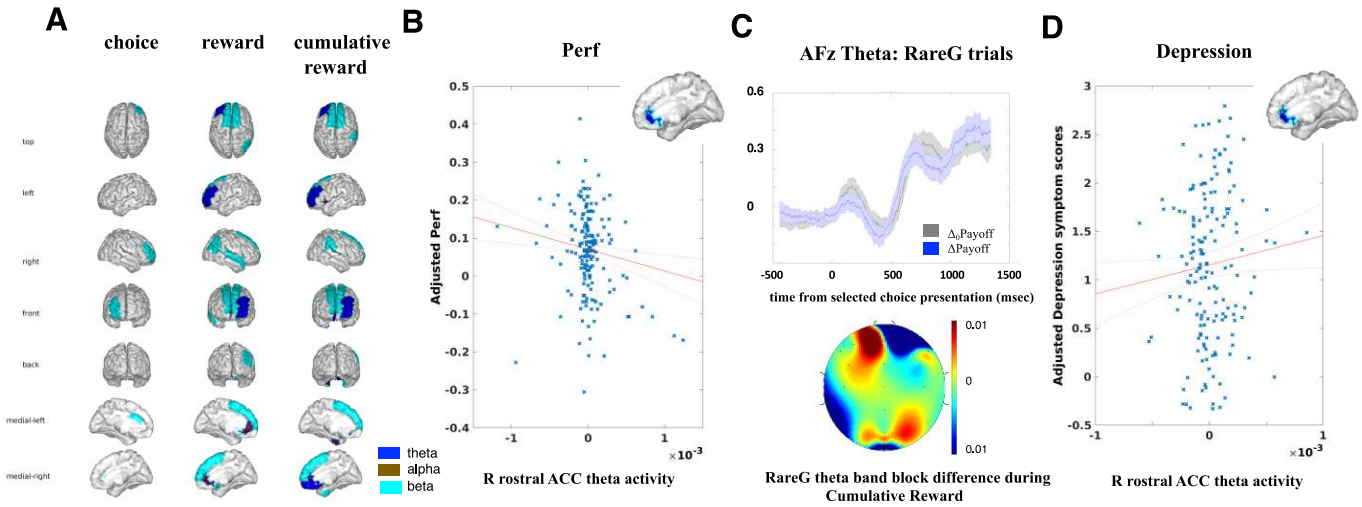

Figure 3: Neural correlates of payoff-based decision making in humans. (A) Independent neural correlates in theta (blue), alpha (brown), and beta (cyan) bands found to be inversely related to *Perf* in the choice selection, reward and cumulative reward epochs ($p<0.0055$). **(B)** Only right rostral anterior cingulate cortex (rACC) theta activity in the cumulative reward period uniquely relates to *Perf* after controlling for *Bias* and time scale of reward prediction. **(C)** Corresponding theta band temporal activity on RareG trials at mid-frontal scalp channel AFz showing differences between payoff blocks (top), and scalp topography of RareG theta band trial difference across Δpayoff and $\Delta_0$payoff blocks in the significant cumulative reward period (bottom). **(D)** *Perf* related rACC theta positively predicts depressed mood symptoms. The scatter plots in (B) and (D) are presented on an adjusted axis as obtained from the multivariate robust regression models; the x-axes for these plots are 10^{-3} source activity units.

Discussion

Reinforcement learning models suggest healthy human choices ideally tend to maximize long-term beneficial outcomes [24,25,27,36,37]. However, many existing neuropsychological measures of decision-making that optimize for long-term payoffs don't reliably estimate the participant's ability to integrate rewards and make foresighted decisions, and instead suffer from biased predispositions [16,20,38,39]; this tendency to choose lesser local minima based gain over long-term gain is also referred to as melioration [18]. In our paradigm, we study gain frequency driven bias separate from payoff-

266 based decision-making by introducing a $\Delta_0$payoff block with no payoff difference between
267 choice options wherein decisions are purely based on gain frequency [15]. Comparing
268 choices within the Δpayoff experimental block, designed to have similar reward
269 distribution structure as the $\Delta_0$payoff block but differing only on the long-term outcome
270 between options, allows measurement of long-term payoff maximization strategy.
271 Therefore, our study design is able to distinguish reward maximization from melioration
272 and further leverages these measures to inform mental health behaviors.

273 The behavioral outcomes of our experiment varied based on individual subject
274 characteristics. We found that payoff-based performance was significantly related to
275 individual bias for observed frequency of gains; this is in line with prior studies of decision-
276 making but wherein gain frequency decisions are often conflated with expected value [1,15].
277 We further modeled subjective differences in reinforcement learning based decision
278 making and by using RL modeling framework, we extracted subjective sensitivity to risks
279 (α) and time scale of reward prediction (γ) to explain each subject's behavior. The
280 parsimonious RL models suggested that the observed behavior is preferably explained
281 by the differences in the extent of reward prediction over time (γ) between individuals.

