Model-based analysis of a novel variant of the balloon analogue risk task improves sensitivity to risk-taking behavior.

Adam W. Fenton
Richfield, OH

BS, Psychology, The Ohio State University, 2015

A Predissertation Work presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Master of Arts

Department of Psychology
University of Virginia
May, 2021

Readers:

Dr. Per B. Sederberg

Dr. James P. Morris

ABSTRACT: The assessment and quantification of risk-taking behaviors is critical to parsing out individual differences in cognitive performance. The Balloon Analogue Risk Task (BART) is frequently implemented for differentiating risk-taking behaviors between different populations. We developed a new version of the BART that is able to differentiate decision makers' (DM) behavior within a single population. Past BART studies typically utilize summary statistics that disregard the experimental context at the choice-by-choice level, grossly oversimplifying risk-taking behavior. To address this, we present a new methodology and metric for quantifying risk-taking behaviors by comparing a DM's sequence of decisions to those of an optimal performer with knowledge of the popping probability of the balloons. We then propose and assess several possible cognitive mechanisms that may underlie the decision-making process within a generative modeling framework. This culminates in the Predictive Linear Utility Model (PLUM), which accurately predicts observed behavior. Finally, we compare a version of PLUM fit only to choices to a version fit to choices and reaction times in order to assess the utility of reaction times within this task environment.

1. INTRODUCTION

Decision-making is affected by a myriad of internal and external factors. An aspect of decision-making that is quite susceptible to these factors is risk-taking behavior. It is well understood that some actions - such as not getting enough sleep or consuming a mind-altering substance - can have detrimental effects on risk-taking behavior (Acheson et al., 2007; Park et al., 2020). However, risk-taking behaviors need to be examined in the context in which the behaviors are observed. Risk-taking behaviors are often assessed via self-report and/or questionnaires, leaving a need for an experimental analogue of risk-taking (Lejuez et al., 2002). Although people's actions differ greatly when the risk of failure is known versus when it is ambiguous, we aim to examine risk-taking behavior in the latter scenario, as it is often the case that we are forced to make choices when we are uncertain of the outcome. The Balloon Analogue Risk Task (BART) was developed to fill this niche in risk-taking behavior studies.

The BART is widely employed in academic and clinical settings to assess risk-taking behaviors. In the standard task, participants earn fixed amounts of money by incrementally pumping a virtual balloon via a keypress. At any point, the participant is able to stop pumping and collect the earned value. However, if the balloon is pumped too much and it pops, the points accrued for that balloon vanish. When deciding whether to pump or collect, participants must balance the prospect of small gains against the increasing possibility of loss due to the balloon popping. This process of comparing potential gains to potential losses is at the heart of prospect theory models (Kahneman & Tversky, 1979). In prospect theory, when faced with a series of choices in which the probabilities of possible outcomes are unknown, the person must alternate between exploring the outcome probability distribution and then exploiting the knowledge gained to maximize rewards.

The main objective of the BART is to mimic real-world scenarios in which the person must maximize a beneficial reward in an environment where the probability of a negative

outcome is initially completely unknown (Lejuez et al., 2002). The BART is a type of n-armed bandit task - a category of psychological tasks in which there are negative consequences to certain actions. The BART, like other n-armed bandit tasks, creates an experimental environment in which participants must balance between exploration and exploitation of the task. Participants have no prior knowledge of the likelihood of the balloons' popping points and must learn the balloons' popping probabilities through trial and error. The BART has excellent test-retest reliability both behaviorally and neurally (White et al., 2008; Li et al., 2020). However, there are multiple drawbacks with the original BART. There are three different balloon types, each with a different pop probability. The upper limits of these balloon types are vastly different, with the largest pop distribution having an upper limit of 128 pumps before exploding. This makes it very difficult for participants to learn the probability of the balloon popping and unnecessarily increases task completion time. The performance metric used in various BART studies (the adjusted score) ignores the information in popped balloons, yet it is still in use today (Plescak et al., 2008; Chandrakumar et al., 2018).

BART performance has been closely correlated with various self-report measures of risk-taking behaviors in real-world settings, including alcohol abuse, drug abuse, and risky sexual behaviors (Lejuez et al., 2003). The task has also been useful in assessing differences between various clinical populations and drug trials (Reddy et al., 2014). As we will see in the results, most healthy participants demonstrate slightly conservative behavior on the BART (Trepel et al., 2005). Our new version is aimed at quantifying differences between individuals across the entire spectrum of risk processing and examining the sequential dependencies of the decision-making process. Here, we introduce a new version of the task that includes variable rewards and more identifiable balloon categories that induce meaningful choices on each trial, providing more room for response variability. Greater response variability between individuals allows us to better capture individual differences even among a cognitively healthy population, providing support for a more attuned representation of the cognitive process at play during risk-taking.

We created a new variant of the BART with five goals in mind: 1) Encourage participants to learn the pop probabilities of the balloons, 2) Provide opportunity for participants to explore different risk-taking behaviors while minimizing task length, 3) Develop a new methodology for assessing BART performance that uses all data, 4) Compare several possible cognitive mechanisms that could be at play during the decision-making process, and 5) Assess the efficacy of incorporating reaction times into our generative cognitive BART model.

2. METHODS AND MATERIALS

2.1. Participants

85 participants (age range 18-43 years; mean = 20 years) were recruited via posted fliers from the University of Virginia and greater Charlottesville, VA community. As mandated by the Institutional Review Board for Social & Behavioral Sciences (IRB-SBS), participants provided written informed consent. Participants were compensated $10 per hour.

2.2 Experiment Testing Environment & Apparatus

A desktop computer with Windows OS and a 1920 x 1080 24" 144Hz display was used to display the computer-based tasks. Participants performed a series of cognitive tasks multiple times in a cognitive battery (see SUPREME). All experiment instructions and task stimuli were presented via the State Machine Interface Library for Experiments (SMILE; https://github.com/compmem/smile). SMILE records participant responses, response times, and the timings of all experiment-related stimuli. Participants made "pump" or "collect" responses using the outer-two buttons of a four-button Black Box Toolkit response pad.

2.3 Task Design

Participants read the BART task description and instructions and then perform a brief practice session in which three gray-colored balloons could be pumped or collected. These practice balloons could also pop if pumped to their popping point. Once the practice session finished, the participants would be informed that they would start the actual task.

As illustrated in Figure 1, the decision maker (DM) sees a colored balloon on the screen. The balloon initially begins with a value of $0.00 if collected, and the bank (which is to one side of the screen) begins with $1.00. A fixation cross appears in the box below the balloon, followed by a monetary amount selected from a uniform reward distribution of $0.05 to $0.25. The participant must either press the '1' button or the '4' button on a button response box. Counterbalanced across participants, one of the buttons performs the pump action, the other performs the collect action. Selecting the pump action causes the pump value in the box below the balloon to be added to the balloon's total net value. The balloon will then either grow slightly or it will pop. If the balloon pops, the balloon will pop visually, and a new balloon will then appear after 750 ms. If it does not pop, then the participant is given another opportunity to pump or collect. If the participant selects the collect action, the pump value is not added to the balloon. Instead, the balloon moves over to the bank, and the balloon's net value is added to the total within the bank. It is important to note that once the pump value appears in the box below the balloon, 1 cent will be lost from the bank every 450 ms and the bank font color flashes to yellow. This rapid, intermittent deterioration of the bank's value incentivizes participants to make decisions quickly, thereby increasing the sense of urgency and making the task more emotionally salient, thereby increasing the sense in urgency with the intention of making RTs more meaningful.

Figure [1]. Image of the BART task.

There were a total of 18 balloons in each block of the BART. There were three different types of balloons, with six balloons for each type. We refer to these different types of balloons as different "bags" of balloons in the context of this experiment in the task instructions. Each balloon within a bag has a popping point that is selected from the same uniform distribution. Balloons from the same bag have the same color, thus providing participants the opportunity to learn the popping point range of each bag type through trial and error. The colors for each bag are selected at random. Participants were not told what the maximum and minimum popping points were ahead of time. They were told that each bag had a different "quality" of balloons, in that some balloon types could possibly be pumped more than other types. The different bags of balloons had the following uniform distributions from which a balloon's popping point was selected: [0, 8], [8, 16], and [0, 16]. Each balloon was given a popping point that was randomly generated from its respective pop point distribution. These distributions were selected to best differentiate participant behaviors. It has been noted that the outcomes of the first balloons can significantly impact a DM's perception of the balloons' pop probabilities (Koscielniak et al., 2016; Schürmann et al., 2019). The presence of three different pop probability distributions helps to mitigate this effect, as a balloon pop early a bag is likely to only affect the DM's interpretation of that bag's pop distribution. The presence of three different bags also encourages the DM to learn and differentiate those distributions. The [8,16] balloons provide an opportunity for the DM to pump more often, with no chance of the balloons popping until the eighth pump at the earliest. The presence of these different pop distributions allows participants who are so inclined to explore these distributions, indicating that they are willing to take risks, at least to a certain extent. DMs who consistently collect the [8,16] balloons prior to the eighth decision point are performing conservatively. This balloon structure allows us to tease out differences in risk-taking behavior without need of more time for the participant to pump balloons with a much larger distribution of pop points.

3. ANALYSIS

3.1 Assessing Performance: A New Performance Metric

Performance on the BART has conventionally been assessed via the adjusted BART score, which is calculated by taking the average number of pumps made on balloons that were collected (Lejuez et al., 2002). The adjusted score is exceedingly useful in detecting large-scale differences between healthy and unhealthy participants (Lejuez et al., 2003, Reddy et al. 2014),. However, we have found that there is much room to further refine our model-based approaches and gain a deeper understanding of the trial-level decision-making process within the context of the BART. We identify three issues with the adjusted score. Firstly, all information from the popped balloons is ignored (i.e., censored) (Pleskac et al., 2008; Chandrakumar et al., 2018; Young & McCoy, 2019). Secondly, the adjusted score does not take into consideration the fact that participants are learning the distribution of the balloons' popping points. For example, if a DM pumped the previous four balloons 20, 15, 9,and 25 times prior to the balloon popping or collection. There is then evidence that they should pump on the 6th trial of the next balloon since they have not experienced a pop prior to pumping the balloon 9 times. Even if the balloon were to pop on that 6th trial, it is informative that the DM made the attempt to pump in light of past experience. This brings us to the third oversight of the adjusted score: When a balloon pops, it is impossible to know for certain at what point the participant would have collected the balloon. The adjusted score does not account for this. Indeed, the creators of the BART themselves acknowledge that the adjusted score is biased, as balloons may terminate early (Pleskac et al., 2008). In response, they developed a BART variant in which a DM must decide at the onset of each trial how many times to pump each balloon. However, evidence suggests that making risky decisions a priori versus sequentially and in the moment can result in different risk-taking behaviors (Figner et al., 2009; Dijkstra et al., 2020). In order to gain a more nuanced understanding of the differences between individuals, we developed a novel metric that is sensitive to risk-taking differences within a healthy participant population.

