Development of Novel Generative Models of Sequentially Dependent Decision Making in Risk-Ambiguous Environments.

Adam W. Fenton
Richfield, OH

BS, The Ohio State University, 2015
MA, University of Virginia, 2021

A Dissertation Presented to the Graduate Faculty of the University of Virginia

Department of Psychology
University of Virginia
September, 2023

Committee Members

Per B. Sederberg, Ph.D. (Chair)

Nicole M. Long, Ph.D.

James P. Morris, Ph.D.

Tanya M. Evans, Ph.D.

Abstract

In this dissertation, I present multiple generative cognitive models of sequentially dependent decision-making in risk-ambiguous environments. Cognitive models provide a means by which we can test theorized latent mechanisms that give rise to observed behavior. Utilizing variants of the Balloon Analogue Risk Task (BART), we investigate risk-related decision-making when the likelihood of success and failure is unknown to participants. We present a novel performance metric for the BART that compares participant performance to optimal performance, eliminating the issue of data censorship associated with traditional performance metrics. We evaluate a model of BART decision-making and use it to generate representations of subjective risk, reward, and value in order to relate these model-predicted time series with hemodynamic response via fMRI. We then investigate asymmetries in behavior in the contexts of risk taking versus risk avoidance.

Table Of Contents

1.1 INTRODUCTION

Decision-making is affected by a myriad of psychological and environmental 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. Self-report measurements may be skewed due to 1) the person self-evaluating believing an undesirable outcome could occur by reporting risky behaviors, and/or 2) a skewed perception of their own risk-related behaviors (Lejuez et al., 2002). Many tasks leave a need for an experimental analogue of risk-taking (Lejuez et al., 2002). It is often the case that we are forced to make choices when we are uncertain of the outcome, and doing so sub-optimally can have lasting repercussions, such as poor decision making financial or economical decisions (Grinblatt & Han, 2005). 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. The Balloon Analogue Risk Task (BART) was developed to emulate real-world risky decision making under uncertainty, making it an ideal experimental paradigm to build upon (Lejuez et al., 2002).

The BART is widely employed in academic and clinical settings to assess risk-taking behaviors. In the original task, participants see a virtual balloon on a computer screen. The balloon is labeled with either a point or monetary value, usually starting at zero. Participants are instructed to earn as many points or as much money as possible by incrementally pumping "air" into a series of balloons. At any point, the participant is able to stop pumping and collect the

balloon and add its value to a grand total/ bank, where the points/money are secure. However, if the balloon is pumped too much and it pops, the points/money accrued in that balloon vanish. When deciding whether to pump or collect, participants must balance the prospect of sequentially risky, incremental 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.

In this dissertation, I present three studies that altogether build, test, and utilize generative cognitive models of the BART in order to gain insight into the possible cognitive mechanisms at play during sequentially dependent risk-taking behavior and uncover individual differences in risky decision making. The objective of the first study is to develop a novel version of BART, performance metric, and underlying computational model that are more sensitive to individual differences than prior BART versions, measurements and models. The second study is an initial exploration into using a BART model to make continuous predictions of brain activity. The third study outlines the development of a new model that captures behavior of the BART and a new, mirror-image version of the task called the Balloon Risk Avoidance Task (BRAT). The BRAT can be conceptualized as the inverse of the BART, in which deflate balloons in order to prevent them from popping while still trying to earn as much money/points as possible. I have developed a model in which both BART and BRAT can be fit independently and simultaneously.

2. Model-based Analysis of a Novel Variant of the Balloon Analogue Risk Task Improves Sensitivity to Risk-taking Behavior

Authors: Adam W. Fenton[1], Per B. Sederberg[1]

Acknowledgements: This research was supported by Air Force Research Labs contract FA8650-16-1-6770. The authors do not note any conflicts of interest.

Summary

In this chapter, I present a study in which we developed a new version of the BART with the purpose of making a fast yet sensitive paradigm for assessing risk taking behavior. We also explaining the shortcomings of traditional methods for assessing BART performance and present a new metric based upon deviance from optimal performance. We present a the development process of a new generative cognitive model of the BART called PLUM (Predictive Linear Utility Model), then demonstrate its ability to fit and recreate observed BART behaviors.

2.1 Introduction

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. In the task, participants must pump up a balloon untilParticipants have no prior knowledge of the likelihood of the balloons' popping points and must learn the balloons' popping probabilities through trial

[1] Department of Psychology, University of Virginia

and error. The BART has excellent test-retest reliability both behaviorally and neurally (White et al., 2008; Li et al., 2020).

BART performance has been closely correlated with various self-report measures of risk-taking behaviors in real-world settings, including greater tendencies to engage in 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, such as the administration of oxytocin to patients with bulimia nervosa and binge eating disorder engaging in less risky decision making compared to placebos (Reddy et al., 2014; Leslie et al., 2019). Most healthy participants demonstrate risk-averse behavior on the BART (Trepel et al., 2005). In the BART, people with lower self-reported impulsivity are more risk-averse when potential rewards are higher due to the potential loss growing in step while people with higher impulsivity are less affected by the increasing levels of risk (Bornovalova et al., 2009).

However, there are multiple drawbacks with the 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 popping. This vast range of popping points makes it very difficult for participants to learn the probability of the balloon popping (de Groot, 2020) and increases task completion time. A second issue lies in the method by which BART performance is assessed. The performance metric used in various BART studies - the adjusted score - is calculated by taking taking the average number of pumps across only the balloons that were collected (Lejuez et al., 2003). This metric discards the information in popped balloons, yet it is still in use today (Plescak et al., 2008; Chandrakumar et al., 2018). Removing popped trials can bias the adjusted score lower due to trials in which risks were taken by pumping the balloon more are censored (Chandrakumar et al., 2018).

Here, we present a new version of the BART that is aimed at quantifying differences between individuals across the entire spectrum of risk processing and examining the sequential dependencies of the decision-making process. In order to do so, we have made modifications to the tasks reward and balloon pop probability structures. BART that includes variable rewards and balloon categories with more distinctive pop probability distributions that induce meaningful choices on at each decision point, 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 and learn multiple distinctive probability distributions while minimizing task length, 3) Develop a new methodology for assessing BART performance that uses all data, and 4) Compare several possible cognitive mechanisms that could be at play during the decision-making process.

2.2 METHODS AND MATERIALS

2.2.1 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.

2.2.2 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 Weichart et al., 2020). 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.

2.2.3 Task Design

Participants read the BART task description and instructions and then perform a brief practice session in which three gray-colored balloons were sequentially presented. These 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 participant 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 to make the task more engaging..

Figure 1. Image of the BART task.

There were a total of 18 balloons in each of the two blocks of the BART performed by participants. 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]. These distributions are similar to those from a prior BART study, with the difference being the [8, 16] range balloon type (Schonberg et al., 2012). 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 and were based off similar distributions used by Schonberg and colleagues (2012). The outcomes of the first balloons can significantly impact a participant'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 in a bag is likely to only affect the participant's interpretation of that bag's pop distribution. The presence of three different bags also encourages the participant to learn and differentiate those distributions. The [8,16] balloons provide an opportunity for the participant 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 to explore these distributions, indicating that they are willing to take risks, at least to a certain extent. participants 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.

2.3. ANALYSIS

2.3.1 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 collected balloons (Lejuez et al., 2002). The adjusted score is useful for detecting large-scale differences between healthy and unhealthy participants (Lejuez et al., 2003, Reddy et al. 2014). However, we have found that there is room to enhance the sensitivity of this task to be able to differentiate smaller differences within an individual over time or between cognitively similar people. We turn to model-based approaches in order to gain a deeper understanding of the individual choice-level decision-making process within the context of the BART. However, there are three limitations of the adjusted score. Firstly, all information from the popped balloons is ignored (Pleskac et al., 2008; Chandrakumar et al., 2018; Young & McCoy, 2019). There is still useful information about behavior in popped balloons, especially if the participant was purposefully engaged in exploratory behavior. 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). 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 participant 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 decision point 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 choice, it is informative that the participant made the attempt to pump in light of past experience. Third, In response, they developed a BART variant in which a participant 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., 2008; 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 participant 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 participant experienced in the task. This E value is calculated from the perspective of an omniscient participant 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 participant and O for every choice. 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.

2.3.2 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 participant compares the potential gain to the potential loss. Since humans have various biases, the participant's perception of value for 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 participant's personal interpretation of value as the *expected value*, E. In this study, we present several possible mechanisms by which E may be modulated.

Estimating the Probability of a Balloon Popping

One of the mechanisms governing the value of E is the participant's estimation of the likelihood of the balloon popping on a given decision. An assumption of this model - and many other prospect theory models in general - is that the participant possesses an initial estimate of this outcome likelihood. This initial belief is subjective and be influenced by extraneous factors such as stress (Molins et al., 2022). In the case of the BART, that likelihood is based upon the participant's estimate of how many times they can pump the balloon until it pops. We call this estimated popping point n. A participant could assume that the pop probability, p, grows in an

exponential fashion (as it truly does) or in a linear fashion. For a participant that assumes the former, the popping point *p* at choice *i* is defined as:

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

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

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

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 participant considers only the current rewards, the subjective value of pumping on at choice/decision point *i* is:

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

where *q* is 1-*p* (the probability of the balloon not popping if the participant chooses to pump), *r* is the amount of money that would be added to the balloon if the participant chooses to pump, and *b* denotes the current monetary value of the balloon. θ is a free parameter for loss aversion. γ is a free parameter modulating nonlinear perceptions of monetary value. When γ is higher, the

difference between monetary values increases exponentially. 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 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 participant 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:

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

E(future) is the summation of the mean reward $\bar{r}$ seen thus far multiplied by 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 participant must decide whether to pump the balloon or to collect the balloon. Our models posit that the participant compares the potential gains against the

potential losses. By this reasoning, the participant 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).

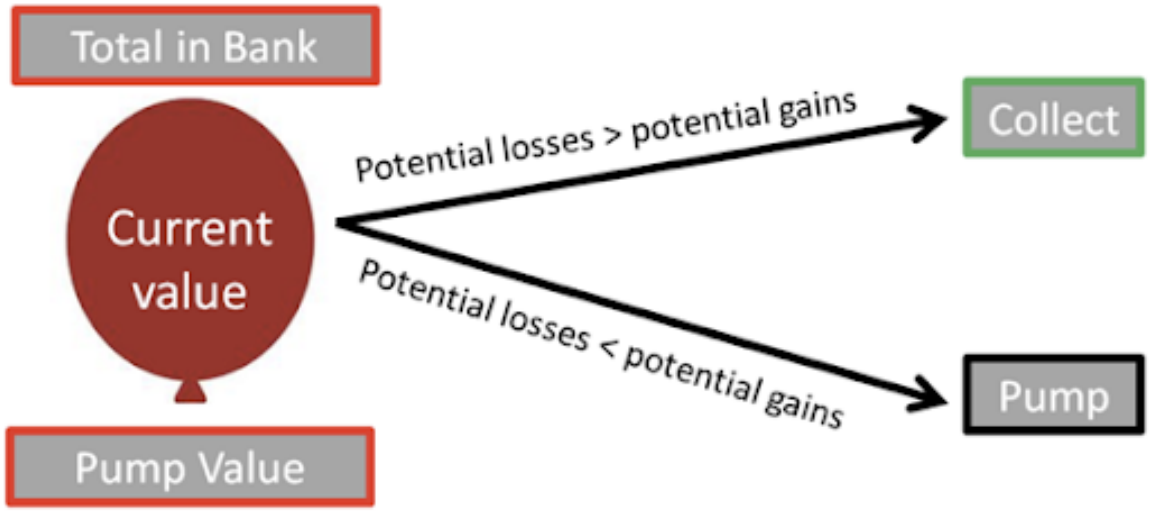


Figure 2: Prospective Gains v Losses. Prospect theory in relation to the BART suggests that when perceived potential losses are greater than perceived potential, the optimal decision is to stop pumping and collect the balloon. When perceived potential gains are greater than perceived potential losses, the participant should pump the balloon.

Choice Probability Function

The mechanism by which the likelihood of the participant's choice *i* is determined by the following sigmoid equation:

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

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

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

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 ranges. 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 participant 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 at decision point/choice *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 participant's perspective, pump decisions in which the balloon does not pop may slightly reinforce their confidence in their estimate of the likelihood of the balloon popping. How different participants 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 choice to variants in which n_h is updated only at the conclusion of a balloon. When n_h is updated on every choice, it slowly increases over the course of sequential pumps, causing choices that do not result in a pop to provide slight evidence/confidence for pumping again.

3.2.5 Modeling Sequentially Dependent Choices

Generative cognitive models provide an ideal test-bed for assessing the validity of hypothesized cognitive processes. Like other cognitive models, generative models 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 participant'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 participant 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/trial number, *N* is the number of balloons, *i* is the choice number, and Ω is the choice at which point the last decision was made. Table 1 presents descriptions of each parameter. Parameter priors were selected to be uninformatively wide.

Parameter	Description	Prior
α	Learning rate	$logit^{-1}(\mathcal{N}(0.0, 1.4)$
γ	Shape of reward curve	$logit^{-1}(\mathcal{N}(0.5, 0.35) * 2)$
θ	Loss aversion	$\mathcal{N}(0.0, 5.0)$
n_0	Starting estimated popping point for each bag	$logit^{-1}(\mathcal{N}(-0.71, 0.80))$
τ	Choice randomness	$\exp^{\mathcal{N}(1.0,1.0)}$

Table 1: PLUM Model Parameters

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

2.3.3 Fitting Models

All models were fit to each subject and block independently. We define a set of priors for each parameter (Table 1). 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 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.

2.4 RESULTS

We first examine the behavioral results and evaluate the utility of our proposed BART performance metric *M*. We then perform a model comparison to investigate which of the proposed cognitive mechanisms are the most likely to give rise to the observed behavior.

2.4.1 Behavioral Results

Performance on the BART task was on average below optimal performance (first block: average $M = 0.46$; second block: $M = 0.45$) indicating that participants were likely to behave conservatively in that they were, on average, not reaching the optimal stopping point on the balloons. Figure 3 shows he distributions of performance for each block were very similar. 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 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 that had higher upper pump limits 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 cautions of pumping too many times, indicating very conservative behavior even when the opportunity for greater exploration presents itself.