282 Uniquely, we then investigated the neural correlates of payoff performance while
283 accounting for individual differences in gain frequency bias and extent of reward
284 prediction. Earlier studies have suggested complex dynamics mediated by dopamine and
285 serotonin neuromodulation in the cortico-basal ganglia circuit to underlie reward and risk
286 based decision making [25,26,29,40–44]. In our study, we focused on three different time
287 periods of the task, the first associated with processing of the selected choice, the ensuing
288 reward presentation period and the cumulative reward period to understand how neural

dynamics in these periods affect payoff-based performance. The selected choice period captures the processing associated with presentation of the chosen door after the actual decision period. We did not analyze the actual decision period since two different choice options are shown on the screen during this time and the signal associated with every choice option was difficult to explicitly assess. These analyses showed distinct neural correlates of payoff based performance in the core frontal executive region of the right rACC during the cumulative reward period. Relatedly, most prior studies on probabilistic reward processing have suggested this medial prefrontal cortex (mPFC) region to be critical for mediating decision performance [1–3,30–32].

More specifically, analyses showed theta activity in right rACC negatively associated with payoff-based performance. This finding is aligned with prior evidence for reward-based theta processing and its widely studied relationship to long-term risk or uncertainty [45–47]. This may be one reason why we observe a negative relationship between cumulative reward period rACC theta and effective payoff, *Perf*, because long-term payoff magnitude has been shown to inversely relate to uncertainty but positively to choice utility [48–51]. Notably, reward period bilateral rACC alpha also related to payoff maximization, but this was not a distinct correlate of payoff as it also significantly explained gain frequency bias or melioration in our task.

Translational neuroscience studies show that reward based decision processing deficits are found in depression and in attention disorders, leading to difficulty in reward integration and foresighted choice-behaviors [24,41,52–56]. Such individuals then focus on the immediate reward outcome in the short-term, characterized by a prolonged attenuation of temporal discounting of rewards [57,58]. Interestingly, rACC theta activity in our paradigm

during the cumulative reward period, which negatively correlated with payoff performance, was a positive predictor of depressed mood. This result is aligned with prior research using fMRI showing that the volume and activation patterns of rACC correlates with depressed mood [59,60]. Related research also suggests that rACC theta activity is a significant baseline marker for depression treatment outcomes [61]. These results are still limited by our study in healthy adults and need to be replicated in clinical populations, and with greater channel density electrophysiology recordings.

Altogether, our study presents the importance of controlling for melioration biases for immediate reward frequency and individual differences in learning while assessing advantageous, i.e. foresighted, decision-making ability in humans. Our findings of payoff-relevant right rostral ACC theta activity in the cumulative reward feedback period, could be important for clinical translational application, particularly for depression, suggesting a plausible neural target for interventions that engage reward processing.

Methods

Participants. 198 adult human subjects (age mean ± standard deviation 35.44 ± 20.30 years, range 18-80 years, 115 females) participated in the study. The study was performed in accordance with the relevant guidelines and regulations approved by the University of California San Diego institutional review board (UCSD IRB #180140). All participants provided written informed consent for the study protocol approved by the University of California San Diego institutional review board. Twelve of these participants were excluded from the study as they had a current diagnosis for a psychiatric disorder and current/recent history of psychotropic medications for a final sample of 186 healthy adult participants. All participants reported normal/corrected-to-normal vision and hearing and no participant reported color blindness. For older adults >60 years of age, participants were confirmed to have a Mini-Mental State Examination (MMSE) score >26 to verify absence of apparent cognitive impairment (Arevalo-Rodriguez et al. 2015). All data was collected prior to the COVID-19 period of restricted human subjects research.

Surveys. All participants provided demographic information by self-report including age, gender, race (in a scale of 1 to 7: Caucasian; Black/African American; Native Hawaiian / Other Pacific Islander; Asian; American Indian / Alaska Native; More than one race; Unknown or not reported) and ethnicity; socio-economic status (SES) was measured on the Family Affluence Scale from 1 to 9 (Boudreau and Poulin, 2008), and any current/past history of clinical diagnoses and medications were reported. For older adults >60 years of age, participants completed the Mini-Mental State Examination (MMSE) and scored >26 to verify absence of apparent cognitive impairment [62]. All participants completed subjective mental health self-reports using standard instruments, ratings of inattention and hyperactivity obtained on the Adult ADHD Rating Scale (New York University and Massachusetts General Hospital. Adult ADHD-RS-IV with Adult Prompts. 2003; : 9–10), Generalized Anxiety Disorder 7-item scale GAD-7 [63] and depression symptoms reported on the 9-item Patient Health Questionnaire, PHQ-9 [64]. Symptoms for these psychiatric conditions were measured because they have been related to changes in reward processing [65–67].

Task Design. We investigated a two-choice decision-making task that enabled a rapid assessment and was easy to understand across the adult lifespan. In this task that we refer to as *Lucky Door*, participants chose between one of two doors, either a rare gain door (RareG, probability for gains p=0.3, for losses p=0.7) or a rare loss door (RareL, probability for losses p=0.3, for gains p=0.7). Participants used the left and right arrow keys on the keyboard to make their door choice. Door choice was monitored throughout the task. In two separate blocks, we investigated whether the overall expected value (payoff) of the choice door can influence individual behavior. In the baseline block with $\Delta_0$payoff (no-payoff difference), the two choice doors did not differ in payoff (RareL door, p=0.3 for -70 coins and p=0.7 for +30 coins, payoff=0; RareG door, p=0.3 for +70 coins and p=0.7 for -30 coins, payoff=0). In the experimental difference block with Δpayoff (payoff difference), expected value or payoff was greater for the RareG door (p=0.3 for +60 coins, p=0.7 for -20 coins, payoff=+40) than for the RareL door (p=0.3 for -60 coins, p=0.7 for +20 coins; payoff=-40). Manipulation of payoff, with greater expected value tied

to the RareG door, allowed for investigating individual propensities to prioritize long-term (or cumulative) vs. short-term (or immediate) rewards. The RareG door was assigned greater payoff because selecting this door suggests payoff-based decision processing in subjects as opposed to simply choosing based on frequency of gains, in which case the RareL choice should be preferred. 40 trials were presented per block and block order was randomized across participants; two practice trials preceded the main Δpayoff or $\Delta_0$payoff blocks. **Figure 1A** shows a schematic of the task stimulus sequence and **Supplementary table 1** shows the reward distribution that was shuffled and updated after every 10 trials had been sampled from that set.