The performance metric we present here is based upon the model mechanisms rooted in prospect theory itself. Our metric, denoted throughout as *M*, is on a scale from 0.0 to 1.0, with 0.5 being optimal performance. We define optimal performance, O, as that exhibited by an omniscient DM who knows the underlying pop probability distributions of the balloons and the reward distribution ahead of time. To calculate *M*, we first calculate the expected value E for every possible decision point in the balloons the DM experienced in the task. This E value is calculated from the perspective of an omniscient DM that already knows the true average of the reward distribution and the popping point ranges for the different bags. E is determined strictly by monetary values and the probability of popping. E diminishes as the number of pumps increases, eventually becoming a negative value. When E is positive, the optimal choice O is to pump the balloon and realize E. When E is negative, the optimal choice is to stop and collect the balloon, thus not realizing E and instead realizing zero. We take the difference between the E values realized by the DM and O for every trial. It is important to note with this methodology, if pumping a balloon results in a pop, the decision to pump is the optimal decision as long as E is a positive value.

$O = E$, when $E > 0$ (1)
$O = 0$, when $E <= 0$ (2)

Thus, deviation from optimal performance is the cumulation of the difference between optimal and realized values We calculate this sum for each balloon *b* as in Equation 3 below.

$$m_b = \sum_{i=0}^{stop} O - E(realized)_i \quad (3)$$

We then scale each m_h score by the following:m_b score by the following:

$$M_b = 0.5 - \frac{0.5 * m_b}{E_{initial}} \text{, when } m_b < 0 \quad (4)$$

$$M_b = 0.5 + \frac{0.5 * m_b}{E_{pop}} \text{, when } m_b > 0 \quad (5)$$

$E_{initial}$ is the optimal expected value of pumping on the first pump of a given balloon. The first pump of a balloon has the largest optimal expected value, as there is no penalty for popping a balloon that does not yet have any value in it, and there are potential rewards to be gained by pumping. E_{pop} is the optimal expected value at the popping point, and is a negative value as the only outcome for pumping at this point would be to lose the value of the balloon. Our performance metric *M* is then the overall average performance of all M_b across all balloons. This performance metric not only shows how suboptimal a participant is performing, but it also shows in what direction they are performing suboptimally. Scores below 0.5 indicate that the participant was more risk-averse and less willing to risk popping the balloon. Scores above 0.5 indicate that the participant was willing to take greater risks.

3.2 Model Mechanisms

Our model of the BART is based upon a prospect theory Bayesian sequential risk-taking model (BSR) account of the task proposed by Wallsten and colleagues (Wallsten et al., 2005). This prospect theory account posits thus: Given an ambiguous probability of a negative net gain, the DM compares the potential gain to the potential loss. Since humans have various biases, the DM's perception of value for the potential gains and losses is not perfect. Instead, it is modulated by their own predisposition to avoid or engage in risk-taking behavior. We refer to the DM's personal interpretation of value as the *expected value*, *E*. In this study, we present several possible mechanisms by which *E* may be modulated.

3.2.1 Estimating the Probability of a Balloon Popping

One of the factors governing the value of *E* is the DM's estimation of the likelihood of the balloon popping on a given decision/trial. An assumption of this model - and many other prospect theory models in general - is that the DM possesses an initial estimate of this likelihood. In the case of the BART, that likelihood is based upon the DM's estimate of how many times they can pump the balloon until it pops. We call this estimated popping point *n*. A DM could assume that the pop probability, *p*, grows in an exponential fashion (as it truly does) or in a linear fashion. For a DM that assumes the former, the popping point *p* at trial *i* is defined as:

$$p = \frac{1}{n - i + 1} \quad (6)$$

If the DM assumes *p* grows in a linear fashion, it is defined as:

$$p = \frac{i}{n} \quad (7)$$

3.2.2 Thinking Ahead: Calculating Expected Value and Considering Future Rewards

An individual making sequentially dependent decisions can do so in one of two ways: 1) consider the potential reward immediately available in comparison to the potential loss, or 2) consider the immediate potential reward in conjunction with future possible rewards, comparing the total sum of immediate and future rewards to the potential loss. If the DM considers only the current rewards, the subjective value of pumping on trial *i* is:

$$E(\text{pump})_i = qr_i^{\gamma} - \theta pb^{\gamma} \quad (8),$$

where *i* denotes the current trial, *q* is 1-*p* (the probability of the balloon not popping if the DM chooses to pump), *r* is the amount of money that would be added to the balloon if the DM chooses to pump, and *b* denotes the current value of the balloon. θ is a free parameter for loss aversion. γ is a free parameter modulating nonlinear perceptions of monetary value. We present model variants with both one γ and two separate γ parameters, denoted by γ^{+} for modulating rewards and γ^{-} for modulating losses, respectively. There is much considerable evidence supporting the view that participants can have asymmetric valuations of rewards and losses (Trepel et al., 2005; Schonberg et al., 2011).

If the DM does, however, take into account future possible rewards, then the value of pumping is

$$E(\text{pump})_i = qr_i^{\gamma} + E(\text{future}) - \theta pb^{\gamma} \quad (9)$$

The equation for calculating *E(future)* is as follows:

$$E(\text{future})_i = \sum_{k=i+1}^{n_h} q_k \, \bar{r}^{\gamma} \quad (10)$$

E(future) is the summation of the mean reward $\bar{r}$ seen thus far times *q*, up to the estimated stopping point *n*. This method discounts possible future rewards based upon the likelihood of reaching that pump number *i* in the balloon. Simply put, less likely future rewards are down weighted in value. To summarize, the equation above shows that *E* is the comparison of the current and possible future rewards that this balloon could yet provide with the potential loss if the balloon were to pop. *E(future)* diminishes as *i* approaches *n*. $\bar{r}$ seen thus far times q, up to the estimated stopping point n. This method discounts possible future rewards based upon the likelihood of reaching that pump number i in the balloon. Simply put, less likely future rewards are down weighted in value. To summarize, the equation above shows that E is the comparison

of the current and possible future rewards that this balloon could yet provide with the potential loss if the balloon were to pop. E(future) diminishes as i approaches n.

With *E* calculated, the DM must decide whether to pump the balloon or to collect the balloon. Our models posit that the DM compares the potential gains against the potential losses. By this reasoning, the DM should be more likely to pump if the expected reward is greater than the potential loss and be more likely to collect when the perceived loss is greater than the potential reward. Figure 2 illustrates this concept.

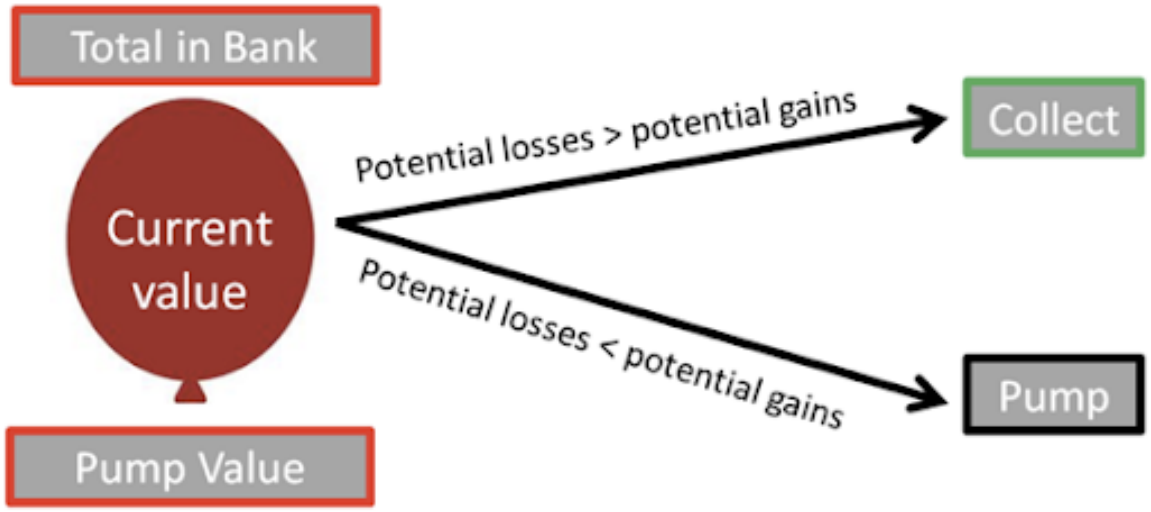

Figure 2: Prospective Gains v Losses

3.2.3 Choice Probability Function

The mechanism by which the likelihood of the DM's choice on a trial *i* is determined by the following sigmoid equation:

$$d_i = \frac{1}{1 + e^{-\tau * E_i}}$$ if the DM pumped, or (11) if the DM pumped, or

$$d_i = 1 - \frac{1}{1 + e^{-\tau * E_i}}$$ if the DM collected. (12) if the DM collected.

where τ is a temperature parameter representing the DM's predisposition to selecting an action at random.

3.2.4 Modelling Choices and Reaction Times

As stated prior, a main objective of this work is to assess the utility of incorporating choice reaction times into a BART model. To do so, we change the pump probability function to be a Wiener first passage of time (WFPT) sequential sampling model (Stone, 1960, Navarro & Fuss, 2009). In the WFPT, there is one noisy accumulator with two boundaries, each representing a different decision. In this case, the distance between the two boundaries is a free parameter *a*.

The initial starting point for the accumulator is free parameter *w* and represents bias toward a particular choice. The drift rate, *v*, is calculated via the following:

$$v = \beta - E_i, \quad (13),$$

Where β is a free parameter modulating E_i. Non-decision time, t_0, represents perceptual processing and motor engagement time. RTs are calculated from the onset of the reward appearing on the screen up until the participant makes a response. Figure 3 provides a visual representation of how the accumulation process works.

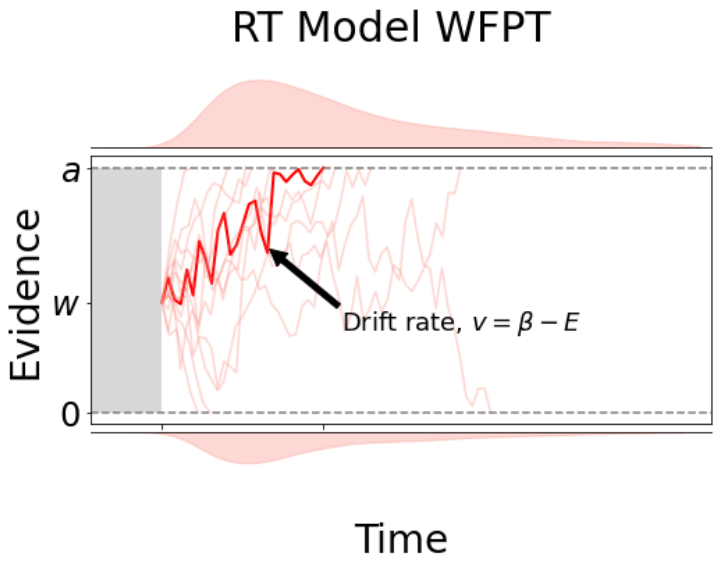

Figure 3: BART Implementation of WFPT. β, a, *w*, and t_0 are free parameters. β biases drift rate. *w* is the starting point of the evidence accumulator (i.e., initial choice bias). a is the distance between both decision boundaries. t_0 is non-decision time

3.2.5 Updating the Perceived Balloon Pop Probability Distribution

Once a decision is made, the outcome of the decision is considered. In this version of the BART, we have three different categories of balloons, each with different popping point distributions. Our model has three separate initial *n* parameters that are updated independently of one another, providing a mechanism by which the model can account for the DM learning these different pop distributions. *n* is a free parameter. We explore model variants in which each initial *n* is fit independently as well as when each *n* is set to the same initial value. To distinguish the *n* parameters, we denote them with *h* for a bag type, resulting in three *n* values: $n_{0,8}$, $n_{0,16}$, and $n_{8,16}$. The degree to which n_h deviated from the resulting popping point on trial *i* is calculated. $n_{0,8}$, $n_{0,16}$, and $n_{8,16}$. The degree to which n_h deviated from the resulting popping point on trial i is calculated.