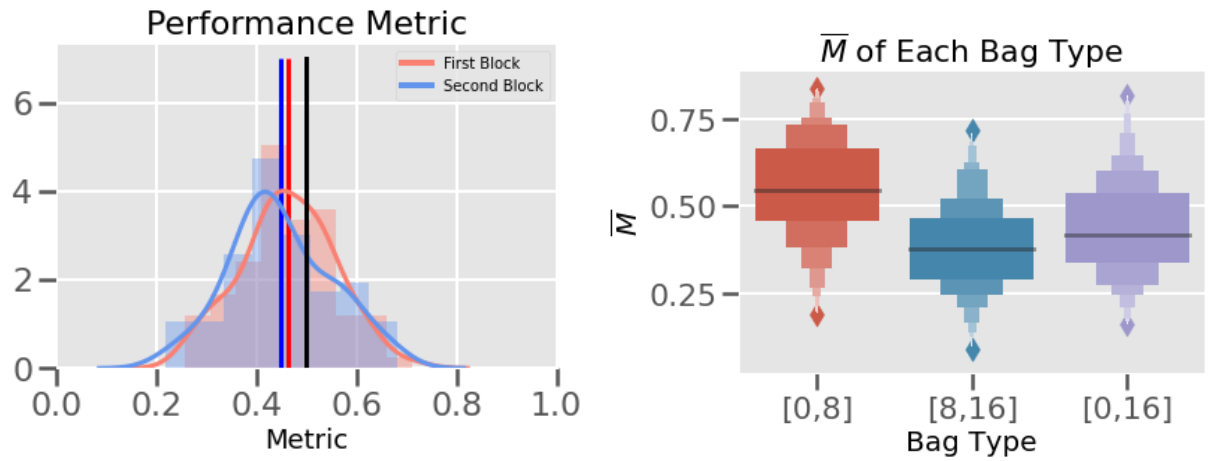


Figure 3: Performance Metric: The left plot shows participants (N=84) scored on average below optimal performance of 0.5 (the black vertical line). First block mean: 0.46 (red vertical line); Second block mean: 0.45 (blue vertical line). The right plot shows the average performance for each balloon type. We see that, on average, participants tok greater risks on the smallest range balloon [0,8] and were conservative on both the [0,16] and [8,16] balloon types on average.

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 participant's specific trial-level context.

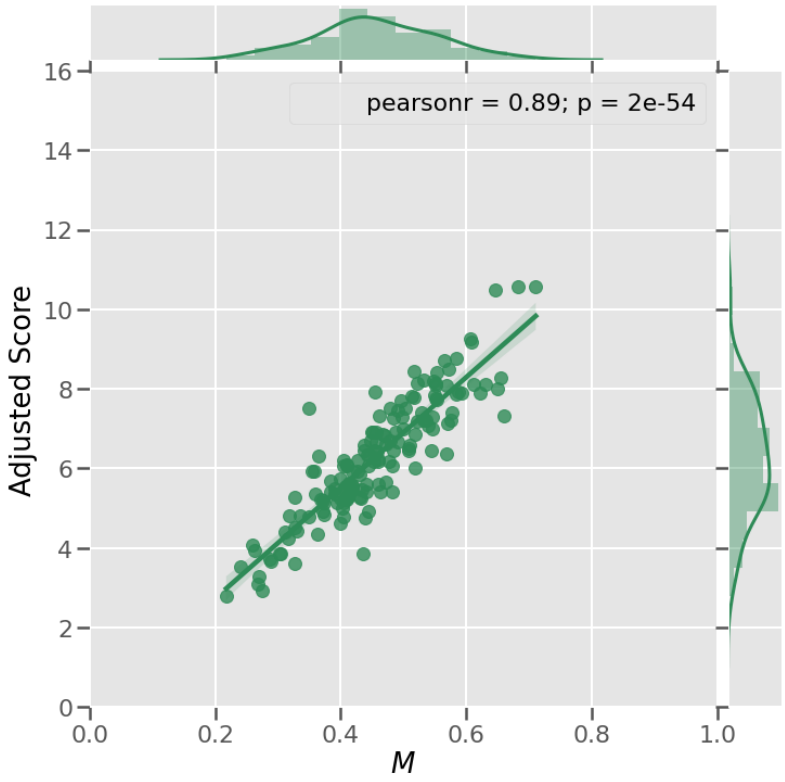


Figure 4: Adjusted Score vs. Deviation from Optimal (M) Across All Blocks. (N=163) The black vertical line represents optimal performance ($M = 0.5$). *M* correlates strongly with adjusted score, indicating that it would be useful in diagnostic studies that utilize adjusted score whilst being more sensitive to true performance.

2.4.2 Modeling Choices: Model Results and Comparisons

2.4.2.1 Model BPIC Comparison

After fitting each model to all subjects within each block independently, 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.

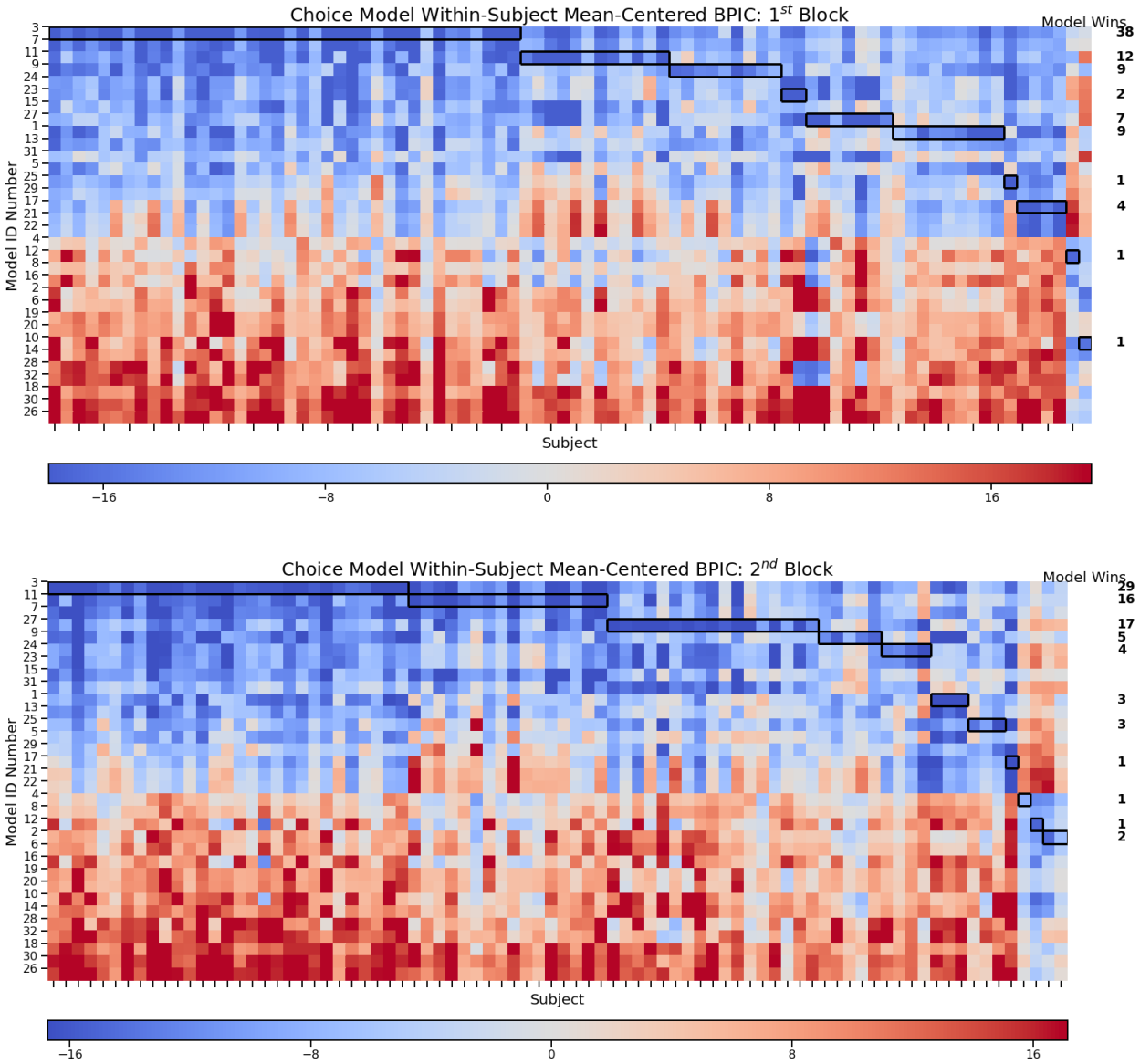


Figure 5: Comparing Model Fits. Within-subject mean-centered BPIC scores are presented. The models are arranged in descending order of model average BPIC across participants. Lower BPIC values (blue) represent a better fit. For a given model, any participants that it fit to better than the other models 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.

Model 3 was the top performing model. We refer to model 3 as the Predictive Linear Utility Model (PLUM). The assumptions of the PLUM model are as follows:

1. The participant 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 participant is taking into account the prospect of future gains on consecutive pumps (equations 9 & 10).
4. The participant updates n_h only at the conclusion of every balloon.

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

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

$$E(\text{pump})_i = qr_i^{\gamma} + \sum_{k=i+1}^{n_h} q_k \overline{r}^{\gamma} - \theta p b^{\gamma} \quad (18)$$

Figure 6 below puts equation 16 into the experimental context.

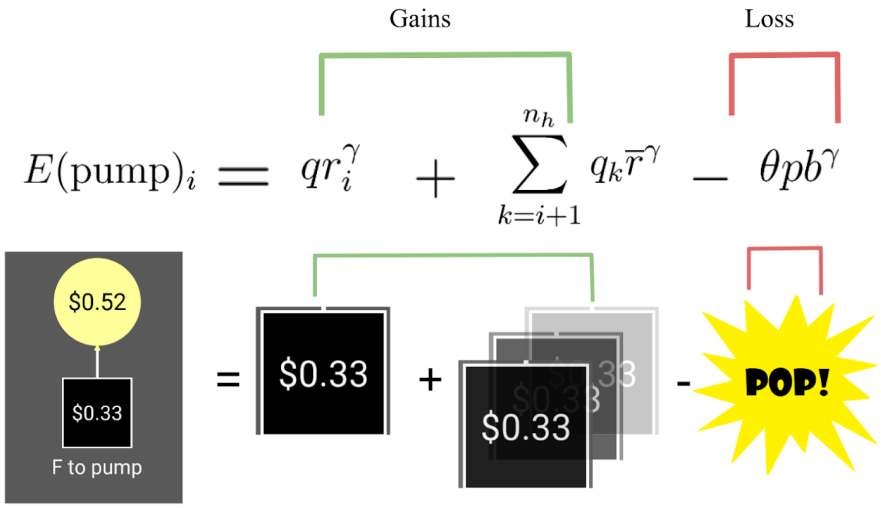


Figure 6: Representing Expected value in PLUM.

2.4.2.3: 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*. The correlations of the simulated average and true observed *M* ($r=0.98$, $p<0.001$) for both blocks are very strong (Figure 7), suggesting that the model is capturing the behavior of the posited cognitive mechanisms well.

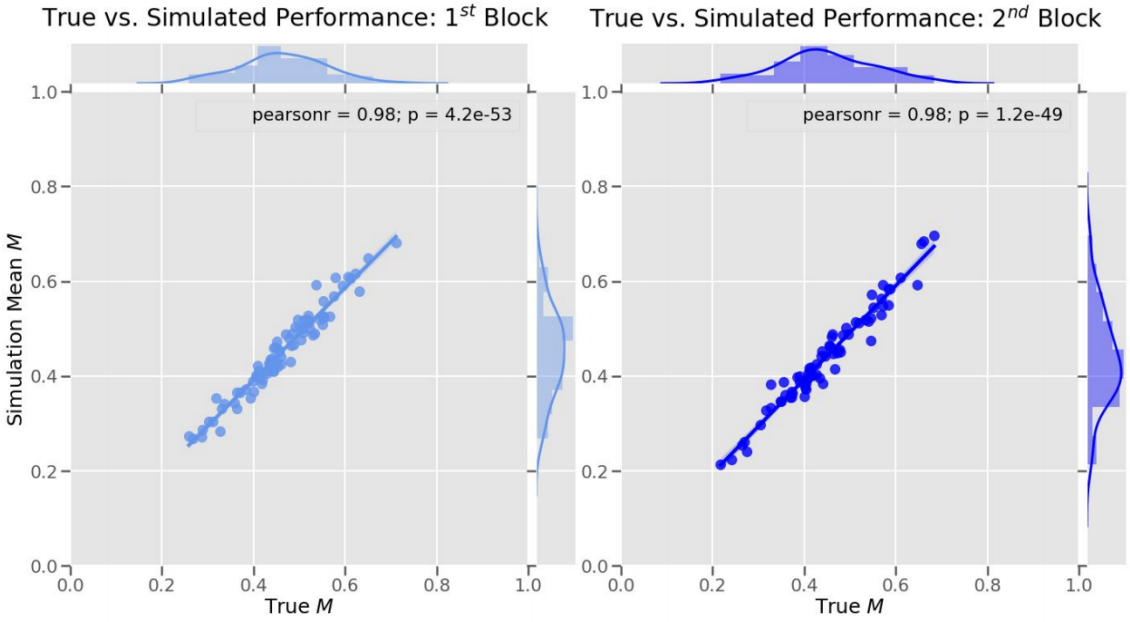


Figure 7: True vs. Simulated Performance for PLUM: Simulations using best-fitting parameters were robustly accurate in recreating observed patterns of pumping and collecting.