The *Lucky Door* task was deployed in Unity as part of the assessment suite on the *BrainE* (short for Brain Engagement) platform [68]. The Lab Streaming Layer (LSL [69]) protocol was used to time-stamp each stimulus/response event during the task. Study participants engaged with the assessment on a Windows 10 laptop sitting at a comfortable viewing distance.

Electroencephalography (EEG). EEG data was collected simultaneous to the *Lucky Door* task using a 24-channel Smarting device with a semi-dry and wireless electrode layout (Next EEG—new human interface, MBT). Data were acquired at 500 Hz sampling frequency at 24-bit resolution. Cognitive event markers were integrated using LSL and data files were stored in xdf format.

Behavioral analyses. Task speeds were calculated as log(1/RT), where RT is response time in seconds. We computed the payoff sensitive performance response (*Perf*) as the difference in proportion selection of the RareG door between the Δpayoff and the $\Delta_0$payoff blocks; RareG vs. RareL EVs differed only in the Δpayoff block. We computed gain frequency bias (*Bias*) as the difference in selected proportion of RareL and RareG doors in the $\Delta_0$payoff block where the payoff for both the doors was the same. While *Perf* is indicative of subjective payoff maximization based selection of advantageous choices, *Bias* informs the choice bias and melioration for higher immediate gain frequencies. For N fraction of responses in each block, we calculated:

$$\begin{aligned} Perf &= N_{exptRareG} \quad N_{baseRareG} \\ Bias &= N_{baseRareL} \quad N_{baseRareG} \end{aligned} \tag{1}$$

We also calculated Win-Stay and Lose-Shift performance on both Δpayoff and $\Delta_0$payoff blocks. Win-Stay was computed as the proportion of times the subject repeated the same choice option in the next trial after obtaining a gain for choosing that option in the current trial. Lose-Shift was computed as the proportion of times the subject would shift away from the current choice option in the next trial, on obtaining a loss in the current trial.

Reinforcement Learning (RL) Model. We simulated three parsimonious RL models with up to two free parameters, which have a key role in reward-based decision-making, to explain the observed individual subject behavior; see **Figure 2A** for model schematic [25,40]:
1) Model A optimized the time scale of reward prediction (γ) for every subject;
2) Model B optimized risk sensitivity (α) for every subject; and

3) Model C optimized both γ and α to for every subject.
The time scale of reward prediction parameter (γ) represents whether reward prediction is myopic or long-sighted, lower values $\gamma \in (0\ 1)$, $\gamma \rightarrow 0$ suggest myopic reward prediction leading to impulsive decisions while higher values, $\gamma \rightarrow 1$ suggest long-sighted integration of rewards for decisions.
The risk sensitivity parameter (α) measures the extent to which expected uncertainty associated with the door influences the decision utility, the smaller the parameter value $\alpha \in (-1\ 1)$, $\alpha \rightarrow -1$ the higher is risk seeking, while a larger value, $\alpha \rightarrow 1$ indicates high risk aversiveness.

The simulation agent had reward distributions as in the real experiment but scaled down by multiplying with a parameter 0.1, and varying with blocks (Δpayoff, $\Delta_0$payoff) that were randomly ordered. There were as high as 50,000 trials in each block for letting model performance converge.

The agent has to choose between two doors each of which (stimulus, *s*) was represented by a radial basis function (Φi) as below:

$$s = \exp\frac{(x \quad _s)^2}{2} \tag{2}$$

Here, the μ_s and σ denotes the mean ($s \in [1\ 2]$; door1 = 1; door2 = 2) and standard deviation of the inverse attention parameter, respectively. σ is set to 1 in our models.

The door stimulus is multiplied with the weight matrix *wv* for computing its value function, *Q*, and *wr* for constructing its risk function, $\sqrt{h}$.

Utility associated with any state at a trial, t, is the combination of value and risk function [70,71], where the risk function is modulated by a risk sensitivity parameter α. Higher the *α*, higher the risk aversiveness of the subject.

$$U(s,t) = Q(s,t) \qquad \sqrt{h(s,t)}$$

Where

$$Q(s,t) = wv(s,t) \quad (s)$$

$$h(s,t) = wr(s,t) \quad (s) \tag{3}$$

The door choice selection is performed using the SoftMax principle defined as below. According to SoftMax, the probability for choosing a door at trial, t, is P(s,t):

$$P(s,t) = \frac{exp(\ \ U(s,t))}{\sum_{i=1}^{n} exp(\ \ U(i,t))} \tag{4}$$

Here, n is the total number of doors available, and β is the exploration index. β is set to 1 in our models.
After choice selection, the weight functions are updated using the below principles. The choice value function Q at trial t+1 for door, s, may be expressed as,

$$Q(s,t+1) = Q(s,t) + \ _Q \quad (s) \tag{5}$$

where η_Q is the learning rate of the value function ($0 < \eta Q < 1$) for the stimulus variable, Φ(s). δ is the temporal difference error represented as

$$= r + \max_{s'} Q(s',t) \quad Q(s,t) \tag{6}$$

where r is the reward associated with taking an action, a, for stimulus, s, at time, t, and γ is the time scale of reward prediction. Similar to the value function, the risk function h has an incremental update as defined by the below equation. Optimizing the risk function in addition to the value function is shown to capture human behavior well in a variety of cognitive tasks involving rewards, punishments and risk [25,33].