If the balloon does not pop,

$$\Delta n = p * (n_h - i). \quad (14)$$

If the balloon does pop,

$$\Delta n = q * (i - n_h). \quad (15)$$

This makes Δn a greater magnitude of change when the previous estimate of n_h differs greatly from *i*. Also, this rule asserts that more information is available for updating n_h when the balloon pops than when it does not. This is reasonable, as from the DM's perspective, pump decisions in which the balloon does not pop only slightly reinforce their estimate of the likelihood of the balloon popping. How different DMs respond to prediction errors/reinforcements is dependent upon the individual. To capture this difference, we use the following:Δn a greater magnitude of change when the previous estimate of n_h differs greatly from i. Also, this rule asserts that more information is available for updating n_h when the balloon pops than when it does not. This is reasonable, as from the DM's perspective, pump decisions in which the balloon does not pop only slightly reinforce their estimate of the likelihood of the balloon popping. How different DMs respond to prediction errors/reinforcements is dependent upon the individual. To capture this difference, we use the following:

$$n_h = n_h + \alpha \Delta n, \quad (16)$$

where α is a free learning rate parameter. α scales the magnitude of the prediction error/reinforcement. This scaled difference is then added to n_h. We compare model variants in which the n_h is updated after every trial to variants in which n_h is updated only at the conclusion of a balloon.α is a free learning rate parameter that scales the magnitude of the prediction error/reinforcement. This scaled difference is then added to n_h. We compare model variants in which the n_h is updated after every trial to variants in which n_h is updated only at the conclusion of a balloon. When n_h is updated on every trial causes it to slowly increase over the course of sequential pumps. This implies that pump choices that do not result in a pop provide slight evidence/confidence for pumping again.

3.2.5 Modelling Sequentially Dependent Choices

Generative cognitive models provide an ideal test-bed for assessing the validity of hypothesized cognitive processes. Like other cognitive models, they can be used to extract parameter estimates that most likely gave rise to the observed behaviors (Busemeyer & Stout, 2002; Huys et al., 2016). The likelihood of this A critical factor in many BART models based upon prospect theory is the sequentially-dependent nature of choices. The likelihood of a DM's specific pattern of behavior becomes increasingly low as the number of choices made increases. This is further compounded when one considers the number of events in which the DM is forced to update their estimated outcome probabilities (i.e., update their pop probabilities after a choice). Therefore, an accurate model of the BART should, given a set of priors enveloping the correct posterior distribution of true parameter values, be able to reliably predict the exact sequence of choices made once fit to the observations. Our analytic likelihood function is the product of all the observed choice likelihoods. The likelihood function for our observed data *X* for our choice models is thus:

$$L(X \mid \alpha, \tau, \gamma, \theta, n_h) = \prod_{B=1}^{N} \prod_{i=1}^{\Omega} d_{i,B} \quad , \ (17) \ ,$$

where *B* is the balloon number, *N* is the number of balloons, *i* is the trial number, and Ω is the trial at which point the last decision was made. Ω is the trial at which point the last decision was made.

For the RT Models, the likelihood function is an implementation of WFPT in which the drift rate v, observed RT, and observed choice are passed into a WFPT model that generates a distribution of plausible RTs and choices, thereby calculating the likelihood of the observed choice and RT. For trials in which the RT was faster than 200 ms or slower than 5 seconds, the likelihood was not updated due to WFPT being too inflexible to capture dramatically fast or slow RTs.

Table 1 below presents descriptions of each parameter. Parameter priors were selected to be uninformatively wide.

Parameter	Description	Prior	PLUM Type
α	Learning rate	$logit^{-1}(\mathcal{N}(0.0, 1.4)$	Both
γ	Shape of reward curve	$logit^{-1}(\mathcal{N}(0.5, 0.35) * 2)$	Both
θ	Loss aversion	$\mathcal{N}(0.0, 5.0)$	Both
n_0	Starting estimated popping point for each bag	$logit^{-1}(\mathcal{N}(-0.71, 0.80))$	Both
τ	Choice randomness	$\exp^{\mathcal{N}(1.0,1.0)}$	Choice
β	Drift rate bias	$\exp^{\mathcal{N}(1.0,1.0)}$	RT
a	Decision boundary	$logit^{-1}(\mathcal{N}(-1.6, 1.5) * 25)$	RT
w	Decision starting bias	$logit^{-1}(\mathcal{N}(0.0, 1.4))$	RT
t_0	Non decision time	$\mathcal{N}(1.0, 3.0)t_0$	RT

Table 1: Model Parameters

Table S1 in the supplementary materials contains a breakdown of which mechanisms are present in each model.

3.3 Fitting Models

All models were fit to each subject and block independently. As in the Table 1, we define a set of priors for each parameter. Using differential evolution with Markov Chain Monte Carlo (DE-MCMC), we generate parameter samples from the priors, calculating the likelihood of the observed data *X* given the parameter values (Turner & Van Zandt, 2012; Turner et al., 2013; see Weichart et al., 2020). To do so, we used the Python library RunDEMC (https://github.com/compmem/RunDEMC) . We utilized 50 chains for the choice model and 80 chains for the RT model, performing 600 iterations of burn-in to find the maximum a posteriori (MAP) estimate, followed by 1400 iterations of sampling with a purification step every 5 iterations.

4. RESULTS

We first examine the behavioral results and the proposed performance metric *M*. We then perform a model comparison to investigate which of the proposed cognitive mechanisms are the most informative in the choice-only models. We then do the same for the RT model versions. Finally, we will compare the winning choice-only model to the winning RT model.

4.1 Behavioral Results

We see that performance on the BART task was on average below optimal performance, indicating that participants were likely to behave conservatively. This is a common finding in risk-taking studies that cognitively normal participants are more likely to be risk-averse (Tom et al., 2007). Figure 4 shows that the distributions of performance for each block were very similar, and that, on average, participants performed in a slightly conservative fashion. Figure 4a (84 participants, 2 blocks each, $n = 164$) shows that, on average, participants performed nearly optimally on the [0, 8] bag. Participants were increasingly suboptimal on average for the other bags and tended to behave conservatively. This indicates that many participants were unwilling to fully explore the full popping point distributions of the [0, 16] and the [8, 16] bags. It is especially interesting to note that when no balloons popped prior to the eighth pump in the [8, 16] bag, many participants were still leery of pumping too many times.

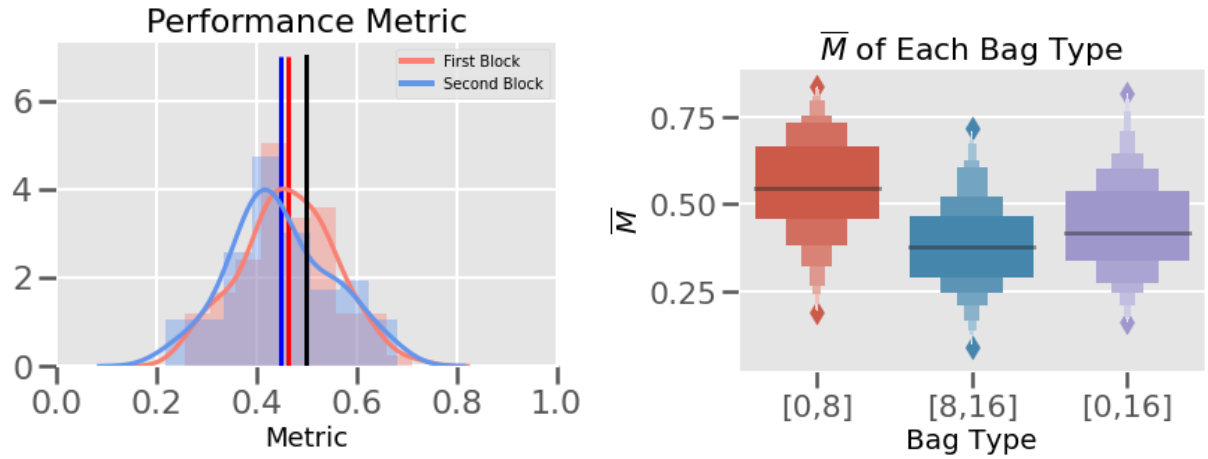

Figure 4: Performance Metric: 4a). Participants (N=84) scored on average below optimal performance of 0.5 (the black vertical line). First block mean: 0.46; Second block mean: 0.45. 4b) Th

To assess the sensitivity of our performance metric, *M*, we compare it to the adjusted BART score. Figure 5 shows how *M* strongly correlates with the adjusted score, but is able to differentiate deviance from optimal performance. This suggests that *M* could be useful in BART studies assessing individual differences in risk taking without the need to remove data. It also provides a sense of in the direction (i.e., conservative or risk-taking) of the risk taking behavior within the DM's specific trial-level context.

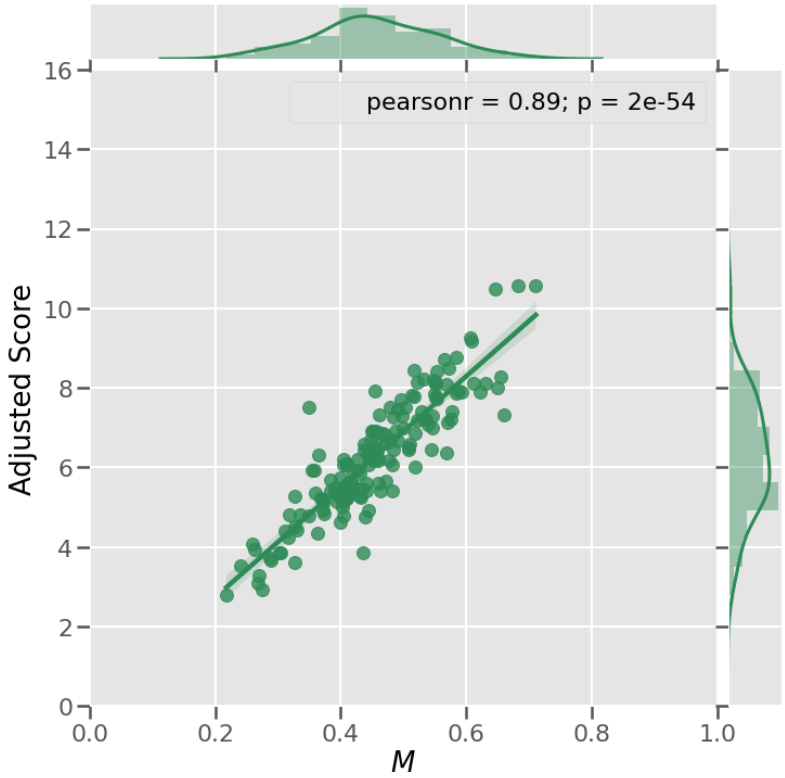

Figure 5: Adjusted Score vs. Deviation from Optimal (M) Across All Blocks. (N=163)The black vertical line represents optimal performance ($M = 0.5$). *M* correlates strongly with adjusted score, indicating that it would be useful in diagnostic studies that utilize adjusted score whilst being more sensitive to true performance.