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

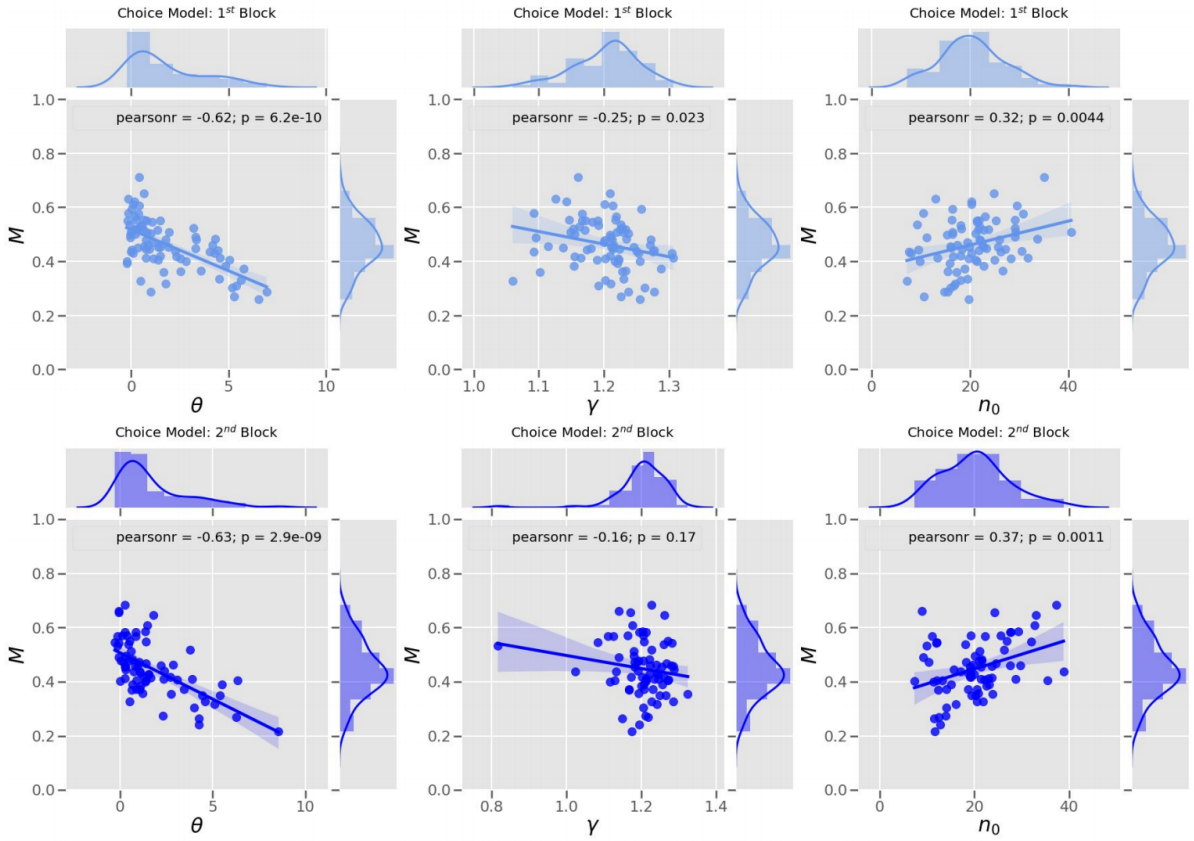


Figure 8: Parameters predicting performance: The parameters θ, γ (for the first block), and n_0 are significantly correlated with *M*. See table S2 in the Appendix for all pairwise comparisons.

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.

A a posterior predictive distribution (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 9 shows the PPDs for one subject fit with 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.

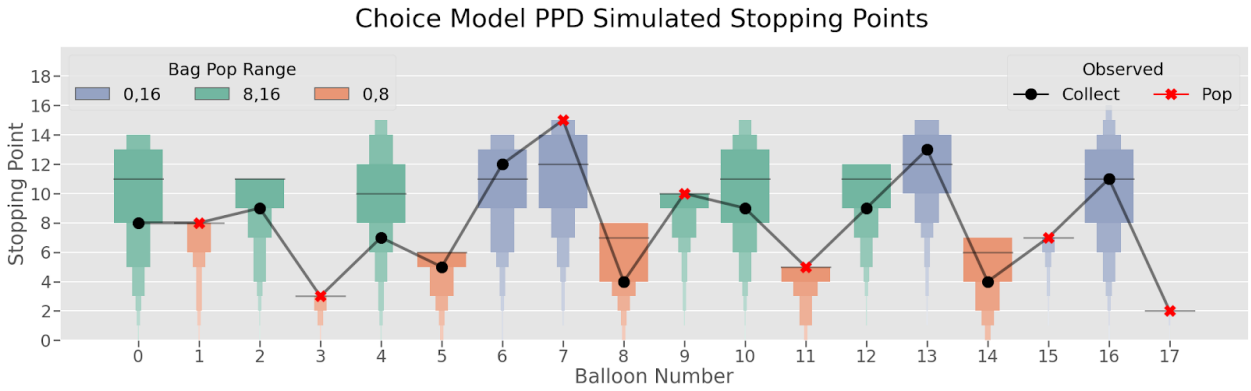


Figure 9: PPD Balloon-Level Performance. For this subject, we are able to see the distribution of points at which the simulations either popped or collected the balloon (box plots). We can compare this to the participant's pop and collection points (black dots for collection points, red x's for the popping points).

2.5 Model comparisons: Comparing PLUM to other modern models of BART decision-making.

Comparing different computational models of the BART is critical for refining model development and improving our understanding of the possible cognitive mechanisms at play during risk-taking. Majority of the BART models are based in prospect theory, stating that participants evaluate the expected value when deciding whether or not to pump a balloon or stop and collect. Thus, many BART models share similar model mechanisms and free parameters such as loss aversion. Given that we have a novel BART variant that is more complex than many

other versions of BART, comparing PLUM to other models not based upon the same prospect theory framing would be useful for future model development.

Our new BART version is quite complex in that it contains multiple balloon probability structures with staggered probability distributions (pump ranges of 0-8, 0-16, and 8-16) and is shorter than most BART tasks, thus speeding up task time. Our variant has only 18 balloons (6 balloons from each category) instead of the more common 30 balloons, with the largest balloon types having a maximum pump upper limit of only 16 pumps as opposed to the more common 128 pump limit. The presence of three different types of balloons encourages the participant to differentiate them and learn their underlying pop probability distributions in order to maximize earnings. The BART variant presented here also has random rewards for pumping the balloon, which could drastically impact the performance of models that do not properly account for this. PLUM was created with multiple balloon types and variable rewards in mind. Comparing PLUM to other BART models will allow us to explore multiple cognitive mechanisms and determine which of these mechanisms are most likely at play in a participant's decision-making process.

2.5.1 Competing Models Descriptions

In recent years, two other models of BART performance have been developed to evaluate individual differences in decision-making during the BART. In this section, we present each of these two models, discussing their proposed cognitive mechanisms and how they compare to one another.

2.5.1.1 The Exponential-Weight Mean-Variance Model

The exponential-weight mean-variance model (EWMV) is based upon the notion that the subjective utility of pumping a balloon is determined by the linear combination of the variance and expected value of potential outcomes (Park et al., 2021). The EWMV model is similar to the

BART model by Wallesten and colleagues presented in 2005 (Park et al., 2021; Wallesten et al., 2005), which is also the inspiration for PLUM as well, making this comparison very applicable. The EWMV model's central mechanism is based upon mean-variance analysis, which can be used to mimic prospect theory (see Park et al, 2021; Boorsman & Sallet, 2009). A critical distinction from PLUM is that the EWMV operates under the assumption that the participant's estimation of the pop probability remains static during the pumping of a balloon (Park et al., 2021). The participant instead begins the task with a prior belief of the pop probability, ψ. At the end of each balloon, the participant compares the number of successful pumps to the total number of pumps made in order to assess the observed probability of popping:

$$P_{k-1} = \frac{\sum_{i=0}^{k-1} \left(n_i^{\text{pumps}} - n_i^{\text{success}}\right)}{\sum_{i=0}^{k-1} n_i^{\text{pumps}}}, \quad (19)$$

and updates their estimation of the pop probability, p_k^{burst}, by taking into consideration ψ. The following equation shows how this pop probability estimate is calculated.

$$p_k^{\text{burst}} = e^{-\xi \sum_{i=0}^{k-1} n_i^{\text{pumps}}} \psi + \left(1 - e^{-\xi \sum_{i=0}^{k-1} n_i^{\text{pumps}}}\right) P_{k-1}$$
$$\text{with } 0 < \psi < 1, \xi > 0. \quad (20)$$

The free parameter ξ is an updating exponent that determines the degree to which observations are factored into assessing p_k^{burst}. The subjective utility for pumping on trial *k* (aka balloon number) and pump number, *l*, is given by:

$$U_{kl}^{\text{pump}} = \left(1 - p_k^{\text{burst}}\right) r - p_k^{\text{burst}} \lambda(l-1)r + \rho p_k^{\text{burst}} \left(1 - p_k^{\text{burst}}\right) \{r + \lambda(l-1)r\}^2 \text{ with } \lambda > 0, \quad (21)$$

$$U_{kl}^{transfer} = 0 \quad (22)$$

The value *r* is the reward for pumping the balloon. λ is the participant's loss aversion. The parameter ρ is the participant's risk preference. Values of ρ less than 0 indicate the participant

prefers choices with a larger variance in potential outcomes. A ρ of 0 indicates that the expected value is the determining factor in assessing the subject utility. A ρ greater indicates the participant prefers to choose options with smaller outcome variance. The likelihood of the participant pumping on a given trial is determined by the following softmax:

$$p_{kl}^{\text{pump}} = \frac{1}{1 + e^{\tau\left(U_{kl}^{\text{trunser}} - U_{kl}^{\text{pump}}\right)}} \text{ with } \tau \geq 0 \quad . (23)$$

τ is an inverse temperature parameter.

2.5.1.2 The Scaled Target Learning Model

A second modern BART model is the Scaled Target Learning model (STL) developed by Zhou, Myung, and Pitt (2021). Notably, this model does not utilize prospect theory, making it a fitting model to compare with the PLUM and EWMV models which both are based in prospect theory. Indeed, this model is based upon a target model of the BART. Target models operate by assuming the participant is updating a subjective ideal number of pumps over the course of the BART (Zhou et al., 2021). After a successful balloon collection, the ideal stopping point for pumping is increased. After a balloon pop, the stopping point decreases. This target number of pumps is denoted as ω_k, and is initially a free parameter, similar to how PLUM has the initially free parameter n. After the initial balloon, ω_k is updated in one of two ways depending on the balloon's outcome:

$\omega_k = \omega_{k-1} \times \left(1 + v \text{ win} \cdot \frac{\text{npump }_{k-1}}{nmax}\right)$, (24) if the participant collects, and

$\omega_k = \omega_{k-1} \times \left(1 - \text{ vloss } \cdot \left(1 - \frac{npump_{k-1}}{n \max}\right)\right)$, (25) if the balloon pops.

$v$win and $v$loss are free parameters that are greater than 0. $nmax$ represents the task design's maximum pop point. The STL utilizes w_k to ascertain the likelihood that the participant will pump at a given choice point l via the following softmax equation:

$$p_{kl}^{\text{pump}} = \frac{1}{1 + e^{\beta(l-\omega_k)}} \quad (\beta \geq 0), \text{(26)}$$

β represents choice randomness, or "behavioral consistency" (Zhou et al., 2021).

The presence of the experimental design parameter $nmax$ in the STL differentiates it from the the other two models. In the BART used in this study, participants have no knowledge of the balloons' maximum pump ranges. To address this, we include an additional model called STLnmax that makes $nmax$ a free parameter such that $nmax > 1$.

2.5.2 Comparing Model Mechanisms

A critical distinction separating PLUM from the EWMV model is that PLUM assumes the participant understands that every pump increases the likelihood that the balloon is going to pop. In the EWMV model, the participant is assumed to have a static representation of the pop probability while pumping the balloon (Park et al., 2021). Notable work with the BART by Schürmann and colleagues provides evidence that BART participants do *not* have static internal estimations of pop probability as a function of pump. Self-report measures show individuals perceive the balloon's pop probability to increase with each consecutive pump and that the subjective pop probabilities can drastically differ from the true, objective pop probability distribution (Schürmann et al., 2019). The STL model implements this increasing likelihood of the balloon popping indirectly via its choice likelihood softmax equation, which compares the distance of the current choice, l, from the target number of pumps, w_k.

As aforementioned in chapter 2, PLUM makes explicit use of the possible current and future rewards in its utility function for calculating expected value at each pump opportunity. It is able to account for variable rewards, as it tasks into consideration the current and future rewards in it's calculator of expected value (eq 9). The STL model, however, ignores rewards entirely, suggesting that the participant is strictly focusing on ascertaining ideal target numbers of pumps. If the STL model performs better than the PLUM and EWMV models, it may suggest that participants are completely ignoring rewards themselves. Given the design of this new version of BART, I hypothesized PLUM would perform best, followed by the EWMV model and then the STL model.

2.5.3 Model Fitting, Comparison Methods, & Results

The EWMV, STL, and STL_{nmax} models were fit in an identical manner as PLUM. Table S3 in the appendix provides the priors selected for the free parameters in these models. We again utilize the BPIC score for comparing model goodness of fit to the observed behavioral data.

Model Comparison Results: BPIC Scores & Simulations

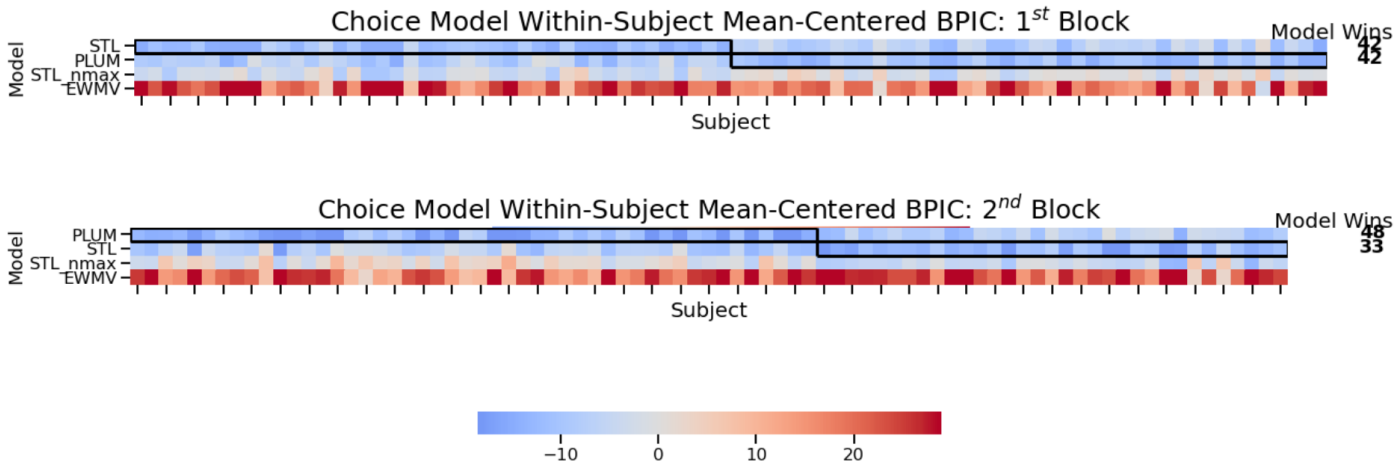


Figure 10: Comparison of PLUM, STL, STL_{nmax}, and EWMV: Models are ordered vertically based upon average within-participant centered BPIC score across participants. To the right are the counts of the number of participants that a given model fit the best out of all other models.