$$h(s,t+1) = h(s,t) + {}_{h} \quad (s) \tag{7}$$

where η_h is the learning rate of the risk function ($0 < \eta h < 1$), and ξ is the risk prediction error expressed by the below equation.

$$= {}^{2} \quad h(s,t) \tag{8}$$

For simplicity, we model as $\eta_h = \eta_Q = 0.1$ as an initial optimization for our subjects for η provided a median of 0.1. The weights *wv* and *wr* are set to a small random number from set [-0.0005 0.0005] at trial = 1. The weights are normalized by dividing by their norm.

The cost function optimizes the frequency of selections of rare gain and rare loss options in $\Delta_0$payoff and Δpayoff blocks for every subject after running the simulation agent for 10 instances of 100,000 each, and inferring the optimal parameters for every participant in our study using *fmincon* function in MATLAB. Cost function = sum of squares of the difference for observed actual (Proportion# $\text{RareG}_{\text{expt}}$ + Proportion# $\text{RareL}_{\text{expt}}$ + Proportion# $\text{RareG}_{\text{base}}$ + Proportion# $\text{RareL}_{\text{base}}$) – simulated actual (Proportion# $\text{RareG}_{\text{expt}}$ + Proportion# $\text{RareL}_{\text{expt}}$ + Proportion# $\text{RareG}_{\text{base}}$ + Proportion# $\text{RareL}_{\text{base}}$). Optimization is carried out for either one (Models A,B) or two (Model C) parameters, using *fmincon*(). We ran fmincon() 100 times to choose the parameter set with least cost for any subject.

The AIC (Akaike Information Criteria) for these models were built using the likelihood function, which was estimated as the average correlation coefficient between the simulated and the observed key task behaviors - 1) payoff performance or *Perf* measure (eqn. 1); 2) gain frequency *Bias* measure (eqn. 1); and 3) the block differences between Δpayoff and Δ0payoff blocks in Win-Stay results for the RareG door.

Neural data processing. We applied a uniform processing pipeline to all EEG data acquired simultaneous to the reward task. This included: 1) data pre-processing, 2) computing event related spectral perturbations (ERSP) for all channels, and 3) cortical source localization of the EEG data filtered within relevant theta, alpha and beta frequency bands.

1) Data preprocessing was conducted using the EEGLAB toolbox in MATLAB[72]. EEG data was resampled at 250 Hz, and filtered in the 1-45 Hz range to exclude ultraslow DC drifts at <1Hz and high-frequency noise produced by muscle movements and external electrical sources at >45Hz. EEG data were average referenced and epoched to the chosen door presentation during the task, in the -.5 sec to +1.5 sec time window (**Figure 1**). Any missing channel data (one channel each in 6 participants) was spherically interpolated to nearest neighbors. Epoched data were cleaned using the autorej function in EEGLAB to remove noisy trials (>5sd outliers rejected over max 8 iterations; 0.91±

2.65% of trials rejected per participant). EEG data were further cleaned by excluding signals estimated to be originating from non-brain sources, such as electrooculographic, electromyographic or unknown sources, using the Sparse Bayesian learning (SBL) algorithm[73,74], https://github.com/aojeda/PEB) explained below in the cortical source localization section.

2) For ERSP calculations, we performed time-frequency decomposition of the epoched data using the continuous wavelet transform (cwt) function in MATLAB's signal processing toolbox. Baseline time-frequency (TF) data in the -250 ms to -50 ms time window prior to chosen door presentation were subtracted from the epoched trials (at each frequency) to observe the event-related synchronization (ERS) and event-related desynchronization (ERD) modulations [75].

3) Cortical source localization was performed to map the underlying neural source activations for the ERSPs using the block-Sparse Bayesian learning (SBL) algorithm [73,74] implemented in a recursive fashion. This is a two-step algorithm in which the first-step is equivalent to low-resolution electromagnetic tomography (LORETA) [76]. LORETA estimates sources subject to smoothness constraints, i.e. nearby sources tend to be co-activated, which may produce source estimates with a high number of false positives that are not biologically plausible. To guard against this, SBL applies sparsity constraints in the second step wherein blocks of irrelevant sources are pruned. This data-driven sparsity constraint of the SBL method reduces the effective number of sources considered at any given time as a solution, thereby reducing the ill-posed nature of the inverse mapping. In other words, one can either increase the number of channels used to solve the ill-posed inverse problem or impose more aggressive constraints on the solution to converge on the source model. The two-stage SBL produces evidence-optimized inverse source models at 0.95AUC relative to the ground truth while without the second stage <0.9AUC is obtained [73,74]. Prior research has also shown that sparse source imaging constraints can be soundly applied to low channel density data [77,78]. Furthermore in Balasubramani et al. (42) we have shown that the ROI estimates resulting from this cortical source mapping have high test-retest reliability (Cronbach's alpha = 0.77, $p<0.0001$).