4.2 Modelling Choices: Choice Model Results and Comparisons

4.2.1 Choice Model BPIC Comparison

After fitting each model to all subjects within each block indepently, we calculate the Bayesian Predictive Information Criterion (BPIC) of each fit. BPIC is a useful tool for assessing goodness of fit, as it assesses model fit while penalizing for more parameters (Ando, 2007). Lower BPIC scores indicate a better model fit.

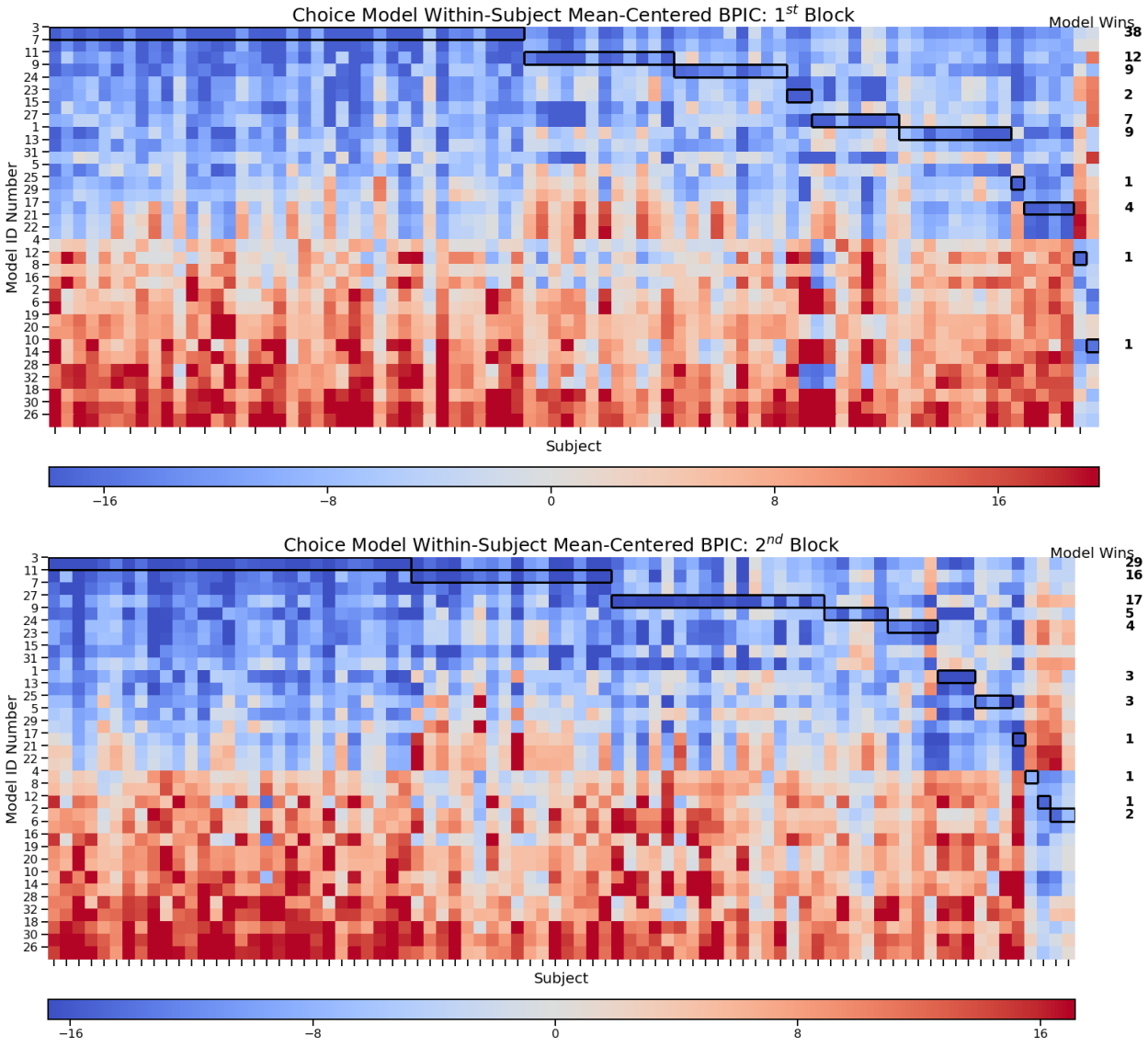

Figure 5: Comparing Choice Model Fits. Within-subject mean-centered BPIC scores are presented. The models are arranged in descending order of model average BPIC across participants. Lower BPIC values (blue) represent a better fit. For a given model, any participants that it fit to better than the other models are outlined in black. The number of participants a model had the best fits for are at the right of the heatmap. Refer to Table S1 in the supplementary materials for a breakdown of each model, listed by Model ID Number.

Model number 3 was the overwhelming winner, as it had the lowest BPIC score for the most subjects across both blocks. We refer to model 3 as the Predictive Linear Utility Model (PLUM). The assumptions of the PLUM model are as follows:

1. The DM assumes the likelihood of a balloon popping increases linearly on successive pumps (equation 7)
2. The initial estimated popping point n_0 for each bag type is same.
3. The DM is taking into account the prospect of future gains on consecutive pumps (equations 9 & 10).

4. The DM updates n_h only at the conclusion of every balloon.

In summary, the PLUM model is composed of the following mechanisms:

$$p = \frac{i}{n} \quad (7)$$

$$E(\text{pump})_i = qr_i^{\gamma} + \sum_{k=i+1}^{n_h} q_k \overline{r}^{\gamma} - \theta p b^{\gamma} \quad (16)$$

Figure 6 below puts equation 16 into the experimental context.

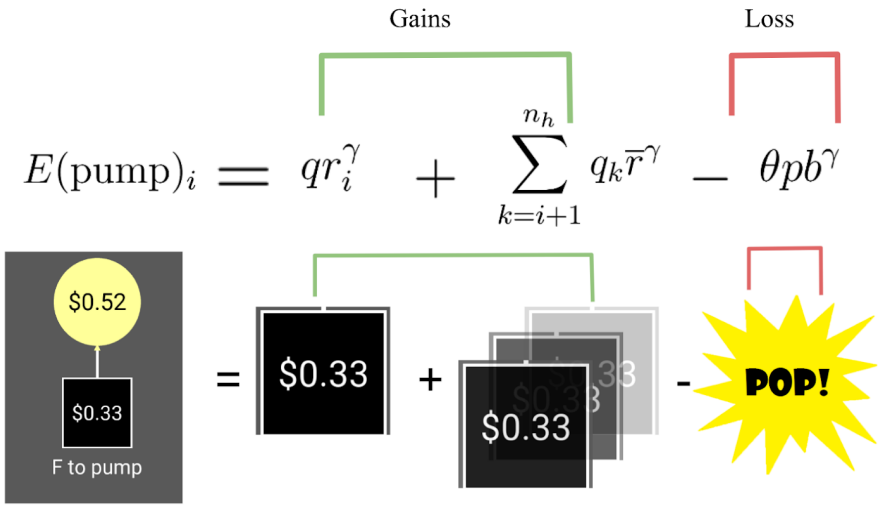

Figure 6: Representing Expected value in PLUM.

4.2.3: Simulations

Using the best fitting parameters from our model fits, we are able to simulate the sequence of choices made by participants in the exact order they were made. We simulated each participants' behavior 1000 times. We then score each simulation via our performance metric *M* and compare the average *M* of the simulations to the participant's true *M*. Figure 7 shows the correlations of simulated average*M* and true observed *M*. The correlations (r=0.98, p<0.000) for both blocks are very strong, suggesting that the model is capturing the mechanisms well.

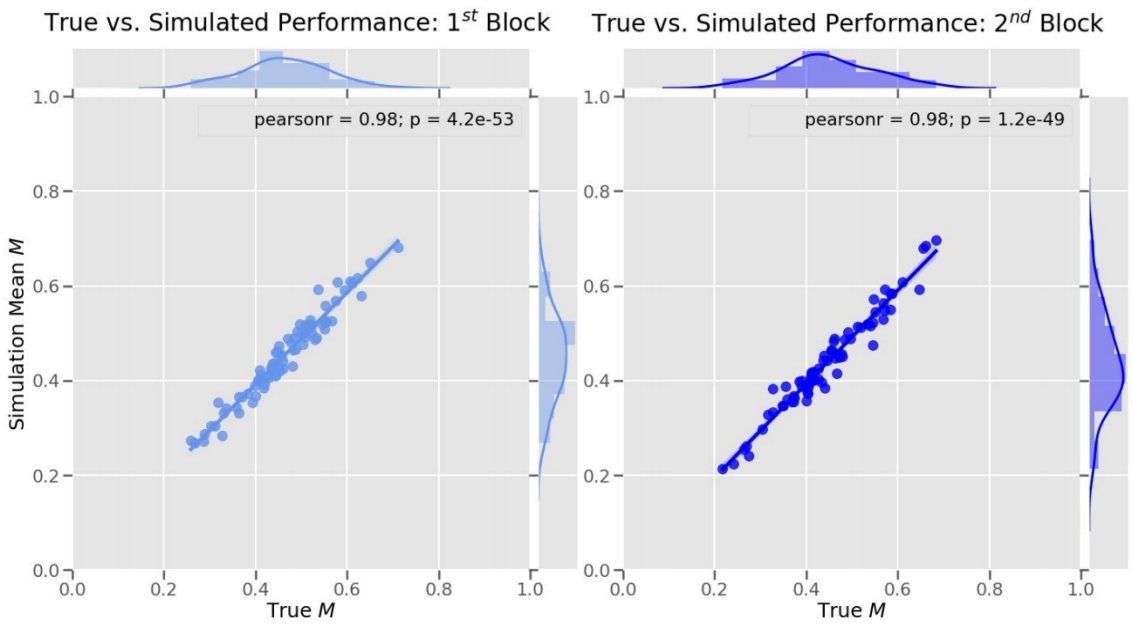

Figure 7: True vs. Simulated Performance for Choice Model: Simulations using best-fitting parameters were robustly accurate in recreating observed patterns of pumping and collecting.

We next examine the relationships between model parameters and *M*. We correlate the best fitting parameters for each participant's *M* for each block.