STL consistently outperformed the EWMV model, contrary to our hypothesis that it would outperform the STL. The STL model and PLUM were very close in terms of BPIC performance. For the first block, the STL model had the lowest BPIC score on average, with PLUM being second, STL_{nmax} being third, and EWMV being last. However, STL and PLUM BPIC scores were not significantly different from one another (t = 0.39, $p = 0.70$), and PLUM BPIC scores were significantly lower than STL_{nmax} (t =-12.72, $p < 0.001$) STL and PLUM both fit 42 participants the better than the other models. In the second block, PLUM had the lowest BPIC score overall, followed by STL, STL_{nmax}, and EWMV models. PLUM also fit the best to 48 participants, with STL fitting the remaining 33 the best. PLUM scores were not significantly different than STL scores on the second block (t = -1.99, $p = 0.0505$). PLUM BPIC was significantly lower than STL_{nmax} in the second block (t = -10.32, $p < 0.001$). The addition of the free parameter in n_{max} in the STL_{nmax} model, however, reduced the goodness of fit two third place in this comparison, due to BPIC penalizing for more parameters. This distinction is critical, as it means that when the STL model takes into account that participants do not know the upper limit of the pump range of the balloons, the model's performance suffers.

We assessed each model's ability to recreate the observed behavior using best-fitting parameters as we did with PLUM. Each model was simulated 1,000 times, with each simulation being scored using performance metric *M*. Average simulation performance was then compared to the participant's behavior (Figure 11). All models show good simulation accuracy. PLUM-generated performance correlated the strongest with observed behavior (1st block: $r = 0.98$; p < 0.001 , 2nd block: $r = 0.98$, p < 0.001), followed by STLnmax (1st block:

$r = 0.96$; p < 0.001 , 2nd block: $r = 0.97$, p < 0.001), STL (1st block: $r = 0.95$; p < 0.001 , 2nd block: $r = 0.95$, p < 0.001), and EWMV (1st block: $r = 0.92$; p < 0.001 , 2nd block: $r = 0.92$, p < 0.001).

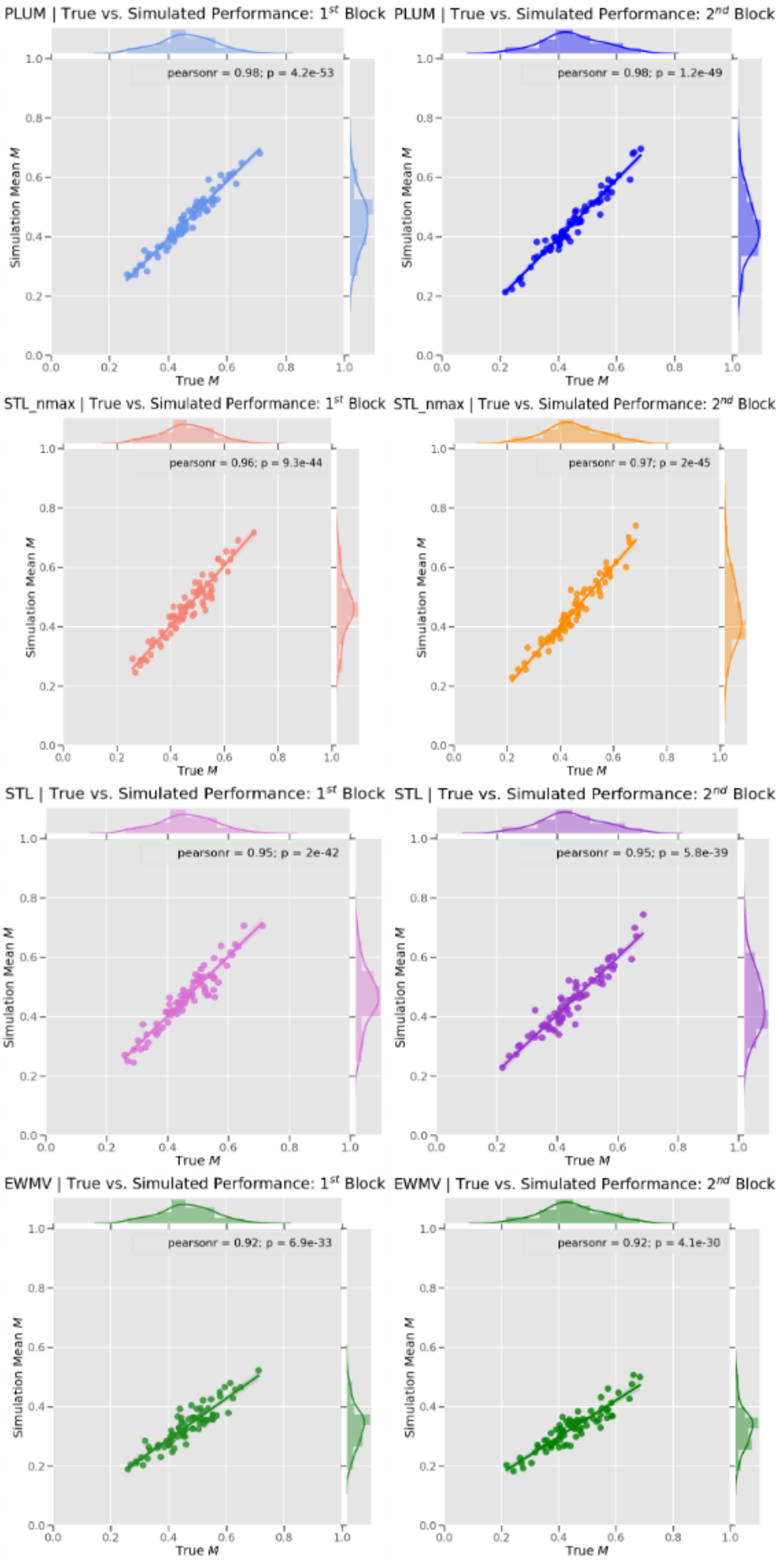

PLUM | True vs. Simulated Performance: 1st Block
pearsonr = 0.98; p = 4.2e-53
PLUM | True vs. Simulated Performance: 2nd Block
pearsonr = 0.98; p = 1.2e-49
STL_nmax | True vs. Simulated Performance: 1st Block
pearsonr = 0.96; p = 9.3e-44
STL_nmax | True vs. Simulated Performance: 2nd Block
pearsonr = 0.97; p = 2e-45
STL | True vs. Simulated Performance: 1st Block
pearsonr = 0.95; p = 2e-42
STL | True vs. Simulated Performance: 2nd Block
pearsonr = 0.95; p = 5.8e-39
EWMV | True vs. Simulated Performance: 1st Block
pearsonr = 0.92; p = 6.9e-33
EWMV | True vs. Simulated Performance: 2nd Block
pearsonr = 0.92; p = 4.1e-30
Simulation Mean M
True M

Figure 11: Simulation Performance of Models. The left column shows model performance in the first block, while the right column shows simulation performance in the second block. Models are arranged in descending order of best two worst simulation accuracy.

2.6 DISCUSSION

Our new version of the BART puts a participant in an ambiguous risk-taking environment that encourages learning the likelihood of failure of different categories of balloons. The performance metric *M* provides more insight than the adjusted score into a participant's actual level of risk taking performance by comparing their performance to that of an optimal participant. Critically, this metric does not censor any data and considers the trial-by-trial sequence of choices. We hypothesized several possible cognitive mechanisms that could be involved in the BART. We identify PLUM as the winning model. This model assumes that the participant 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 participant is not merely focusing on just the immediate possible gains, but also considering future gains. Reinforcing this is the fact that the top three performing models contain the mechanism for considering future rewards (eq 10). PLUM also assumes the participant 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 participants are 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). Despite PLUM being the most complex model out of those compared, it was tied in BPIC scores with the STL that contained information outside of the participants' knowledge and fit better than the STL_{nmax} model. PLUM also demonstrated the greatest simulation accuracy.

3. Relating neural activity to model-predicted values of risk and reward in the Balloon Analogue Risk Task

Authors: Adam W. Fenton[1], Rebecca Waugh[1], Per B. Sederberg[1]

Acknowledgments: This research was supported by Air Force Research Labs contract FA8650-16-1-6770. The authors do not note any conflicts of interest.

Abstract

In this chapter, I conduct an exploratory analysis in which I compare model-predicted values of risk and value to their objective counterparts. I then assess if PLUM's predictions of expected value are correlated with brain activity. For this analysis, I fit the PLUM model to behavioral data. I then use those best fitting parameters to generate model-based predictions of participants' perceptions of risk and value as a timeseries, correlating these with fMRI recorded brain activity.

PLUM is based in prospect theory, and it evaluates the expected value of a risky choice (Kahneman & Tversky, 1979). This expected value (E) is a comparison between potential gains and losses, each of which are a product of their point/monetary value and the likelihood of that event occurring. It is likely that PLUM has a defined mechanisms that captures the participant's projection of attaining future consecutive pumps via expected future rewards and factoring this in during deliberation. This mechanism is useful for modeling choices in which a reward is higher than average but likelihood of current and future pump success is low (especially towards the balloon's estimated popping point), as well as when the reward for pumping is low but the estimated risk of current and immediate future pops is also low (often near the beginning of a balloon).

3.1 Introduction

The decision-making process when taking risks is profoundly influenced by the participant's interpretation of outcome likelihoods. Ambiguity in the likelihood of known outcomes or being uncertain of the what the outcome actually entails can impact an individual's behavior. It has long been known that people behave quite differently when the odds of success/failure are ambiguous. Often in ambiguous circumstances people will tend to perform overly conservatively and thus far from optimality (Ellsberg, 1961; Camerer & Weber, 1992; Trepel et al., 2005; Levy et al., 2009; Wu et al., 2021; Gärling et al., 2009). It is worth investigating the distinction between decision-making in risky environments (situations in which the probability of success is a known value) and ambiguous environments (situations in which the probability of success is unknown) because ambiguous risk-taking is a common occurrence in our lives. The majority of research investigating risk-taking behaviors with fMRI examines the differences in decision-making between conditions of varying degrees of known risk (see Wu et al., 2021 for review). The human brain likely utilizes much of the same brain regions in formulating a course of action between known risk and ambiguous risk conditions, with evidence to suggest that there are distinct differences as well (Levy et al., 2009; Wu et al., 2021). Studies examining the differences between risky conditions and ambiguous conditions have shown increased orbitofrontal cortex (OFC) (Hsu et al., 2005), dorsal medial prefrontal cortex (dmPFC), anterior insula, posterior parietal cortex, amygdala, and inferior frontal gyrus (Huettel et al., 2005) in ambiguous conditions compared to conditions in which risk is more explicit (Wu et al., 2021; Levy et al., 2009).

Functional imaging studies with the BART have shown profound effects of risk-taking behaviors in hemodynamic responses. Broadly, the mesolimbic-frontal pathway, parietal, subcortical, ventral striatum, and prefrontal brain regions all play a role in risk-ambiguous

decision-making (Preuschoff et al., 2006; Rao et al., 2008; Li et al., 2020). The ventral striatum has been shown to have greater activity in risk-averse people compared to risk-seekers in anticipation of high-risk gambles (Rudorf et al., 2012). The dorsal lateral prefrontal cortex (dlPFC) has been linked with active risk-taking as opposed to passive observation of risk-taking in the BART (Rao et al., 2008; Aupperle et al., 2015). Schonberg and colleagues have shown that during the sequential BART, the anterior insula, dlPFC, and dorsal anterior cingulate cortex (dACC) activity increased the more times participants pumped a given balloon (Schonberg et al., 2012). These results are further supported by later findings in which ACC and dlPFC activity increased with the amount of risk in the BART (Li et al., 2020). These results suggest that the participant is paying attention to and considering the changes in expected value at each choice point. A more recent study examining the effects of immediate vs. delayed decision outcome presentation when taking risks demonstrated a distinctive network made up of the dlPFC and dACC associated with prediction-error monitoring when taking risks under ambiguity (Fiore and Gu, 2021). The Schonberg and colleagues study also showed ventromedial prefrontal cortex (vmPFC) activity decreases as a function of pumps in a balloon, suggesting that the vmPFC may be engaged in representing increasing potential loss and/or decreasing expected value (Schonberg et al., 2012). Supporting this is the fact that the vmPFC is highly interconnected with the ventral striatum - a region of dopaminergic neurons that adaptively encode for subjective value and expected reward in both humans and (Tobler et al., 2005; Hiser & Koenigs, 2018; Helion et al., 2019; Gallistel et al., 2014).