Source space activity signals were estimated and their root mean squares were partitioned into (1) regions of interest (ROIs) based on the standard 68 brain region Desikan-Killiany atlas [35], using the Colin-27 head model [79] and (2) artifact sources contributing to EEG noise from non-brain sources such as electrooculographic, electromyographic or unknown sources; activations from non-brain sources were removed to clean the EEG data. Cleaned subject-wise trial-averaged EEG data were then specifically filtered in theta (3-7 Hz), alpha (8-12 Hz), and beta (13-30 Hz) bands and separately source localized in each of these bands to estimate their cortical ROI source signals. The source signal envelopes were computed in MatLab (envelop function) by a spline interpolation over the local maxima separated by at least one time sample; we used this *spectral amplitude* signal for all neural analyses presented here. For analyses, we focused on theta, alpha and beta signals in the relevant selected choice period (0-500 ms after selected choice presentation), trial reward period (500-

1000 ms after selected choice presentation), and the cumulative reward period (1000-1500 ms after selected choice presentation).

Statistical Analyses. We fit robust multivariate linear regression models in MATLAB to investigate the behavioral relationships between the *Perf* measure and demographic variables (age, sex, race, ethnicity and SES), while controlling for *Bias* and order of block presentation. For any linear regression model, the response variable was log-transformed for normality and we identified significant factors contributing to the main effects. For regression models, we report the overall model R^2 and p-value, and individual variable β coefficients, t-statistic, degrees of freedom, and p-values.

Channel-wise theta, alpha, beta ERS and ERD modulations for significant spectral activity were computed relative to baseline by first processing for any outliers; any activations greater than 5MAD from the median were removed from further analyses. The significant average activity across all trials were found by performing t-tests ($p<0.05$) across subjects, followed by false discovery rate (FDR, alpha = 0.05) corrections applied across the three dimensions of time, frequency, and channels [80].

For computing source level activity correlates of the behavioral *Perf* measure, we first found the difference in RareG door specific neural activations between Δpayoff and $\Delta_0$payoff blocks in three frequency bands – theta, alpha and beta and in three relevant trial periods: selected choice presentation, trial reward and cumulative reward period. We again used robust linear regression fits for identifying individual ROIs that relate to the *Perf* measure, while also accounting for individual *Bias* and variations in RL as predicted by the best fitting RL model. The results were family-wise error rate corrected for multiple comparisons for 3 trial periods and 3 frequency bands (FWER correction, $p<0.0055$). The independently identified ROIs were further factored in a unified multivariate linear regression model to account for comparisons across ROIs; significant ROIs in this final multivariate model were reported after controlling for *Bias*, and the RL model-fit parameter ($p<0.05$).

Finally, we used robust multivariate linear regressions to model the self-reported mood symptoms of Anxiety, Depression, Inattention, Hyperactivity using predictors of demographic variables, *Perf*, *Bias*, RL model-fit parameter along with the identified neural correlates of *Perf*. Adjusted responses from robust multivariate models were plotted using the plotAdjustedResponse function in MATLAB.

References

1. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Deciding advantageously before knowing the advantageous strategy. *Science* **275**, 1293–1295 (1997).
2. Brevers, D., Bechara, A., Cleeremans, A. & Noel, X. Iowa Gambling Task (IGT): twenty years after – gambling disorder and IGT. *Front. Psychol.* **4**, (2013).
3. Bechara, A., Tranel, D. & Damasio, H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain* **123**, 2189–2202 (2000).
4. Bembich, S. *et al.* Differences in time course activation of dorsolateral prefrontal cortex associated with low or high risk choices in a gambling task. *Frontiers in human neuroscience* **8**, 464 (2014).
5. Christakou, A., Brammer, M., Giampietro, V. & Rubia, K. Right ventromedial and dorsolateral prefrontal cortices mediate adaptive decisions under ambiguity by integrating choice utility and outcome evaluation. *Journal of Neuroscience* **29**, 11020–11028 (2009).
6. Garon, N., Moore, C. & Waschbusch, D. A. Decision Making in Children With ADHD Only, ADHD-Anxious/Depressed, and Control Children Using a Child Version of the Iowa Gambling Task. *J Atten Disord* **9**, 607–619 (2006).
7. Gupta, R., Koscik, T. R., Bechara, A. & Tranel, D. The amygdala and decision-making. *Neuropsychologia* **49**, 760–766 (2011).
8. Mueller, E. M., Nguyen, J., Ray, W. J. & Borkovec, T. D. Future-oriented decision-making in Generalized Anxiety Disorder is evident across different versions of the Iowa Gambling Task. *Journal of Behavior Therapy and Experimental Psychiatry* **41**, 165–171 (2010).
9. Must, A., Horvath, S., Nemeth, V. L. & Janka, Z. The Iowa GamblingTask in depression – what have we learned about sub-optimal decision-making strategies? *Frontiers in psychology* **4**, 1–6 (2013).
10. Oberg, S. A., Christie, G. J. & Tata, M. S. Problem gamblers exhibit reward hypersensitivity in medial frontal cortex during gambling. *Neuropsychologia* **49**, 3768–3775 (2011).
11. Roca, M. *et al.* Executive Functions in Pathologic Gamblers Selected in an Ecologic Setting. *Cognitive and Behavioral Neurology* **21**, 1–4 (2008).
12. Woodrow, A. *et al.* Decision-making ability in psychosis: a systematic review and meta-analysis of the magnitude, specificity and correlates of impaired performance on the Iowa and Cambridge Gambling Tasks. *Psychological Medicine* **49**, 32–48 (2019).
13. Chiu, Y.-C. & Lin, C.-H. Is deck C an advantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions* **3**, 37 (2007).
14. Chiu, Y.-C. *et al.* Immediate gain is long-term loss: Are there foresighted decision makers in the Iowa Gambling Task? *Behavioral and Brain Functions* **4**, 13 (2008).
15. Lin, C., Chiu, Y. & Huang, J. Gain-loss frequency and final outcome in the Soochow Gambling Task: A Reassessment. *Behavioral and Brain Functions* **5**, 1–9 (2009).
16. Napoli, A. & Fum, D. Rewards and punishments in iterated decision making: An explanation for the frequency of the contingent event effect. in *10th International Conference on Cognitive Modeling. Philadelphia, PA* (Citeseer, 2010).