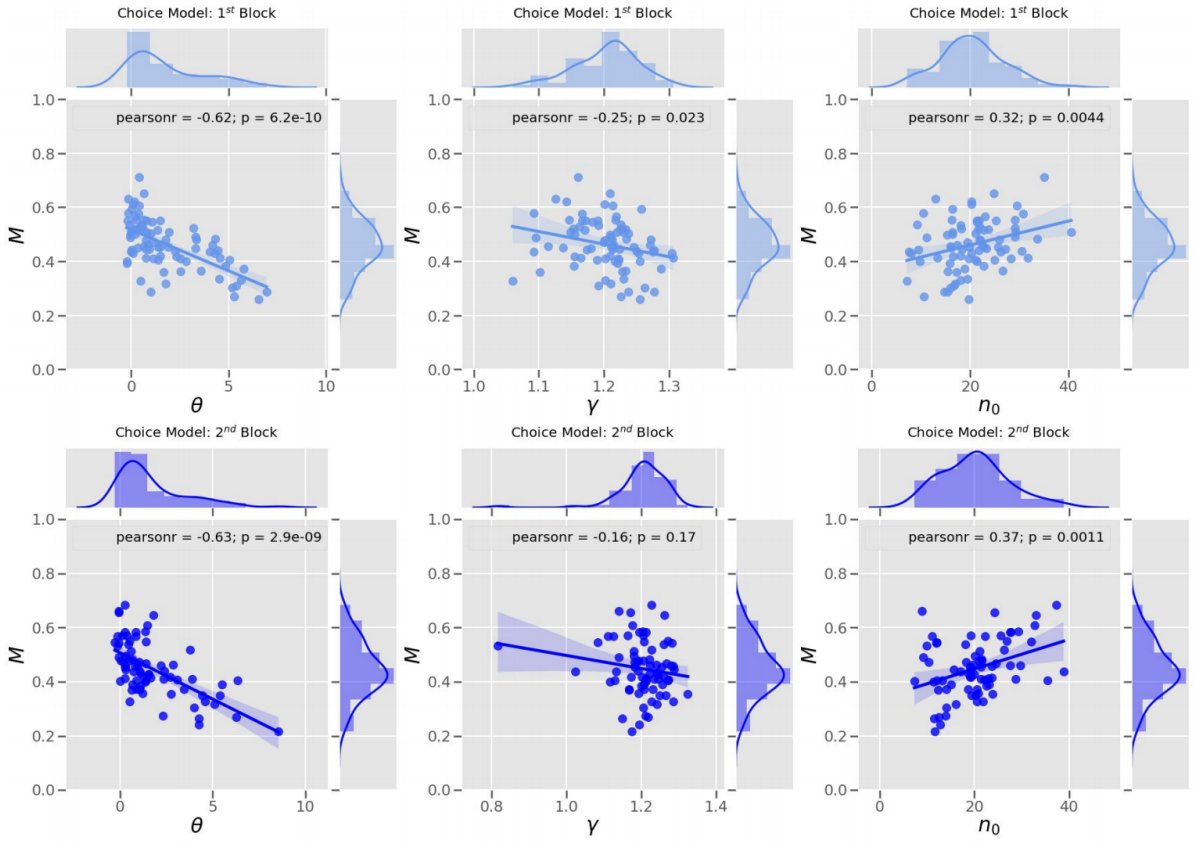

Figure 8: Parameters predicting performance: The parameters θ, γ (for the first block), and n_0 are significantly correlated with *M*. See table S2 in the Appendix for all pairwise comparisons.

As we would expect, loss aversion θ is negatively correlated with *M* for both blocks, indicating that more loss-averse participants performed more conservatively. n_0 is positively correlated with *M*, but was not significantly correlated with θ (r = 0.127, p = 0.12). This suggests that participants similarly loss-averse could have different initial estimates of the pop probability. θ is negatively correlated with M, indicating that more loss-averse participants performed more conservatively.

4.3 Modelling Choices & RTs: Reaction Time Model Results

For the RT model, we replicated the same approaches performed for the choice model. As stated before, the notable exception for fitting the RT model is that the likelihood function was not updated on trials that had a choice RT of less than 200 ms (20% of trials for the first block, 25% of trials for the second). This leads to the RT model being fit to fewer trials when compared to the choice model. Exploring the BPICs below, we see once again that the PLUM was the winner (see Equation 16).

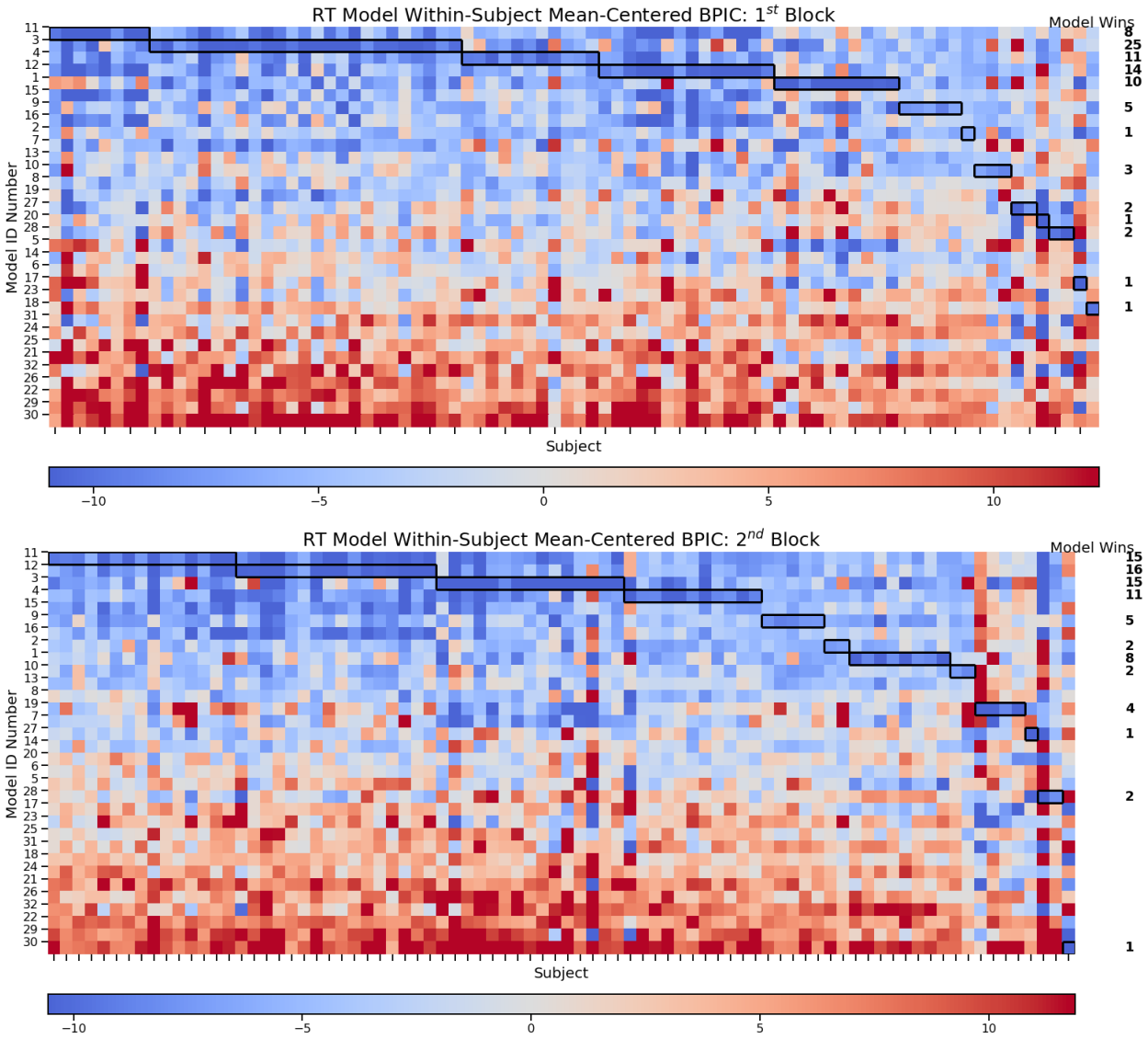

Figure 9: RT Model BPIC comparison. The models are arranged in descending order of model average BPIC across participants. The RT version of PLUM had the lowest score for 25 participants in the first block and 15 participants in the second block.

We performed simulations using best-fitting parameters. Though the RT version of PLUM was not fit to every choice, we still simulated choices for all trials.

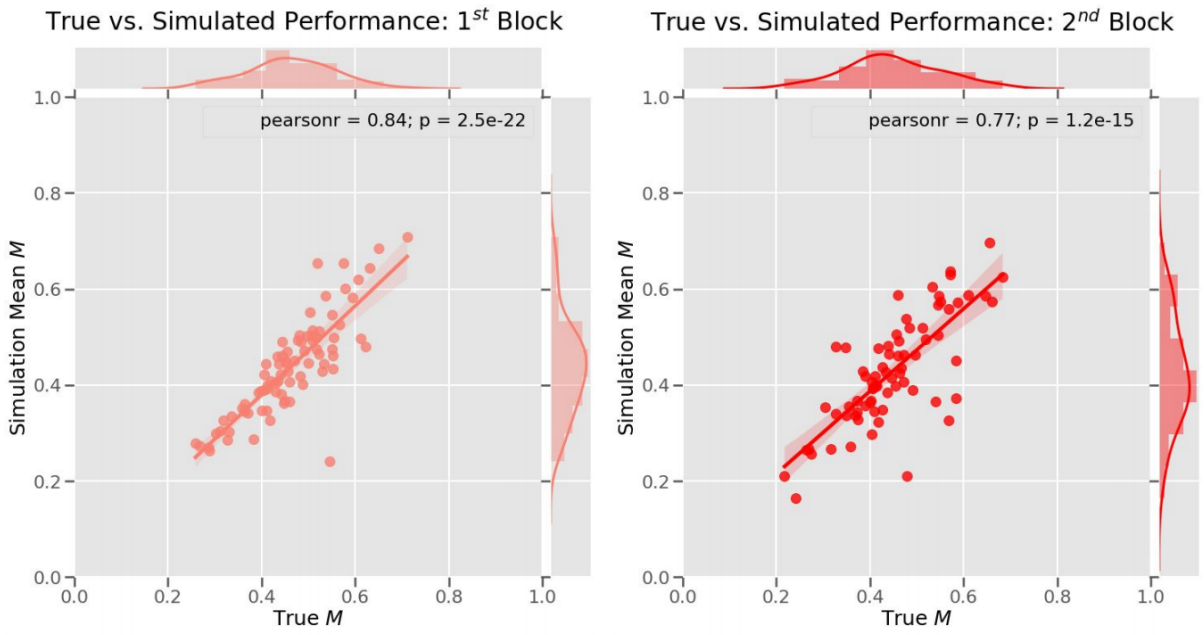

Figure 10: RT Model Predicted Performance. Simulations using best-fitting parameters were reliably accurate in recreating observed patterns of pumping and collecting.

Simulation performance shows that the RT version of PLUM predicted *M* is strongly correlated with true performance. By using WFPT within a generative modelling framework, we are able to simulate RTs at every trial and compare these simulations to the observed RTs. Figure 11 portrays RTs as a function of distance from the collection trial for balloons that were collected. We see a trend for RTs to become longer the closer a DM is to the trial they collect on. We hypothesize that this is due to increased deliberation about the available E_i, for the utility of E_i drops the closer the DM moves toward n_h. Recent evidence suggests that this is the case, as RTs in the BART tend to slow down as the contrast between risk & reward grows (The simulated RTs follow the same trend as the observed RTs. E_i, for the utility of E_i drops the closer the DM moves toward n_h. The simulated RTs follow the same trend as the observed RTs.