Past fMRI BART studies typically use the objective risk, the actual pop probability, as a regressor in analyses using general linear models to examine how neural activity changes with varying levels of risk (Schonberg et al., 2012). However, objective risk and the participant's

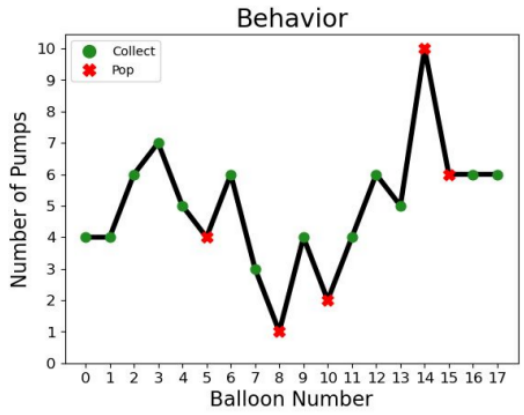
subjective estimate of risk could be very different. Indeed, though the participant understands that the balloon is more likely to pop as consecutive pumps are made, the participant does not know the true underlying pop probability distribution - especially at task onset. Participants can have subjective estimates of the pop probability that are very different from the true underlying pop distribution (Schürmann et al., 2019). Our prior work with PLUM demonstrates that it is possible that participants in BART utilize a more linear representation of increasing pop probabilities, whereas the objective probabilities of the balloons popping as a function of pump number is exponential. Schonberg and colleagues in 2012, in an attempt to circumnavigate using objective measures of risk, took the average number of pumps for each of three balloon types, then took the difference between pump number and the corresponding balloon type mean to get the demeaned number of pumps as a more subjective/participant-specific regressor. However, they only examined the rewarded balloons (Schonberg et al., 2012). This approach likely suffers from the same issues of data censorship associated with the performance measure of adjusted score (Pleskac et al., 2008; Coon and Lee, 2022). Also, many BART imaging tasks are configured such that the reward for every pump is the same, as is the case in the aforementioned study by Schonberg and colleagues. This consistent reward structure could make risk confounded with reward/loss/expected value (de Groot, 2020), leading to risk and reward/loss/expected value being more difficult to differentiate in fMRI signals. In our version of BART, the reward for each pump is drawn from a uniform distribution from $0.05 to $0.25. Ideally, this motivates the participant to be more attentive to rewards for pumping when evaluating the utility of a given choice, given that the expected value is not changing in-step with perceived pop probability.

3.2 Hypotheses

We predict that examining the relationships between fMRI voxel-wise activity and model-based estimations of risk, reward, loss, and expected value as they evolve over the course of the task may be a more direct and insightful approach than using objective/experiment based measures for 1) improving computational models of sequential risk-taking, and 2) investigating roles of brain regions in the context of sequentially-dependent decision making in risk-ambiguous environments. The BART is an ideal task for this venture, as participant's begin with no knowledge of the pop probabilities and must learn from experience (Kóbor et al., 2018; Wu et al., 2021). As participants proceed through the task, individual differences will dictate when and if they transition from an ambiguous, exploratory mode to a more exploitative, methodical risk-taking stratagem in line with risk-taking with known odds (de Groot, 2020). The BART straddles these two disparate conditions of defined risk and ambiguous risk. PLUM is also well-suited for this investigation. Unlike many other models of the BART, PLUM assumes the participant understands that the pop probability of a balloon increases with each pump. Many models of the BART assert that the participant keeps a static representation of pop probability over the course of a balloon, such as the aforementioned EWMV model and earlier BART models (Lejuez et al., 2005; Zhou et al., 2021; Park et al., 2021). We compare the results of using objective regressors to those using model-based regressors. We hypothesize that PLUM-generated regressors will lead to different correlations with hemodynamic responses compared to using experimental, objective regressors such as objective risk, or objective values of gains/losses.

3.3 Methods

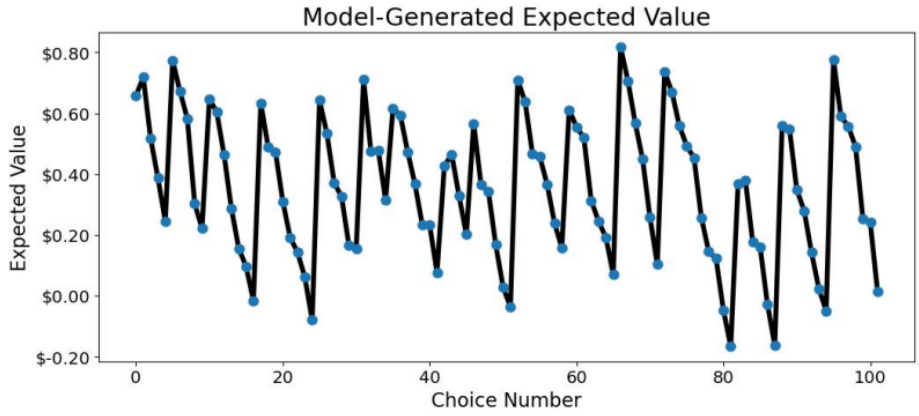
1



2

$$L(Behavior|\alpha, \beta, \gamma, \theta, n$$

3



4

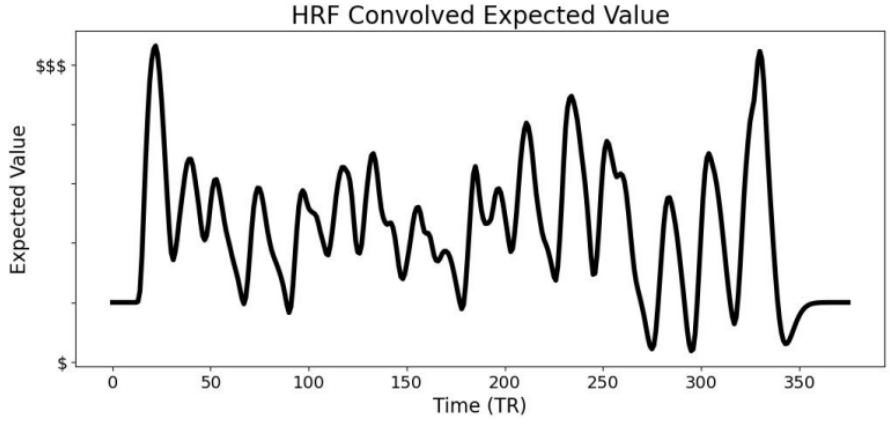


Figure 12: From Behavior to Model-Generated Regressors. 1) Depicts the observed behavior, which is every choice made by the participant. 2) We fit the observed behavior with PLUM to extract best-fitting parameters. 3) Using best-fitting parameters, we generate model-latent values that change over time, in this case expected value. 4) We convolve the generated latent values with an HRF (a Glover HRF model was used in this example).

3.3.1 Participants

27 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.

3.3.2 Task Design

Participants completed one block of the BART task identical to the task previously presented. Participants pumped 18 balloons in total (6 of each balloon type, randomly ordered). Participants made responses via an MRI compatible button box. Brain images were collected using a Siemens Trio Magnetic Resonance Imaging with TIM system. The BART was administered as a part of a larger task battery (Weichart et al., 2021).

3.3.3 fMRI Model-Based Analyses

All fMRI processed fMRI data was analyzed using the NiLearn Python library, which utilizes the Scikit-Learn Python library (Abraham et al., 2014). For each participant, I will instantiate a first-level general linear model (GLM). Each participant's GLM utilizes an event matrix customized for the respective participant. The event matrix contains information regarding the onset and duration of both experiment-related events, objective measures of value and risk, and PLUM-generated estimates of subjective risk, subjective value, and subjective utility. These events were then convolved with a Glover hemodynamic response function (HRF)

to generate a set of explanatory variables (EVs) for the first-level GLMs (Weber et al, 2015). We also include a contrast of between the subjective regressors and the objective regressors in order to detect any differences in significantly correlating voxel activity. Motion artifacts were included as nuisance regressors. A gray matter mask was formed from the aggregate of all participants and used in the first-level GLMs to help differentiate gray matter voxels from other brain tissues. For cluster differentiation, a minimum distance of $8mm^3$ was used.

3.3.4 Model-Generated Subjective VS Objective Explanatory Variables

Experiment-related EVs include rewards for pumping, current balloon value, and objective balloon pop probability. Once individual behavioral data is fit with PLUM, we utilized the best-fitting parameters from each model fit to generate model-subjective regressors. These regressors represent various PLUM-predicted time series of subjective risk, subjective reward, and subjective loss. PLUM will generate these values at each choice point. Thus, we can convolve these values with a model of the HRF to create a continuous time series that can be correlated with individual voxels in the first-level GLM.

Comparison	Description	PLUM Equation
Probability	Probability of balloon popping	Eq 7
Pump Reward	Reward added to total for pumping	reward^{γ}
Balloon Total	Value of balloon	$\boldsymbol{\theta}\text{balloon}^{\gamma}$

Table 2: Comparing Objective and Subjective Values of PLUM Mechanisms

To compare subjective and objective valuations of risk and value, we fit three first-level GLMs for each participant (Table 2). For example, we fit a first level GLM in which the subjective and objective pop probabilities are EVs and compute a contrast between them. Once every subject is fit via a first-level GLM, we model the group-level data using a second-level GLM (Monti, 2011; Gläscher & O'Doherty, 2010, Weber et al., 2015) . A second-level GLM includes the individual-level regression coefficients as the input variables and allows for the identification of brain regions that show consistent activation across subjects as well as the estimation of the variability of the effects across subjects. The issue of multiple comparisons arises when performing voxelwise analyses. To correct for this, we use a false positive rate (FPR) correction in conjunction with cluster-extent thresholding to identify significant clusters (corrected $p < 0.05$, cluster size > 12) (Wu et al., 2021). Cluster-extent thresholding was performed via AFNI's 3DCulstSim tool (Cox, 1996; Cox & Hyde, 1997; Cox et al., 2017a; Cox et al., 2017b). We then compared the subjectexamining any overlapping or distinct significantly correlating brain regions.We compare the maximum correlation statistical values between the

objective regressors with those of the model-based regressors, assessing any differences between the subsequent significant clusters.

3.4 RESULTS

3.4.1 Behavioral Results

Participant performance was measured using the performance metric previously described (*M)*. This performance metric ranges from 0 to 1, with 0.5 being considered optimal performance. Scores above and below 0.5 are considered suboptimal. Scores below 0.5 reflect overly conservative decision making (not taking risks). Scores above 0.5 reflect overly risky decision making. On average, participants performed slightly conservatively with a mean score of 0.44, indicating participants were risk-averse overall.

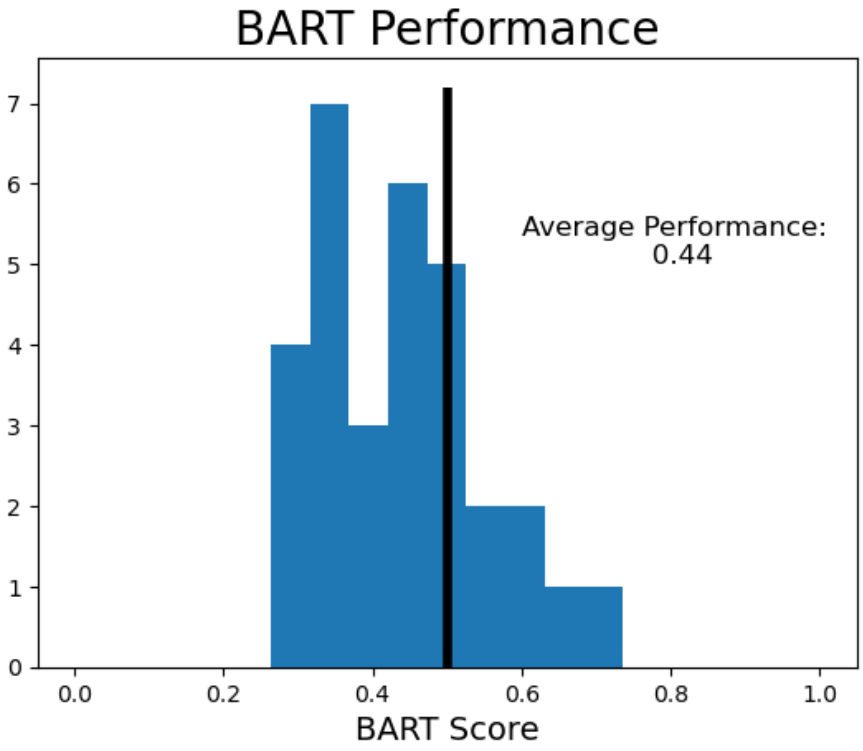


Figure 13: Participant BART Performance. Participants (N=27) performed on average slightly conservatively. The black line indicates optimal performance of 0.5.

3.4.2 Imaging Results

3.4.2.1 Comparing Subjective and Objective EVs

Presented here are the results of our regression and clustering analyses. For full tables of cluster correlations, see appendix S2.

Contrasting subjective and objective pump reward did not demonstrate much differentiation, for both did not portray much correlation with voxel activity after thresholding. The superior parietal lobule (Z = 5.5, size = 416mm^3) contained the highest peak activation for subjective pump reward.. It was also was associated with left middle OFC (Z = 4.2, size = 128mm^3). Objective pump reward was also associated with activity in the superior parietal lobule, but in the opposite direction (Z = -5.4, size = 424). It too was associated with the OFC as well (Z = -4.5, size = 584mm^3).

The contrast for balloon total was similarly stark. The strongest peak activation was most correlated with left middle occipital lobe for subjective totals (Z = 5.6, size = 4272mm^3). The strongest peak for objective total was left superior frontal gyrus (Z = 4.80, size=128mm^3).

Pop probability was the most informative comparison. Contrasting subjective pop probability and objective pop probability shows that modeled subjective probability was correlated with many significant clusters. Objective probability had only two significant clusters, with the strongest positive correlation being in the right superior temporal lobe (Z = -4.8, size = 168mm^3). This region has been associated with risk taking under ambiguity and has been shown to differentiate risk-taking strategies in alternative choice tasks (Paulus et al., 2001). It has also been singled out in differentiating sex and age differences during risky decision making (Lee et

al., 2009). Subjective probability was correlated with multiple regions, with the strongest being the frontal gyrus (max Z = 6.96, size = 5464.0mm^3) with a subcluster in the right inferior OFC (Z = 4.98). ACC activity was correlated with subjective probability. ACC activity is associated with response error monitoring, conflict, and likelihood estimation (Brown & Alexander, 2017; Brown & Braver, 2005; Brown & Braver, 2008; Fukunaga et al., 2012). Subjective probability correlated significantly with various clusters in the motor cortex, possibly indicating a relationship between perceived pop probability and motor planning or inhibitory control. An important mechanistic distinction between the objective probability and the subjective probability is that PLUM makes the assumption that participants perceive the balloon's pop likelihood to be increasing in a more linear manner. The objective probability, however, is exponential. The disparity in number of significantly voxel activity and clusters provides some evidence that subjective probability is a more useful and more fitting predictor of brain activity than objective risk.