17. Singh, V. & Khan, A. Decision making in the reward and punishment variants of the Iowa gambling task: evidence of “foresight” or “framing”? *Frontiers in neuroscience* **6**, 107 (2012).
18. Sims, C. R., Neth, H., Jacobs, R. A. & Gray, W. D. Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review* **120**, 139–154 (2013).
19. Franken, I. H. & Muris, P. Individual differences in decision-making. *Personality and Individual Differences* **39**, 991–998 (2005).
20. Furl, B. A. The influence of individual differences on the Iowa Gambling Task and real-world decision making. (Wake Forest University, 2010).
21. Harman, J. L. Individual differences in need for cognition and decision making in the Iowa Gambling Task. *Personality and individual differences* **51**, 112–116 (2011).
22. Newman, L. I., Polk, T. A. & Preston, S. D. Revealing individual differences in the Iowa Gambling Task. in 1067–1072 (Cognitive Science Society Austin, TX, 2008).
23. Weller, J. A., Levin, I. P. & Bechara, A. Do individual differences in Iowa Gambling Task performance predict adaptive decision making for risky gains and losses? *Journal of Clinical and Experimental Neuropsychology* **32**, 141–150 (2010).
24. Balasubramani, P. P. & Chakravarthy, V. S. Bipolar oscillations between positive and negative mood states in a computational model of Basal Ganglia. *Cogn Neurodyn* (2019) doi:10.1007/s11571-019-09564-7.
25. Balasubramani, P. P., Chakravarthy, S., Ravindran, B. & Moustafa, A. A. 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. *Frontiers in Computational Neuroscience* **8**, 47 (2014).
26. Chakravarthy, V. S., Balasubramani, P. P., Mandali, A., Jahanshahi, M. & Moustafa, A. A. The many facets of dopamine: Toward an integrative theory of the role of dopamine in managing the body’s energy resources. *Physiology & behavior* (2018).
27. Gupta, A., Balasubramani, P. P. & Chakravarthy, S. Computational model of precision grip in Parkinson’s disease: A Utility based approach. *Frontiers in Computational Neuroscience* **7**, (2013).
28. Muralidharan, V., Balasubramani, P. P., Chakravarthy, V. S., Lewis, S. J. & Moustafa, A. A. A computational model of altered gait patterns in parkinson’s disease patients negotiating narrow doorways. *Frontiers in computational neuroscience* **7**, 190 (2014).
29. Doya, K. Metalearning and neuromodulation. *Neural networks* **15**, 495–506 (2002).
30. Balasubramani, P. P. & Hayden, B. Overlapping neural processes for stopping and economic choice in orbitofrontal cortex. *bioRxiv* 304709 (2018).
31. Kennerley, S. W., Behrens, T. E. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature neuroscience* **14**, 1581 (2011).
32. Moccia, L. *et al.* Neural correlates of cognitive control in gambling disorder: a systematic review of fMRI studies. *Neuroscience and Biobehavioral Reviews* **78**, 104–116 (2017).
33. Balasubramani, P. P., Chakravarthy, S., Ravindran, B. & Moustafa, A. A. A network model of basal ganglia for understanding the roles of dopamine and serotonin in