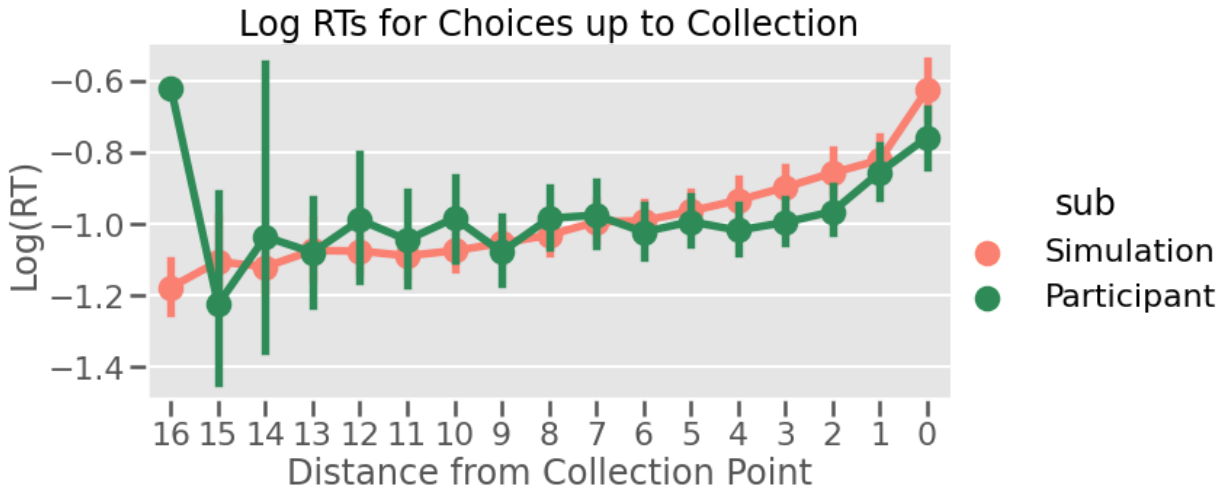

Figure 11: Comparing Observed & Simulated Reaction Times: We see the trend that RTs elongate the closer a DM is to the collection trial, represented here as 0 on the x-axis. There are fewer observed RTs the further away we move from the collection point.

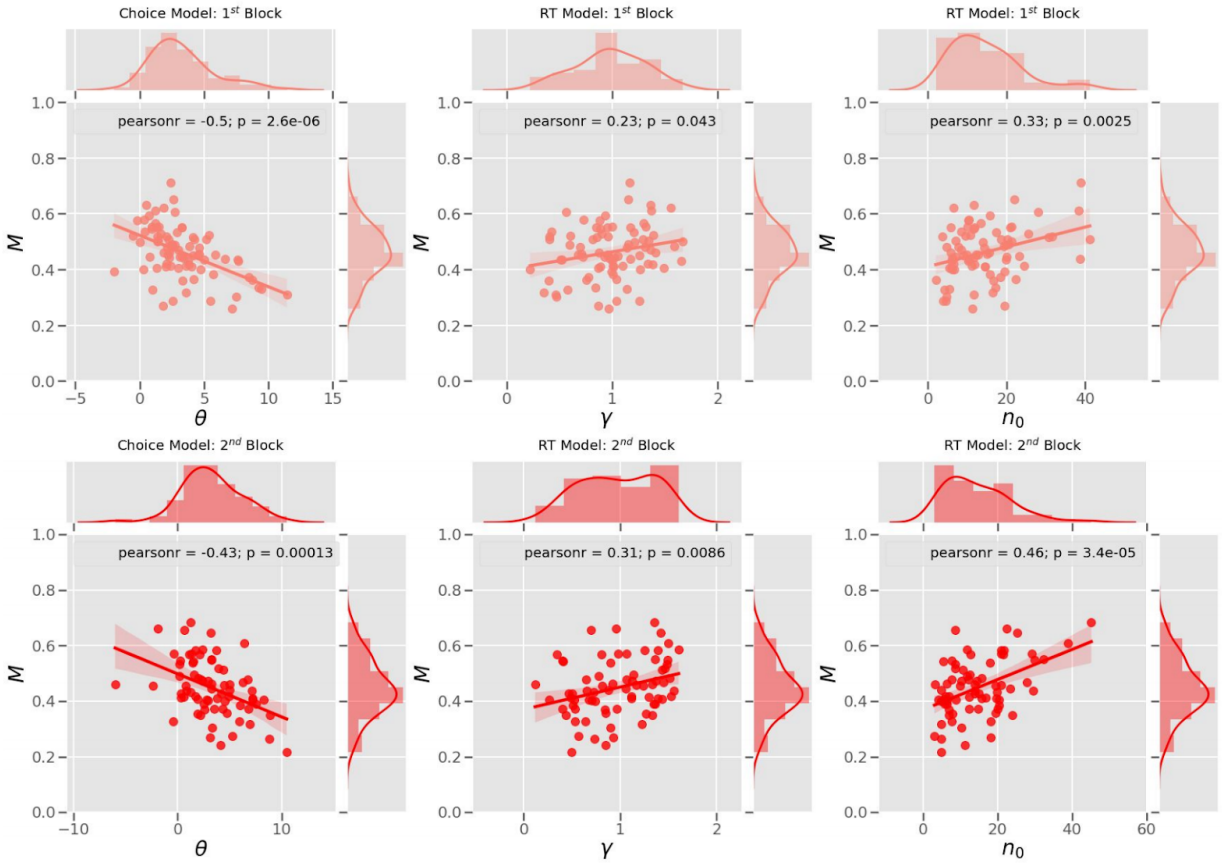

Figure12: RT-PLUM Parameter Fit and Performance Correlations. θ correlates moderately with *M*, while γ and n_0 correlate weakly with *M*. See table S3 for all model pairwise comparisons.

θ has a moderate correlation with *M*, while γ and n_0 are moderate/weakly correlated with *M,* respectively.

4.4 Comparing Models: Choice Model vs. Reaction Time Model

In order to compare the winning choice model to the winning RT model, we must utilize alternative approaches to the standard methodologies for drawing model comparisons. Since the RT model was fit to reaction times, it is being fit to more data per choice. However, the RT model was also fit to fewer trials, as the likelihood was not updated on trials with an observed RT faster than 200ms. BPIC comparisons and other similar methods can only be applied when the models are accounting for the same data. In an effort to partially circumnavigate this, we re-fit the choice-only version of PLUM, only updating the likelihood on the same trials the RT model was fit to. By doing so, we are able to more reliably compare the two models, as the only feature that differs between them is the consideration of the RTs themselves.. Therefore, we 1) compare the model fits' parameter posteriors for the mutual parameters and 2) compare each model's posterior predictive distributions (PPD) of *M*.

We hypothesize that RTs - if informative to the latent processes in the model - will lead to the RT model having more constrained parameter posteriors. To compare the fits between the two models' parameter posteriors, we first calculate the 95% highest density interval (HDI) for each parameter poster for each model. We then compare then compare the difference between the lower and upper bounds of each models' HDI to one another within each participant.

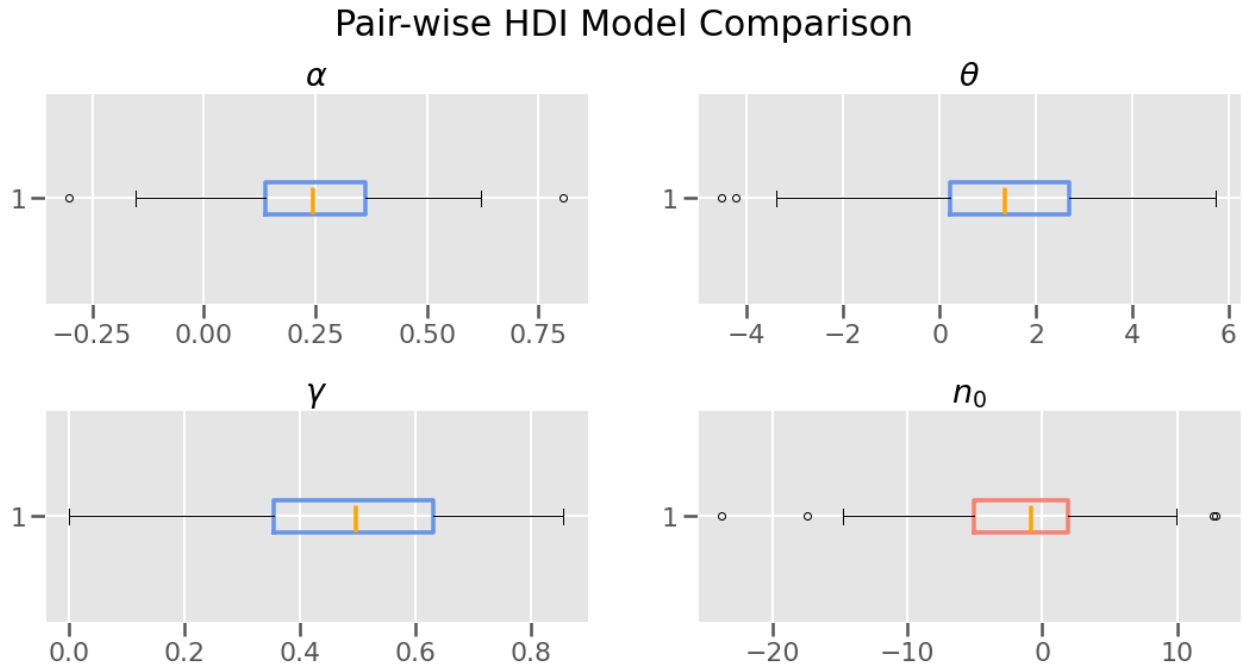

Figure 13: Comparing Model Posteriors. The model with more constrained posteriors indicates better performance. The range of the 95% HDI is calculated for each parameter posterior. We then subtract the choice-only PLUM posterior HDI range from the RT-PLUM posterior range. Blue box plots indicate that the choice model parameters were more constrained (RT HDI range - Choice HDI range > 0). The red box plot indicates that the RT model n_0 posterior was more constrained (RT HDI range - Choice HDI range < 0).

In the figure, the blue box plots indicate that the choice model's parameters were more constrained for α, θ, and γ. The red box plot shows that n_0 was more constrained for the RT-PLUM. The Choice-only version of PLUM had parameters that were more constrained. We conducted Mann-Whitney U tests to assess whether the distributions of HDI ranges were significantly different between the Choice-only PLUM and the RT-PLUM: α: U=12693.0, $p <$ 0.000; θ: U=33863.0, p<0.000; γ: U=1590.0, p<0.000; n_0: U=42589.0, p<0.047. These results show that the degree to which the two models constrained parameter fits were significantly different.

A PPD is a distribution of simulated or predicted observations conditional on the data that was observed (Gelman et al., 1996). To generate a PPD for a given subject, we sample multiple sets of parameters from each parameter's posterior distribution. With each sample, we then simulate behavior. The better-performing model will have more accurate choice-level predictions. For each model, we drew 100 parameter sets from the model fit posteriors. For each parameter set, we simulated behavior 100 times. Figure 14 shows the PPDs for one subject for both choice-only PLUM and RT-PLUM. Notice how the choice model simulated stopping points tend to be both more accurate and more constrained. In a similar manner as to how we compared parameter posteriors, we compare the PPDs between the choice-only PLUM and RT-PLUM. We calculate the PPDs 95% HDI within each subject and block, then evaluate the distribution of differences of max deviations from the true *M* within each subject. Given that the parameter posteriors are more constrained in the choice-only PLUM, it is no surprise that the RT-PLUM is less constrained than the choice-only version (Figure 15).