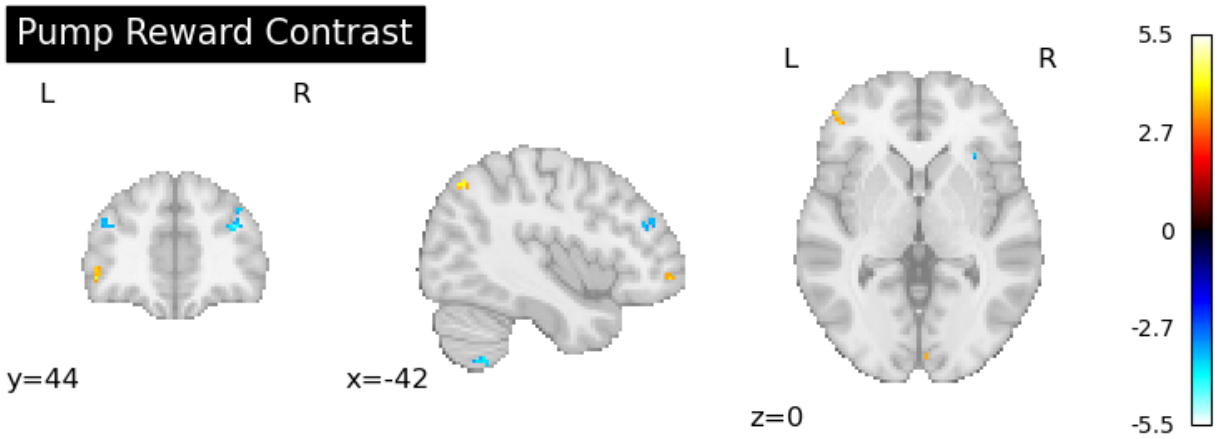


Figure 14: Pump Reward Contrast

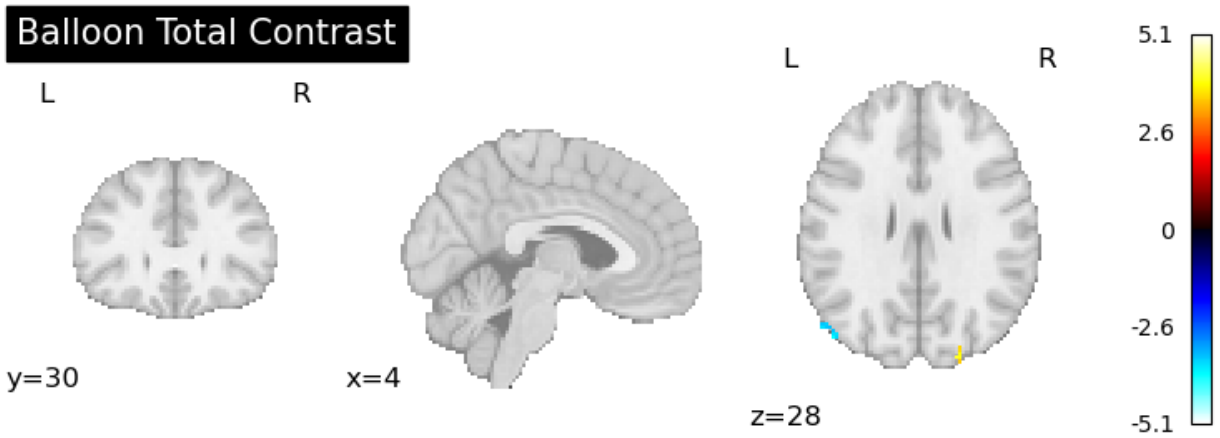


Figure 15: Balloon Total Contrast

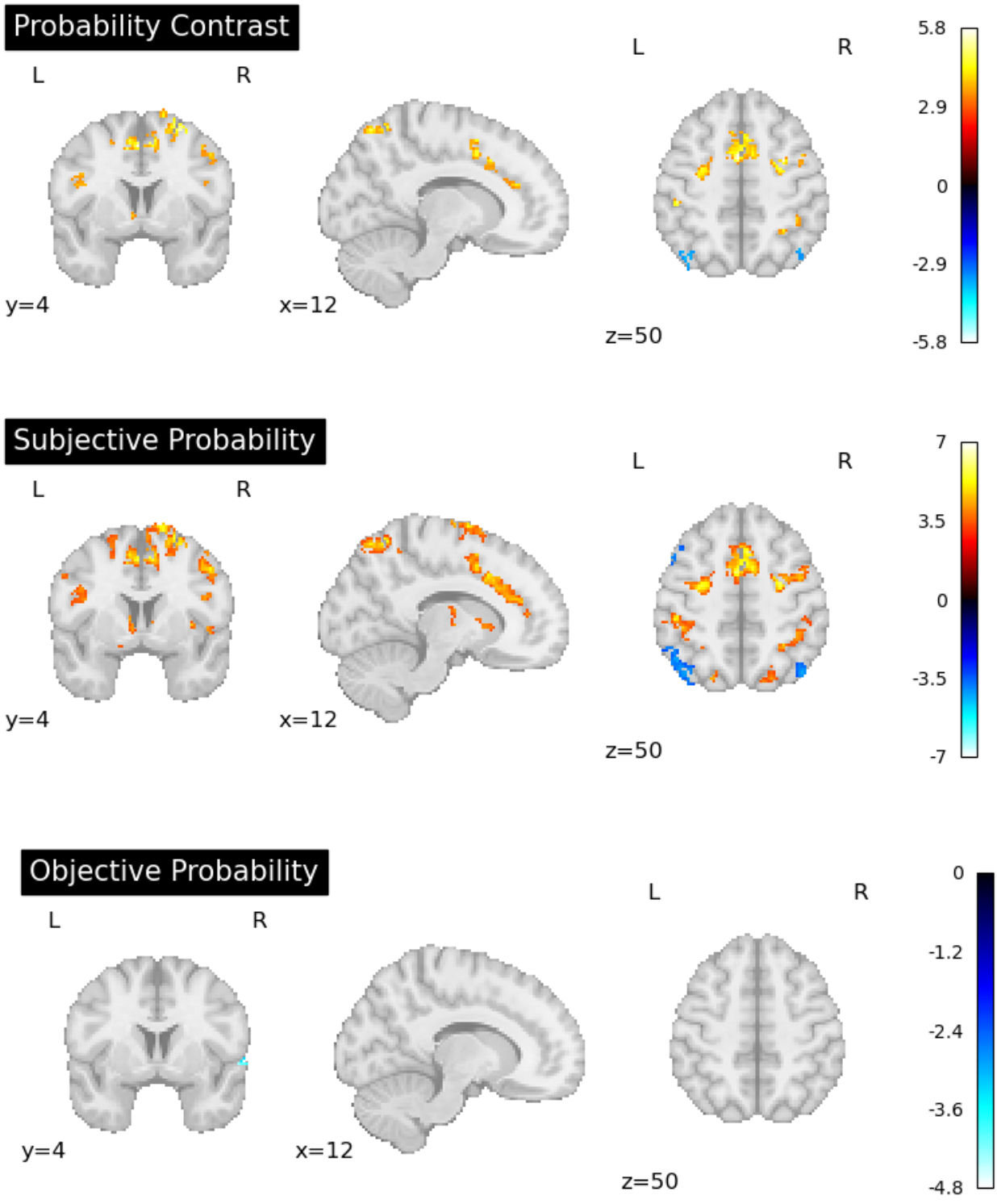


Figure 16: Neural Correlates of Pop Probability

3.4.2.1 PLUM-Estimated Expected Value

We also assessed if PLUM-predicted expected value would be significantly correlated with brain activity. The max peak was located in the left precuneus (Z = 5.14, size = 2904mm^3). The precuneus has been noted to be at the heart of much speculation We also observed some correlated right medial OFC activity (Z=4.35, size=168mm^3). Previous work demonstrates that in youths with greater risk-taking tendencies, precuneus activation is stronger than peers with lower risk-taking tendencies (DeWitt et al., 2014; Eckstrand et al., 2017).

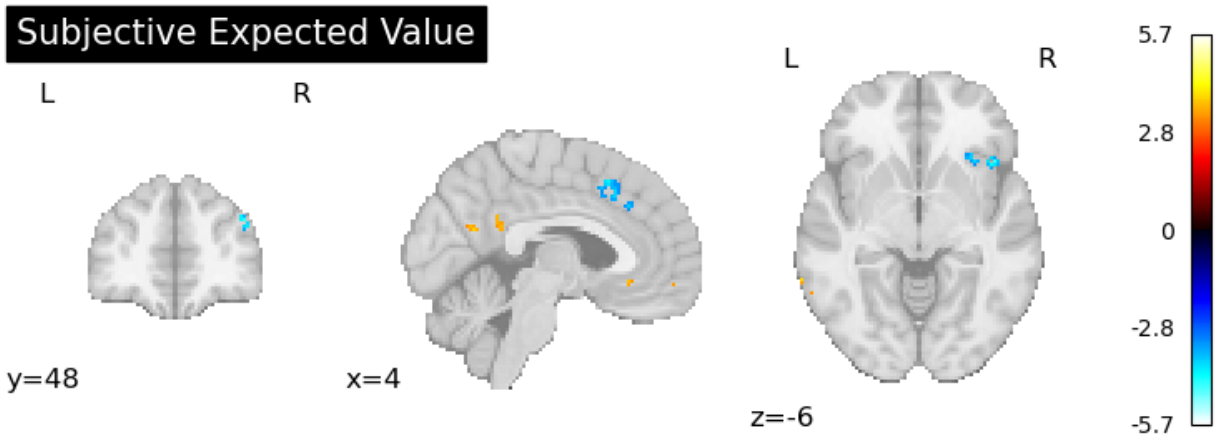


Figure 17: Neural Correlates of Expected Value

3.5 Conclusions and Future Work

This preliminary investigation into neural correlates of model-based cognitive mechanisms provides some support that computational models of cognition can be used to generate subjective values of risk in a neurally meaningful way. This work represents a stepping stone for evaluating models of risky decision making by using models to generate a time series that can then be correlated to neural activity over the course of a task. This work also steps away

from traditional BART studies that typically use parametric forms of risk and average across different conditions (Schonberg et al., 2012; Li et al, 2020). Instead, we convolved a model-produced time series with an HRF to make predictions choice-by-choice predictions of brain activity. Participants are learning a probability space of outcome likelihoods over time through trial and error, having to balance the draw of more reward with the weight of increasing failure (Lejuez et al., 2002; Lejuez et al., 2005). Even with the limited data in this data set, we found evidence that a cognitive model can generate a more accurate predictor of neural activity than using objective ones. This method could also be used to compare and contrast BART models beyond just assessing their goodness of fit to behavioral data alone. Instead, model mechanisms can be singled out if they are capturing a latent mechanism observed in neural data, thus allowing the model builders to modify the task and/or the model to better suit their needs.

In future work, careful consideration should be made to the BART when being set up for fMRI analysis. One consideration is the reward structure. Though we had varying rewards, we did not see a relationship between subjective or objective reward with neural activity. However, this could have played a role in why balloon total was also not affiliated with brain regions normally associated with risk-taking in BART, as the balloon's total no longer increased in-step with risk. If this is so, it means that varying the reward structure should be considered in future BART variants to better disentangle gains, losses, and risk in imaging studies. A future task variant could increase the upper limit of the reward range to perhaps better elucidate a more careful consideration of it. It is also the case that the BART task presented here had a bank that slowly lost value as participants deliberated. This could be detrimental to collecting fMRI data, as the diminishing bank pushes participants to respond faster and possibly not take as much time to deliberate their choices as they normally would.

4. Study 3: Utilization of the Framing Effect to Differentiate Risk-Taking & Risk-Avoidant Behaviors

Authors: Adam W. Fenton[1], Peter J. Castagna[2], Michael J. Crowley[3], Per B. Sederberg[1]

Summary

In this chapter, I present a computational model designed to investigate asymmetrical risky decision making behaviors. We utilize the Balloon Risk Avoidance Task (BRAT) - a task similar to the BART but in which participants deflate balloons instead of inflating them.

4.1 INTRODUCTION

Despite being a time of peak physical health, adolescents experience a spike in morbidity and mortality largely from preventable risky behaviors (e.g., car crashes, unprotected sex, drug use, binge drinking; Eaton et al, 2012). This volatility in youths has led to significant efforts to understand why adolescence is a time of excessive risk-taking (Moore et al., 2013), develop ways to measure an individual's propensity to engage in risky behaviors, and find out how these behaviors might be prevented (Reyna & Farley, 2006).

The Balloon Analogue Risk Task (BART) (Lejuez et al, 2002) was developed as a clinical-oriented measure of risk-taking predispositions, and therefore, aims to identify individuals more likely to engage in risky behaviors (Lejuez.et al, 2002; Lejuez et al, 2005). The BART can successfully identify high risk-taking individuals (Aklin et al 2005; Lejuez et al, 2002; Lejuez et al, 2004). For instance, performance on the BART is related to several constructs highly related to risk-taking such as sensation-seeking, impulsivity, and psychopathy (Hunt et al

[2] Department of Psychology, University of Alabama
[3] Yale Child Study Center, Yale School of Medicine

2005; Lauriola et al 2014), as well as engagement in risky behaviors (Lejuez et al, 2004). BART performance has been linked to social cues such as peer pressure (Kessler et al., 2017).

Whereas a great deal of literature has focused on adolescent risk-taking behaviors (e.g., Lejuez et al., 2005), risk avoidance has received considerably less attention. Risk avoidance has been defined as the action of evading behaviors that expose the individual to loss or harm (Crowley et al., 2021). Adolescence serves as pivotal period for the development of anxiety disorder (e.g., Kessler et al., 2007), making it reasonable for one to expect a rise in risk avoidant behaviors in some adolescents (Lorian and Grisham, 2011).

Given that risk-taking is conceptualized as approach motivation and sensation seeking (Coffey et al., 2011; Collado et al., 2014; Jones & Lejuez, 2005; Shoham et al., 2016), assessing risk avoidance is critical to get a complete picture of avoidance versus approach motivations. Recent evidence suggests that the validity of a novel behavioral measure of risk avoidance, the Balloon Risk Avoidance Task (BRAT) (Crowley et al, 2021). Specifically, avoidance behavior on the BRAT was positively associated with anxiety and fear, negatively related to risk taking as measured by the BART, and unrelated to depressive symptoms, trait surgency (approach motivation), IQ and age.