reward-punishment-risk based decision making. *Name: Frontiers in Computational Neuroscience* **9**, 76 (2015).
34. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019).
35. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
36. Balasubramani, P. P., Chakravarthy, V. S., Ali, M., Ravindran, B. & Moustafa, A. A. Identifying the Basal Ganglia Network Model Markers for Medication-Induced Impulsivity in Parkinson's Disease Patients. *PLoS One* **10**, e0127542 (2015).
37. Balasubramani, P. P., Moreno-Bote, R. & Hayden, B. Y. Using a Simple Neural Network to Delineate Some Principles of Distributed Economic Choice. *Frontiers in Computational Neuroscience* **12**, 22 (2018).
38. Gansler, D. A., Jerram, M. W., Vannorsdall, T. D. & Schretlen, D. J. Comparing alternative metrics to assess performance on the Iowa Gambling Task. *Journal of Clinical and Experimental Neuropsychology* **33**, 1040–1048 (2011).
39. Stocco, A., Fum, D. & Napoli, A. Dissociable processes underlying decisions in the Iowa Gambling Task: a new integrative framework. *Behavioral and Brain Functions* **5**, 1–12 (2009).
40. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction. Adaptive Computations and Machine Learning*. (MIT Press/Bradford, 1998).
41. Balasubramani, P. P., Chakravarthy, V. S., Ali, M., Ravindran, B. & Moustafa, A. A. Identifying the Basal Ganglia Network Model Markers for Medication-Induced Impulsivity in Parkinson's Disease Patients. *PLoS One* **10**, (2015).
42. Balasubramani, P. P., Chakravarthy, V. S., Ravindran, B. & Moustafa, A. A. Modeling Serotonin's Contributions to Basal Ganglia Dynamics. in *Computational Neuroscience Models of the Basal Ganglia* 215–243 (Springer, 2018).
43. Schultz, W. Multiple reward signals in the brain. *Nature reviews. Neuroscience* **1**, 199 (2000).
44. Tobler, P. N., Christopoulos, G. I., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Risk-dependent reward value signal in human prefrontal cortex. *Proceedings of the National Academy of Sciences* **106**, 7185–7190 (2009).
45. Cavanagh, J. F., Frank, M. J., Klein, T. J. & Allen, J. J. Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage* **49**, 3198–3209 (2010).
46. Christie, G. J. & Tata, M. S. Right frontal cortex generates reward-related theta-band oscillatory activity. *Neuroimage* **48**, 415–422 (2009).
47. Zavala, B. *et al.* Cognitive control involves theta power within trials and beta power across trials in the prefrontal-subthalamic network. *Brain* **141**, 3361–3376 (2018).
48. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10**, 1214–1221 (2007).
49. Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X. & Milham, M. P. Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. *NeuroImage* **32**, 477–484 (2006).

50. Paulus, M. P. & Frank, L. R. Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage* **30**, 668–677 (2006).
51. Preuschoff, K., Bossaerts, P. & Quartz, S. R. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* **51**, 381–90 (2006).
52. Gradin, V. B. *et al.* Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* **134**, 1751–1764 (2011).
53. Groen, Y., Gaastra, G. F., Lewis-Evans, B. & Tucha, O. Risky behavior in gambling tasks in individuals with ADHD–a systematic literature review. *PLoS One* **8**, e74909 (2013).
54. Miller, W. R. & Seligman, M. E. Depression and the perception of reinforcement. *Journal of Abnormal Psychology* **82**, 62–73 (1973).
55. Silvetti, M., Wiersema, J. R., Sonuga-Barke, E. & Verguts, T. Deficient reinforcement learning in medial frontal cortex as a model of dopamine-related motivational deficits in ADHD. *Neural networks* **46**, 199–209 (2013).
56. Ziegler, S., Pedersen, M. L., Mowinckel, A. M. & Biele, G. Modelling ADHD: A review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neuroscience & Biobehavioral Reviews* **71**, 633–656 (2016).
57. Eshel, N. & Roiser, J. P. Reward and punishment processing in depression. *Biological psychiatry* **68**, 118–124 (2010).
58. Pizzagalli, D. A., Sherwood, R. J., Henriques, J. B. & Davidson, R. J. Frontal brain asymmetry and reward responsiveness: a source-localization study. *Psychological Science* **16**, 805–813 (2005).
59. Boes, A. D., McCormick, L. M., Coryell, W. H. & Nopoulos, P. Rostral Anterior Cingulate Cortex Volume Correlates with Depressed Mood in Normal Healthy Children. *Biological Psychiatry* **63**, 391–397 (2008).
60. Yoshimura, S. *et al.* Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *Journal of Affective Disorders* **122**, 76–85 (2010).
61. Pizzagalli, D. A. *et al.* Pretreatment Rostral Anterior Cingulate Cortex Theta Activity in Relation to Symptom Improvement in Depression: A Randomized Clinical Trial. *JAMA Psychiatry* **75**, 547–554 (2018).
62. Arevalo- Rodriguez, I. *et al.* Mini- Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* (2015).
63. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Loewe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of internal medicine* **166**, 1092–1097 (2006).
64. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of general internal medicine* **16**, 606–613 (2001).
65. Admon, R. & Pizzagalli, D. A. Dysfunctional reward processing in depression. *Current Opinion in Psychology* **4**, 114–118 (2015).
66. Dillon, D. G. *et al.* Peril and pleasure: An RDOC- inspired examination of threat responses and reward processing in anxiety and depression. *Depression and anxiety* **31**, 233–249 (2014).