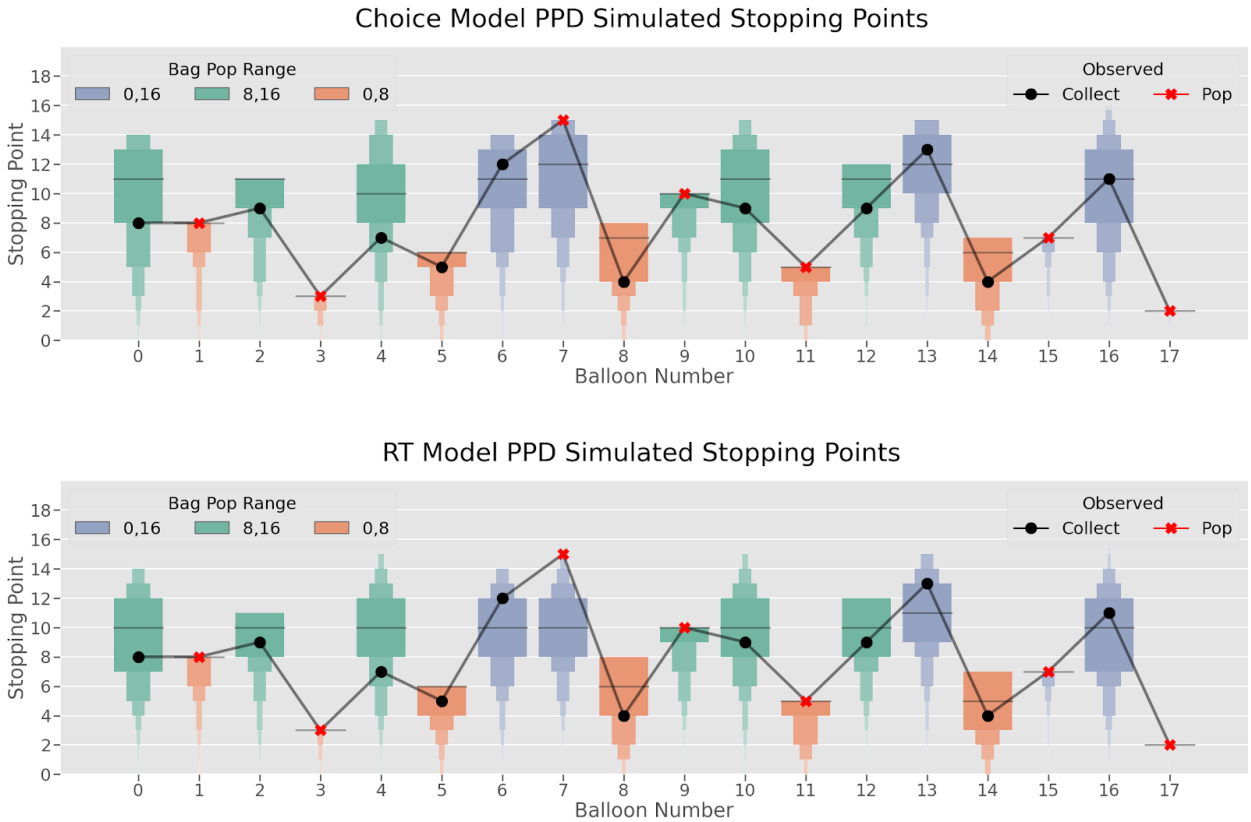

Figure 14: PPD Balloon-Level Performance. For this subject, we are able to see the distribution of points at which the simulations either popped or collected the balloon (box plots). We can compare this to the participant's pop and collection points (black dots for collection points, red x's for the popping points).

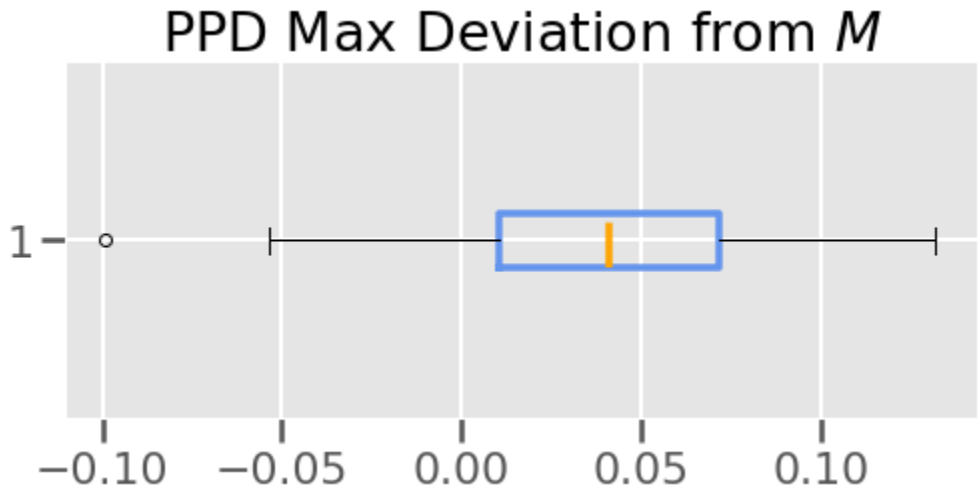

Figure 15: Comparing PPD Accuracy Between Choice-only PLUM and RT-PLUM. The upper and lower ranges of the 95% HDI for PPDs within each subject and block, then find the maximal difference between the upper/lower range of the HDIs and and true *M*. We then calculate the difference of this deviation from *M* between the choice-only and RT-PLUMs. The Choice-only PLUM PPDs were more constrained and more accurate than the RT-PLUM.

5. DISCUSSION

Our new version of the BART puts a DM in a risk-taking environment that encourages learning the likelihood of failure of different categories of balloons. The performance metric *M* provides more insight into whether a participant is performing conservatively or not by comparing their performance to that of an optimal DM. Critically, this metric does not censor any data and considers the trial-by-trial sequence of choices. We hypothesized several possible cognitive functions that could be involved in the BART. We identify PLUM as the winning model. This model assumes that the DM actively compares the potential gains to potential losses at each choice point. Unlike past BART models, PLUM projects potential gains into the future, meaning it assumes that the DM is not merely focusing on just the immediate possible gains, but also considering future gains. Reinforcing this is the fact that - when modelling only choices - the top three performing models contain this mechanism. PLUM also assumes the DM is updating their expected pop probability every choice while delaying in updating the estimated popping point until the end of the balloon. The manner in which the estimated pop probability increases, however, is linear in PLUM, indicating that DMs are not updating the probability of the balloon in a simplified manner and not as it is in reality. However, it should be noted that this could be due to constraints of the task. A future version of this task in which the different bags had pop distributions with greater upper limits would be useful for investigating this, as this would provide longer sequences of choices prior to a resolution (pop or collect).

When fit to RTs as well as choices, PLUM was still the winning model, but its performance was significantly worse compared to when PLUM was fit to choices alone. This is likely due to the noisy nature of the RTs displayed by participants, increasing the difficulty of the model fitting. In future work, changes to the task structure could help make more responses have more meaningful RTs. BART versions with increased monetary rewards/losses can impact risk-taking behaviors (Bornovalova et al., 2009). A possible variation would be adopting a snowballing reward structure in which the reward value would increase every pump. In spite of these disadvantages, the RT version of PLUM was still able to predict performance moderately to strongly. This leads us to believe that in our BART version, some, but not all RTs are meaningful. It is also plausible that the accumulation process in deciding whether to pump or not begins at the moment the previous trial's outcome resolves. Essentially, this would mean the moment a DM receives feedback that the balloon has popped or not, the decision making process for the next trial begins. A future version could jitter the presentation of choice feedback, measuring the RT from the moment of feedback onset.

References

Acheson, A., Richards, J. B., & de Wit, H. (2007). Effects of sleep deprivation on impulsive behaviors in men and women. *Physiology & Behavior*, *91*(5), 579–587. https://doi.org/10.1016/j.physbeh.2007.03.020

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, *94*(2), 443–458. https://doi.org/10.1093/biomet/asm017

Bornovalova, M. A., Cashman-Rolls, A., O'Donnell, J. M., Ettinger, K., Richards, J. B., deWit, H., & Lejuez, C. W. (2009). Risk taking differences on a behavioral task as a function of potential reward/loss magnitude and individual differences in impulsivity and sensation seeking. *Pharmacology Biochemistry and Behavior*, *93*(3), 258–262. https://doi.org/10.1016/j.pbb.2008.10.023

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*(3), 253–262. https://doi.org/10.1037//1040-3590.14.3.253

Chandrakumar, D., Feuerriegel, D., Bode, S., Grech, M., & Keage, H. A. D. (2018). Event-Related Potentials in Relation to Risk-Taking: A Systematic Review. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00111

Dijkstra, N. F. S., Tiemeier, H., Figner, B. C., & Groenen, P. J. F. (2020). A censored mixture model for modeling risk taking. *ArXiv:2002.08146 [Stat]*. http://arxiv.org/abs/2002.08146

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 709.

Gelman, A., Meng, X.-L., & Stern, H. (1996). POSTERIOR PREDICTIVE ASSESSMENT OF MODEL FITNESS VIA REALIZED DISCREPANCIES. *Statistica Sinica*, *6*(4), 733–760. JSTOR.

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. JSTOR. https://doi.org/10.2307/1914185

Koscielniak, M., Rydzewska, K., & Sedek, G. (2016). Effects of Age and Initial Risk Perception on Balloon Analog Risk Task: The Mediating Role of Processing Speed and Need for Cognitive Closure. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00659

Li, X., Pan, Y., Fang, Z., Lei, H., Zhang, X., Shi, H., ... & Wan, Y. (2020). Test-retest reliability of brain responses to risk-taking during the balloon analogue risk task. *NeuroImage*, *209*, 116495.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. https://doi.org/10.1037//1076-898X.8.2.75

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of adolescence*, *26*(4), 475-479.

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, *53*(4), 222–230. https://doi.org/10.1016/j.jmp.2009.02.003

Park, H., Yang, J., Vassileva, J., & Ahn, W. Y. (2019). The Exponential Weight Updating Model: A Novel Computational Model for the Balloon Analogue Risk Task.

Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Experimental and Clinical Psychopharmacology*, *16*(6), 555–564. https://doi.org/10.1037/a0014245

Reddy, L. F., Lee, J., Davis, M. C., Altshuler, L., Glahn, D. C., Miklowitz, D. J., & Green, M. F. (2014). Impulsivity and Risk Taking in Bipolar Disorder and Schizophrenia. *Neuropsychopharmacology*, *39*(2), 456–463. https://doi.org/10.1038/npp.2013.218

Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences*, *15*(1), 11–19. https://doi.org/10.1016/j.tics.2010.10.002

Schürmann, O., Frey, R., & Pleskac, T. J. (2019). Mapping risk perceptions in dynamic risk‐taking environments. *Journal of Behavioral Decision Making*, *32*(1), 94–105. https://doi.org/10.1002/bdm.2098

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260. https://doi.org/10.1007/BF02289729

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, *315*(5811), 515–518. https://doi.org/10.1126/science.1134239

Trepel, C., Fox, C. R., & Poldrack, R. A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research*, *23*(1), 34–50. https://doi.org/10.1016/j.cogbrainres.2005.01.016

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384. https://doi.org/10.1037/a0032222

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*(2), 69–85. https://doi.org/10.1016/j.jmp.2012.02.005

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling Behavior in a Clinically Diagnostic Sequential Risk-Taking Task. *Psychological Review*, *112*(4), 862–880. https://doi.org/10.1037/0033-295X.112.4.862

Weichart, E. R., Turner, B. M., & Sederberg, P. B. (2020). A model of dynamic, within-trial conflict resolution for decision making. *Psychological Review*.