The BRAT is a variation of the automatic BART (Lejuez et al., 2002; Pleskac et al., 2008). In the BRAT, participants are presented with a balloon that they watch grow to be fully inflated on a computer screen. A meter beside the balloon rises vertically from safe to danger. Participants are shown that if no action is taken, the balloon is sure to pop. Participants prevent the balloon from popping by choosing the number of "pumps of air" (between 1-128) they wish to remove from the balloon, thereby deflating it and reducing the likelihood that it pops. The larger the number that they choose, the less points they can earn for that trial (for example,

choosing 20 pumps would result in saving 108 points). The BRAT differs from most other behavioral risk tasks in that it favors risk avoidance. This framing biases participants to move away from risk, making the application of the BRAT in the present study a measure of risk avoidance.

Though risk-taking and risk avoidance may at first glance appear to be two sides of the same coin, it is possible that the underlying cognitive mechanisms operating in such circumstances differ in critical respects. The BRAT and the BART are essentially the same task, but the decision maker (participant) is given a different initial perspective or framing. This framing effect can have important ramifications for how prospective gains and losses are evaluated, even if the risks are identical (Khaneman & Tversky, 1981,1984; De Martino et al., 2006).

In recent years, there have been a number of advances in the computational modeling of behavioral tasks to improve the measurement of underlying cognitive processes (Ahn et al., 2014; Park et al., 2021; Zhou et al., 2021; Ratcliff, 1978). Given the promising behavioral results of the BRAT and success of applying computational models to BART performance (eg., Park et al., 2021), the purpose of the current study is to investigate differences between BART and BRAT risk taking behavior. In doing so, we 1) propose a computational model that is able to account for both risk-taking and risk avoidance scenarios, and 2) gain evidence that the BRAT is tapping into a different aspect of one's cognitive signature compared to the BART.

4.2 METHODS

4.2.1 Participants.

The sample consisted of 127 children (67 female), recruited via a mass mailing list provided by a credit mailing company, Experian, targeting the towns surrounding New Haven, CT. The reported ethnicity among the final sample of children was as follows: 12\% African American, 8.8\% Hispanic, 6.4\% Asian, 3.2\% Native American and 68\% White, 1.6\% other. Four children lacked a BRAT behavioral assessment, and three children lacked a questionnaire assessment. The final dataset consisted of 120 children. Children were fluent in English and had no evidence of serious mental illness (psychosis, autism, bipolar disorder), assessed via a parental telephone screen. Children were considered ineligible for the study if their caregiver, typically their mother, reported that the child in question was being treated for, or carried a diagnosis of psychosis, autism, or bipolar disorder. Children were intellectually in the normal to high normal range (verbal IQ mean = 112.25, SD = 12.88, range 82-139) (Wechsler, 1999). The mean age of the children was 13.82 years (SD = 2.03 years, range = 10.10 - 17.78 years). This research was approved by the Yale University School of Medicine Human Investigation Committee.

4.2.2 Tasks

There were two tasks in this study: The balloon analogue risk task (BART) and the balloon risk avoidance task (BRAT). Participants performed one block of each and were counter-balanced. Before beginning the tasks, each participant viewed a prize book, with prizes values associated with points earned.

4.2.2.1 *The Balloon Risk Avoidance Task (BART)*

The BART involves inflating a computer-generated balloon that can pop if over-inflated. The BART consists of 30 balloons, played one at a time, each worth a maximum of 128 points.

On each trial, the participant is presented with a balloon and must inflate the balloon by inputting a number between 1 and 128 where higher numbers indicate more pumps/inflation and more points. However, by inflating the balloon the participant also increases the potential risk for explosion and a total loss of points for the current trial. Explosion points are predetermined, following a normal distribution (Lejuez.et al., 2002), with a mean of 64.

4.2.2.2 *The Balloon Risk Avoidance Task (BRAT)*

The BRAT involves deflating a computer-generated balloon that is guaranteed to pop if participants do nothing. The BRAT consists of 30 balloons, played one at a time, each worth 128 points. The participant initially sees a question mark in the center of the screen prompting them that the trial is about to begin. Next, the participant watches a balloon inflate to the point of popping, indicating the maximal level of risk. A meter next to the balloon rises from a safe level (blue) to a red level (danger). The participant has 6 seconds to deflate the balloon before it pops. A red, segmented timer indicates the elapsed time. If the participant fails to deflate the balloon, then it immediately pops. The participant can deflate the balloon by inputting a number between 1 and 128 into a dial with a mouse. However, by deflating the balloon the participant also returns some of the points they could possibly earn. For example, if the participant decides to deflate the balloon by 88 points, and the balloon does not pop, then the participant keeps the remaining 40 points. If a balloon is left inflated past its popping point, then the balloon pops and the participant loses all the points for that trial. Explosion points are predetermined, following a normal distribution identical to the BART (Lejuez et al, 2002), with a mean of 64.

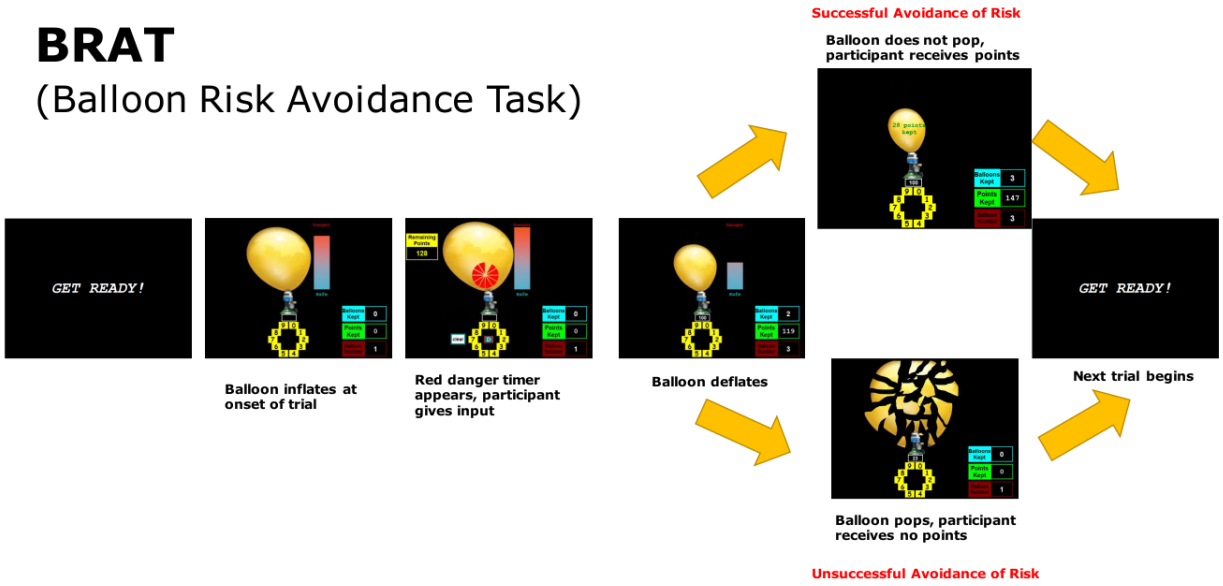


Figure 18: BRAT Design. The BART for this study was identical except for the balloon starting deflated.

4.2.3 Independent Measures

Youth Risk Behavior Survey (YRBS). The adolescent completed the YRBS, a self-report measure of risk-taking behaviors (version 2007; Brener et al., 2002; Center for Disease Control and Prevention 2007). In order to increase adolescent honesty, participants were told that their responses to the substance use questions were confidential and would not be shared with their parents except in cases regarding self-harm or concern for others safety. A certificate of Confidentiality (COC) was obtained, and the adolescents were told that the confidentiality of their reports of illegal behaviors (including substance use) were protected by the COC. The YRBS monitors several categories of health risk behavior including safety, violence-related behaviors, sad feelings and attempted suicide, tobacco use, alcohol use, drug use, sexual behavior, body weight, dietary behavior, physical activity, and other health related topics.

We were particularly interested in questions assessing the use of alcohol, illicit drugs, nicotine, and marijuana, as well as physical fighting. The alcohol use subscale consisted of five questions (e.g., “During the past 30 days, how often did you drink alcohol?”, “During the past 30 days, on how many days did you have at least one drink of alcohol?”). Illicit drug use was assessed by twelve questions (e.g., “During the past 12 months, has anyone offered, sold, or given you an illegal drug?”, “During your life, how many times have you used ecstasy (also called MDMA)?”). Risky sexual behavior was comprised of three questions (e.g., “How old were you when you had sexual intercourse for the first time?”, “The last time you had sexual intercourse, did you or your partner use a condom?”). Nicotine use was assessed by three questions (e.g., “Have you ever tried cigarette smoking, even one or two puffs?”, “During the past 30 days, on how many days did you smoke cigarettes?”). Marijuana use subscale was also comprised of three questions (e.g., “During the past 30 days, how many times did you use marijuana?”, “How old were you when you tried marijuana for the first time?”). Finally, physical fighting was assessed by three questions (e.g., “Have you ever been in a physical fight in which you were hurt and had to be treated by a doctor or nurse?”, “Have you ever been in a physical fight?”).

4.3 MODEL

One of the hallmarks of the BART is that the risk of failure is initially uncertain. The decision maker learns information about the probability space of the task from experience. Performance on the BART is a balancing act between exploration and exploitation. It is advantageous to explore the balloon popping point distribution and then exploit the knowledge gained. Modeling BART behavior necessitates capturing this process of learning an ambiguous

probability distribution and making decisions upon an updating internal representation of the probability of success/failure.

Developing a model of any cognitive task requires acknowledging the constraints presented by the structure of the task itself. In this case, the participant must enter a number for how many pumps they wish to add or remove from the balloon for the BART and BRAT respectively. BART versions in which the participant enters a number of pumps, holds down a key to inflate or presses a key to stop continuous inflation are known as automatic BART versions. Including the original BART, some versions have the participant make each pump of the balloon individually. Behavior on sequential versions compared to automatic versions are essentially identical (Young and McCoy, 2019). The automatic versions of the BART have the distinct advantage of typically being faster to administer than version in which participants must make individual pumps. In BART versions in which the participant makes each pump individually, as long as the participant is keeping track of how many times they have pumped a given balloon, they know precisely the number of pumps it took to pop a balloon if a balloon pops. This knowledge of the true popping point of a balloon is very informative for updating the participant's internal representation of the balloon's pop probability space., in the versions of BART and BRAT presented here, no feedback about the point at which the balloon popped is provided. When the participant enters the number of pumps they desire, they watch the balloon change size accordingly. Once the balloon has reached the particular size, it either pops or is collected. This means that in the event of a pop, the exact point at which the balloon popped is unclear. For example, when performing the BART block in this study, a participant may select for 50 pumps to be added to the balloon that has a popping point of 40 pumps. The balloon will inflate and 50 points will be added, then it will pop, obscuring the fact that the popping point was

at 40 pumps. A model for this type of automatic BART and BRAT must take into consideration the lack of clear feedback.

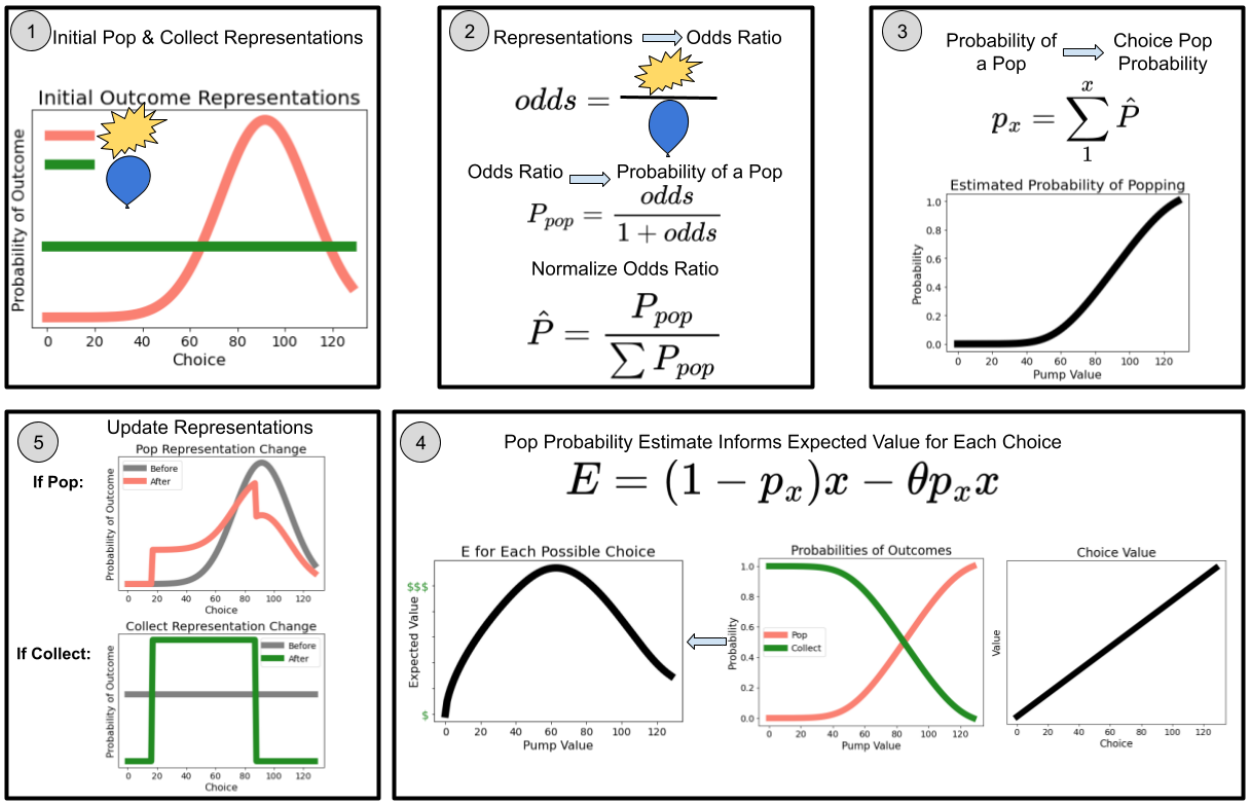


Figure 13: Model Mechanisms. The participant keeps track of observations of pops and collections independently from one another. The participant has an initial bias for expecting pops across choice values (the red line), but has a uniform distribution for expecting collections across all choices (the green line). The participant compares the odds of popping to the odds of collecting at a given choice point via an odds ratio, which is then transformed into a perceived probability of the balloon popping. This pop probability, P, is used to evaluate the expected value (E) of each possible choice. When a choice is made and subsequent outcome observed, the underlying outcome of distribution is updated, capturing the process of learning the probability space of the balloons.