67. Luman, M., Tripp, G. & Scheres, A. Identifying the neurobiology of altered reinforcement sensitivity in ADHD: A review and research agenda. *Neuroscience & Biobehavioral Reviews* **34**, 744–754 (2010).
68. Balasubramani, P. P. *et al.* Mapping Cognitive Brain Functions at Scale. *NeuroImage* 117641 (2020).
69. Kothe, C., Medine, D., Boulay, C., Grivich, M. & Stenner, T. *'Lab Streaming Layer' Copyright*. (2019).
70. Bell, D. E. Risk, return, and utility. *Management science* **41**, 23–30 (1995).
71. d'Acremont, M., Lu, Z.-L., Li, X., Van der Linden, M. & Bechara, A. Neural correlates of risk prediction error during reinforcement learning in humans. *Neuroimage* **47**, 1929–1939 (2009).
72. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9–21 (2004).
73. Ojeda, A., Kreutz-Delgado, K. & Mullen, T. Fast and robust Block-Sparse Bayesian learning for EEG source imaging. *Neuroimage* **174**, 449–462 (2018).
74. Ojeda, A., Kreutz-Delgado, K. & Mishra, J. Bridging M/EEG Source Imaging and Independent Component Analysis Frameworks Using Biologically Inspired Sparsity Priors. *Neural Computation* **33**, 1–31 (2021).
75. Pfurtscheller, G. EEG event - related desynchronization (ERD) and event - releated synchronization (ERS). *Electroencephalography: Basic Principles, Clinical Applications and Releated Fields* 958–967 (1999).
76. Pascual-Marqui, R. D., Michel, C. M. & Lehmann, D. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* **18**, 49–65 (1994).
77. Ding, L. & He, B. Sparse source imaging in electroencephalography with accurate field modeling. *Human brain mapping* **29**, 1053–1067 (2008).
78. Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K. & Hansen, L. K. The smartphone brain scanner: a portable real-time neuroimaging system. *PloS one* **9**, e86733 (2014).
79. Holmes, C. J. *et al.* Enhancement of MR Images Using Registration for Signal Averaging. *Journal of Computer Assisted Tomography* **22**, 324–333 (1998).
80. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage* **15**, 870–878 (2002).

Acknowledgements

This work was supported by University of California San Diego (UCSD) lab start-up funds (DR, JM), the Interdisciplinary Research Fellowship in NeuroAIDS (PB: R25MH081482), the Brain & Behavior Research Fund (PB), the Kavli Foundation (PB, JM), and the Sanford Institute for Empathy and Compassion (JM, PB). We thank Alankar Misra for software development of the *BrainE* software and several UCSD undergraduate students who assisted with data collection. The *BrainE* software is copyrighted for commercial use (Regents of the University of California Copyright #SD2018-816) and free for research and educational purposes. We thank Sabyasachi Shivkumar and Vignesh Muralidharan for their helpful feedback on the study analysis.

Author contributions

PB contributed to conceptualization, task design, data collection, analysis, writing of the manuscript; JD contributed to data analysis; GG,MZ,FA,VM contributed to data collection; DR, JM contributed to conceptualization, task design, writing of the manuscript.

Data Availability

A part of the dataset with 96 of the 186 participants data used in this study is available on the open-access repository link: 10.5281/zenodo.4088951

Conflict of Interest

The authors declare no conflict of interest.

Figure Legends

Figure 1. Reward task and associated behavior A) As per the task schematic, participants fixated for 0.5 sec, then chose from one of two choice doors. Post-response, fixation reappeared for 0.5 sec, followed by presentation of the chosen door for 0.5 sec, then immediate gain or loss feedback provided for 0.5 sec, and finally, cumulative feedback of all gains/losses up to the present trial shown for 0.5 sec. Reward distributions for the door choices are presented in Supplementary Table 1. **B)** Win-Stay behavior for the rare gain RareG door is significantly greater on the ΔPayoff versus Δ_0Payoff block. *:p<.05. **C)** Gain frequency *Bias* significantly predicts payoff performance, *Perf* (r=0.65, p<0.0001).

Figure 2. Reinforcement Learning Model (A) Schematic representing the stimulus, value function and choice selection modules. The model results for number of selections associated with each of the choice door stimuli in each task block are compared against the actual selections made by each subject, for purposes of model optimization. The model uses the utility, U, associated with each choice response for making the decision, where the utility is a function of reward average and reward variance associated with choices. The decision in the model is taken using the SoftMax probability, P, of making the choices. Model parameters are highlighted as α (model agent's differential risk sensitivity to gain and loss outcome uncertainties), and γ (time scale of reward prediction). **(B)** AIC values for three models, Model A (γ optimized), B (α optimized), C (γ and α optimized), show that the γ model was best performing. **(C)** The γ optimized model showed strong correlations between simulated and actual *Perf* (ρ(185)=0.81, p<0.0001), *Bias* (ρ(185)=0.89, p<0.0001) and RareG Win-Stay difference between blocks (ρ(185)=0.17 p=0.004).

Figure 3: Neural correlates of payoff-based decision making in humans. (A) Independent neural correlates in theta (blue), alpha (brown), and beta (cyan) bands found to be inversely related to *Perf* in the choice selection, reward and cumulative reward epochs (p<0.0055). **(B)** Only right rostral anterior cingulate cortex (rACC) theta activity in the cumulative reward period uniquely relates to *Perf* after controlling for *Bias* and time scale of reward prediction. **(C)** Corresponding theta band temporal activity on RareG trials at mid-frontal scalp channel AFz showing differences between payoff blocks (top), and scalp topography of RareG theta band trial difference across Δpayoff and $\Delta_0$payoff blocks in the significant cumulative reward period (bottom). **(D)** *Perf* related rACC theta positively predicts depressed mood symptoms. The scatter plots in (B) and (D) are presented on an adjusted axis as obtained from the multivariate robust regression models; the x-axes for these plots are 10^{-3} source activity units.

Table Legends

873
874 **Table 1. Subject characteristics.** Median ± MAD for subjects demographics variables,
875 mental health self-report scores, and parameter γ from the preferred reinforcement
876 learning model. MAD: median absolute deviation, SES: socioeconomic status score.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BalasubramaniLuckyDoorSupplementSciReports.docx