White, T. L., Lejuez, C. W., & de Wit, H. (2008). Test-Retest Characteristics of the Balloon Analogue Risk Task (BART). *Experimental and Clinical Psychopharmacology*, *16*(6), 565–570. https://doi.org/10.1037/a0014083

Young, M. E., & McCoy, A. W. (2019). Variations on the balloon analogue risk task: A censored regression analysis. *Behavior Research Methods*, *51*(6), 2509–2521. https://doi.org/10.3758/s13428-018-1094-8

APPENDIX

Table S1: Tested Models

Model	Free n parameters	Estimated Probability Function	Free γ parameters	Update n values	E(future)
1	1	Linear, Equation 6	1	Every Trial	yes
2	1	Linear, Equation 6	1	Every Trial	no
3	1	Linear, Equation 6	1	End of Balloon	yes
4	1	Linear, Equation 6	1	End of Balloon	no
5	1	Linear, Equation 6	2	Every Trial	yes
6	1	Linear, Equation 6	2	Every Trial	no
7	1	Linear, Equation 6	2	End of Balloon	yes
8	1	Linear, Equation 6	2	End of Balloon	no
9	1	Exponential, Equation 7	1	Every Trial	yes
10	1	Exponential, Equation 7	1	Every Trial	no
11	1	Exponential, Equation 7	1	End of Balloon	yes
12	1	Exponential, Equation 7	1	End of Balloon	no
13	1	Exponential, Equation 7	2	Every Trial	yes
14	1	Exponential, Equation 7	2	Every Trial	no
15	1	Exponential, Equation 7	2	End of Balloon	yes
16	1	Exponential, Equation 7	2	End of Balloon	no
17	3	Linear, Equation 6	1	Every Trial	yes
18	3	Linear, Equation 6	1	Every Trial	no
19	3	Linear, Equation 6	1	End of Balloon	yes
20	3	Linear, Equation 6	1	End of Balloon	no
21	3	Linear, Equation 6	2	Every Trial	yes
22	3	Linear, Equation 6	2	Every Trial	no
23	3	Linear, Equation 6	2	End of Balloon	yes
24	3	Linear, Equation 6	2	End of Balloon	no
25	3	Exponential, Equation 7	1	Every Trial	yes
26	3	Exponential, Equation 7	1	Every Trial	no
27	3	Exponential, Equation 7	1	End of Balloon	yes
28	3	Exponential, Equation 7	1	End of Balloon	no
29	3	Exponential, Equation 7	2	Every Trial	yes
30	3	Exponential, Equation 7	2	Every Trial	no
31	3	Exponential, Equation 7	2	End of Balloon	yes
32	3	Exponential, Equation 7	2	End of Balloon	no

Table S2: Choice-Only PLUM Pairwise Correlations

	X	Y	method	tail	n	r	CI95%	r2	adj_r2	z	p-unc	BF10	power
3	Metric	θ	pearson	two-sided	153	-0.621355	[-0.71, -0.51]	0.386082	0.377897	-0.727209	1.038267e-17	6.248e+14	1.000000
11	τ	n_0	pearson	two-sided	153	-0.480491	[-0.59, -0.35]	0.230872	0.220617	-0.523623	3.265579e-10	3.17e+07	0.999996
12	γ	θ	pearson	two-sided	153	0.351635	[0.2, 0.48]	0.123647	0.111962	0.367308	8.311811e-06	1892.239	0.994643
4	Metric	n_0	pearson	two-sided	153	0.346367	[0.2, 0.48]	0.119970	0.108237	0.361310	1.155966e-05	1385.076	0.993398
13	γ	n_0	pearson	two-sided	153	0.289969	[0.14, 0.43]	0.084082	0.071870	0.298532	2.771939e-04	70.428	0.956096
2	Metric	γ	pearson	two-sided	153	-0.199887	[-0.35, -0.04]	0.039955	0.027154	-0.202614	1.324043e-02	2.112	0.701499
10	τ	θ	pearson	two-sided	153	-0.168458	[-0.32, -0.01]	0.028378	0.015423	-0.170079	3.738614e-02	0.865	0.551333
9	τ	γ	pearson	two-sided	153	-0.161633	[-0.31, -0.0]	0.026125	0.013140	-0.163063	4.593033e-02	0.728	0.517088
5	α	τ	pearson	two-sided	153	0.155240	[-0.0, 0.31]	0.024099	0.011087	0.156505	5.535249e-02	0.623	0.484971
1	Metric	τ	pearson	two-sided	153	0.135391	[-0.02, 0.29]	0.018331	0.005242	0.136228	9.518599e-02	0.402	0.387210
14	θ	n_0	pearson	two-sided	153	0.127349	[-0.03, 0.28]	0.016218	0.003101	0.128044	1.167190e-01	0.342	0.349393
7	α	θ	pearson	two-sided	153	-0.077187	[-0.23, 0.08]	0.005958	-0.007296	-0.077341	3.429593e-01	0.158	0.157923
6	α	γ	pearson	two-sided	153	-0.062429	[-0.22, 0.1]	0.003897	-0.009384	-0.062511	4.433044e-01	0.135	0.119661
8	α	n_0	pearson	two-sided	153	0.052460	[-0.11, 0.21]	0.002752	-0.010545	0.052508	5.195647e-01	0.124	0.098725
0	Metric	α	pearson	two-sided	153	-0.027849	[-0.19, 0.13]	0.000776	-0.012547	-0.027856	7.325653e-01	0.107	0.063405

Table S3: RT-PLUM Pairwise Correlations

	X	Y	method	tail	n	r	CI95%	r2	adj_r2	z	p-unc	BF10	power
34	a	t_0	pearson	two-sided	153	-0.623327	[-0.71, -0.52]	0.388536	0.380384	-0.730428	7.649510e-18	8.421e+14	1.000000
3	Metric	θ	pearson	two-sided	153	-0.461826	[-0.58, -0.33]	0.213283	0.202794	-0.499630	1.869126e-09	5.878e+06	0.999985
4	Metric	n_0	pearson	two-sided	153	0.397290	[0.25, 0.52]	0.157839	0.146610	0.420427	3.680695e-07	3.673e+04	0.999323
19	β	w	pearson	two-sided	153	-0.380680	[-0.51, -0.24]	0.144917	0.133516	-0.400854	1.208927e-06	1.181e+04	0.998481
17	β	n_0	pearson	two-sided	153	0.347147	[0.2, 0.48]	0.120511	0.108785	0.362196	1.101286e-05	1450.002	0.993597
2	Metric	γ	pearson	two-sided	153	0.269119	[0.12, 0.41]	0.072425	0.060057	0.275913	7.687921e-04	27.442	0.923529
16	β	θ	pearson	two-sided	153	-0.254750	[-0.4, -0.1]	0.064898	0.052430	-0.260486	1.483737e-03	15.025	0.892457
35	w	t_0	pearson	two-sided	153	-0.251747	[-0.39, -0.1]	0.063376	0.050888	-0.257277	1.694585e-03	13.311	0.884983
11	α	n_0	pearson	two-sided	153	0.226537	[0.07, 0.37]	0.051319	0.038670	0.230536	4.865166e-03	5.131	0.808327
24	γ	w	pearson	two-sided	153	-0.196600	[-0.34, -0.04]	0.038652	0.025834	-0.199194	1.486339e-02	1.909	0.686763
6	Metric	w	pearson	two-sided	153	-0.196417	[-0.34, -0.04]	0.038580	0.025761	-0.199003	1.495885e-02	1.899	0.685932
23	γ	a	pearson	two-sided	153	0.192556	[0.03, 0.34]	0.037078	0.024239	0.194990	1.709705e-02	1.69	0.668237
18	β	a	pearson	two-sided	153	0.186673	[0.03, 0.34]	0.034847	0.021978	0.188888	2.086448e-02	1.422	0.640606
15	β	γ	pearson	two-sided	153	-0.185031	[-0.33, -0.03]	0.034236	0.021360	-0.187187	2.203620e-02	1.357	0.632765
22	γ	n_0	pearson	two-sided	153	0.167348	[0.01, 0.32]	0.028005	0.015046	0.168937	3.867649e-02	0.84	0.545779
28	θ	w	pearson	two-sided	153	0.155919	[-0.0, 0.31]	0.024311	0.011302	0.157201	5.428177e-02	0.633	0.488377
31	n_0	w	pearson	two-sided	153	-0.124836	[-0.28, 0.03]	0.015584	0.002458	-0.125490	1.241734e-01	0.326	0.337878
30	n_0	a	pearson	two-sided	153	-0.114044	[-0.27, 0.05]	0.013006	-0.000154	-0.114542	1.604196e-01	0.268	0.290369
20	β	t_0	pearson	two-sided	153	-0.109391	[-0.26, 0.05]	0.011966	-0.001207	-0.109831	1.782884e-01	0.248	0.270966
33	a	w	pearson	two-sided	153	-0.094820	[-0.25, 0.06]	0.008991	-0.004223	-0.095105	2.436698e-01	0.198	0.214986
25	γ	t_0	pearson	two-sided	153	-0.093252	[-0.25, 0.07]	0.008696	-0.004521	-0.093524	2.515875e-01	0.194	0.209424
5	Metric	a	pearson	two-sided	153	0.083687	[-0.08, 0.24]	0.007004	-0.006236	0.083883	3.037302e-01	0.171	0.177539
10	α	θ	pearson	two-sided	153	-0.081990	[-0.24, 0.08]	0.006722	-0.006521	-0.082174	3.136753e-01	0.167	0.172256
0	Metric	α	pearson	two-sided	153	0.077553	[-0.08, 0.23]	0.006014	-0.007239	0.077709	3.406665e-01	0.159	0.158984
29	θ	t_0	pearson	two-sided	153	0.054809	[-0.1, 0.21]	0.003004	-0.010289	0.054864	5.010153e-01	0.127	0.103305
13	α	w	pearson	two-sided	153	-0.047690	[-0.2, 0.11]	0.002274	-0.011029	-0.047727	5.582831e-01	0.12	0.090086
32	n_0	t_0	pearson	two-sided	153	0.041712	[-0.12, 0.2]	0.001740	-0.011570	0.041736	6.086953e-01	0.115	0.080492
9	α	γ	pearson	two-sided	153	0.040944	[-0.12, 0.2]	0.001676	-0.011635	0.040967	6.153086e-01	0.115	0.079359
1	Metric	β	pearson	two-sided	153	-0.033488	[-0.19, 0.13]	0.001121	-0.012197	-0.033501	6.811121e-01	0.11	0.069500
14	α	t_0	pearson	two-sided	153	0.027303	[-0.13, 0.19]	0.000745	-0.012578	0.027310	7.376131e-01	0.107	0.062877
21	γ	θ	pearson	two-sided	153	0.026445	[-0.13, 0.18]	0.000699	-0.012625	0.026451	7.455769e-01	0.107	0.062068
27	θ	a	pearson	two-sided	153	-0.014836	[-0.17, 0.14]	0.000220	-0.013110	-0.014837	8.555716e-01	0.103	0.053707
8	α	β	pearson	two-sided	153	0.014747	[-0.14, 0.17]	0.000217	-0.013113	0.014748	8.564268e-01	0.103	0.053661
26	θ	n_0	pearson	two-sided	153	-0.011801	[-0.17, 0.15]	0.000139	-0.013192	-0.011802	8.848828e-01	0.102	0.052305
12	α	a	pearson	two-sided	153	-0.004658	[-0.16, 0.15]	0.000022	-0.013311	-0.004658	9.544266e-01	0.101	0.050271
7	Metric	t_0	pearson	two-sided	153	-0.001204	[-0.16, 0.16]	0.000001	-0.013332	-0.001204	9.882168e-01	0.101	0.049921