Due to the lack of accurate feedback in regards to balloon popping points, we hypothesize that the participant must then keep track of the outcomes observed in relation to the

pump number given., we suggest that the participant is keeping track of pops and collections separately. For developing a generative model to test this assertion, we turn to signal detection theory (SDT). SDT models are useful for a wide variety of cognitive processes and can represent many 2-choice decision processes (Turner et al., 2011). We instantiate two initial distributions; one representing observations of pops (P) and one representing observations of collections (C). These are not representative of the probability of a balloon popping or being collected. Rather, these represent trials in which participant responses resulted in pops or collections. P is a truncated normal distribution with a mean of μ and a standard deviation of σ. Both parameters are free parameters in the model. C is a uniform distribution from 0 to 128. μ and σ together form the shape of P initially, representing the initial pop observation bias at the start of the experiment. P and C are probability mass functions in that they represent a range of discrete, whole values of pumps (ie, there can't be a half pump). We then utilize the P and C distributions to calculate the estimated pop probability via calculating an odds ratio as seen in Equation 27.

$$odds_i = \frac{P_i}{C_i}, \quad (27)$$

Normalizing this odds ratio, we get P, which is the normalized probability of the balloon popping.

$$P_i = \frac{odds_i}{1 + odds_i}, \quad (28)$$

This allows us to calculate the probability of the balloon popping at each specific choice i by calculating the cumulative sum for each choice up to and including i.

$$p_i = \sum_{i=1}^{128} P_i \quad , (29)$$

We then utilize prospect theory to calculate the expected value E for a given choice. Prospect theory posits that the participant estimates the value of a given choice by comparing the possible reward scaled by the estimated likelihood of success to the value of a punishment associated with loss scaled by the likelihood of losing. Equation 30 shows this comparison.

$$E_i = q_i * r_i^{\gamma} - \theta * p_i * r_i^{\gamma} \quad , (30)$$

γ is a free parameter representing the modulation of value, with higher gamma increasing the disparity between reward values non-linearly. θ represents the participant's overall loss aversion, and thus scales the perceived possible punishment for getting a pop.

Utilizing a softmax equation, we calculate the likelihood of the participant's choice. A softmax equation turns relative strengths into relative probabilities of making a choice - in this case, relative expected values into relative likelihoods of making each respective number of pumps.

$$d_i = \frac{e^{E_i * \tau}}{\sum_{i=0}^{i} e^{E_i * \tau}} \quad , (31)$$

τ is a free temperature parameter. Lower values of tau flatten the perceived differences between choices, thus making the likelihoods of far-apart choices more similar than if a higher τ were used. Once a choice is made, the participant observes the result and the corresponding distribution of observations is updated by means of a kernel, K. If the balloon popped, the P distribution is updated. If the balloon was collected, the C distribution is updated by means of a box kernel, K. We refer to the observed outcome distribution as *O* in equation 33 (refer to Turner and Van Zandt, 2011 for more on this box kernel approach).

$$K\left(\frac{x-x_i}{h}\right) = \begin{cases} \frac{1}{2h} & \text{for } |x-x_i| \leq h \\ 0 & \text{else} \end{cases}, \quad (32)$$

$$\hat{f}_{O,i}(x;h) = (1-\lambda)\hat{f}_{O,i-1}(x;h) + \lambda K\left(\frac{x-x_r}{h}\right), \quad (33)$$

h is a free parameter determining the width of the kernel, and λ is a free parameter determining the degree to which the distribution within the bounds of h is updated, acting as a learning parameter and scales the rate of decay. The box kernel increases values within the boundary h and decays the values outside of that boundary. Figure 14 demonstrates the updating process of the distributions as a result of trial decisions and outcomes.

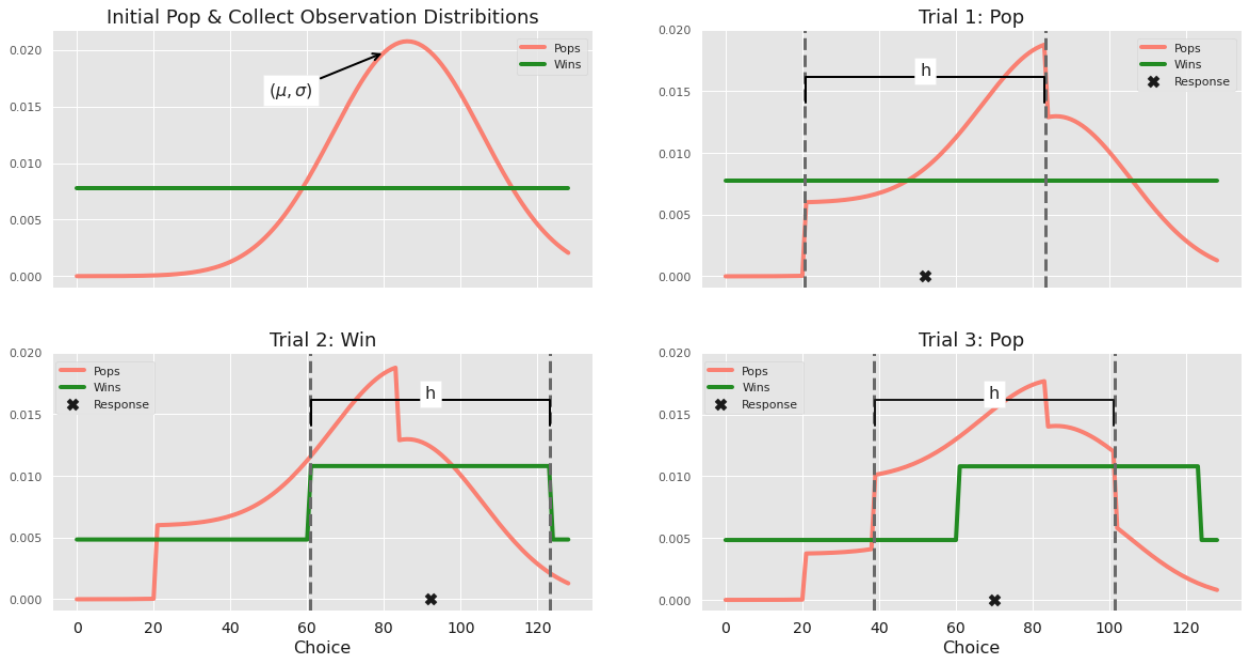


Figure 19: Updating Representations. The initial representation of pops (red) is determined by the free parameters μ and σ. The collection representation (green) is a uniform distribution from 0 to 128. In the first trial, the participant makes a response (denoted by the black "x"), resulting in a win. The representation for collections is then updated, with free parameter h determining the width of the box kernel over the distribution. Free parameter λ scales how much the representation within range h spikes, while the values outside of h decay. Trial 2 results in another win, and so the win representation is updated again. Trial 3 results in a pop, so the pop representation is updated.

4.3.1 Model Fitting

Our model proposed here allows us to derive estimated parameter values that most likely manifested the observed behavior given the model's mechanisms (Busemeyer & Stout, 2002; Huys et al, 2016). Given that we have a generative model, once we fit a participant's data with the model, we can then use the best-fitting parameters to simulate the participant's behavior. A good model should be able to reliably recreate the observed data. We establish a prior distribution for each parameter, then use differential evolution in conjunction with Markov chain Monte Carlo (DE-MCMC) to sample values from each prior distribution for each parameter. Using these sampled parameter values, we calculate the likelihood of observed behavioral data

(Turner & Van Zandt, 2012; Turner et al., 2013). We make use of the freely available Python codebase RunDEMC (https://github.com/compmem/RunDEMC). We performed a total of 1000 iterations with the first 400 being burn-in iterations to reach the maximum a posterior (MAP) estimate and the later 600 iterations sampling with purification once every fifth iteration.

Parameter	Description	Prior
h	Kernel width	$\text{logit}^{-1}(\mathcal{N}(-1, 1.5) * 128)$
λ	Kernel updating scalar	$\text{logit}^{-1}(\mathcal{N}(0.0, 1.4))$
γ	Modulates value	$\exp^{\mathcal{N}(0.56,0.75)}$
τ	Decision likelihood temperature parameter	$\exp^{\mathcal{N}(0.56,0.75)}$
μ	Initial mean of pop representation	$\text{logit}^{-1}(\mathcal{N}(0.0, 0.75) * 128)$
σ	Initial standard deviation of pop representation	$\exp^{\mathcal{N}(2.61,1.0)}$
θ	Loss aversion	$\mathcal{N}(0.5,1.0)$

Table 3: Model Parameters.

4.4 RESULTS

4.4.1 Behavior

We first examined the behavioral differences between participant's performance on BART and BRAT. The metric by which we measure performance in this study is the mean number of pumps across all balloons within a task. Looking at the mean number of pumps each participant made for both tasks, we find that participants generally left more pumps remaining in the balloon for BRAT than they did for BART. This means that participants tended to take greater risks in the BRAT than the BART. Within-subject performance was skewed negatively when calculating the difference between BART and BRAT, whereby the mean pump for BART was X (SD = X) and the mean pump for BRAT was Y (SD = Y) (subtracting BRAT mean pumps from BART mean pumps) (type of t-test, $t(DF)$ = -5.54, $p < 0.001$). While performance on these two tasks is moderately correlated (r=0.492, $p < 0.001$), they are not perfectly correlated. This lack of perfect collinearity provides evidence that, while BART and BRAT are mirror-images of one

another, the perception of moving toward or away from risk could profoundly affect the way participants perceive and react toward risk in this task.

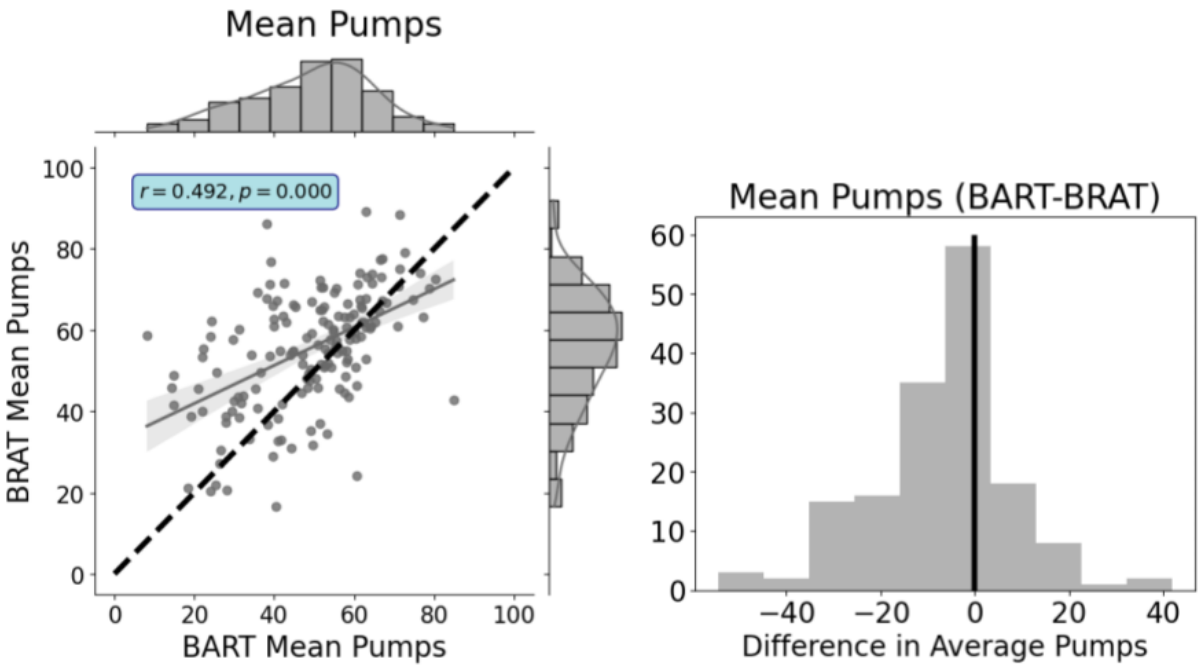


Figure 20: BART vs. BRAT Behavior: Performance on the BART and BRAT is measured via the average amount of pumps of air the participant chooses to have remain in the balloon. The left plot shows that performance on BART and BRAT is moderately correlated with one another. The right panel shows the distribution of within-subject performance differences between BART and BRAT. A two-sample t-test shows that participants were significantly more conservative on the BART than they were on the BRAT (t = -5.54, $p < 0.001$).

4.4.2 Model Simulations

To assess the efficacy of the model, we fit the model to each participants' BART & BRAT data independently. Using the best-fitting parameters from each fit, we then simulate each respective task. For a given participant, we performed 1000 simulations for each task, with each simulation using the exact same balloons the participant experienced in the same order. This provides a distribution of simulated choices and overall behavior. Figure 16 shows an example of

simulating two participants' behaviors. = Since the participants enter the number of pumps/deflations at the onset of a balloon, we need not be concerned by issues associated with sequential BART tasks in which participants make individual pumps such as data censorship (Pleskac et al., 2008; Coon and Lee, 2022). Figure 17 shows that the model was able to reliably reproduce the observed behaviors. Model generated BART performance correlated very strongly with observed BART performance (r(DF) = 0.975, $p < 0.001$), and generated BRAT performance correlated strongly with observed BRAT performance (r = 0.966, $p < 0.001$). We also examined how similar the standard deviation of simulated choices was to that of the observed choices. We found that the simulated choice standard deviation for both tasks was strongly correlated with the observed choice standard deviations (BART: r = 0.962, $p < 0.001$; BRAT: r = 0.923, $p < 0.001$).

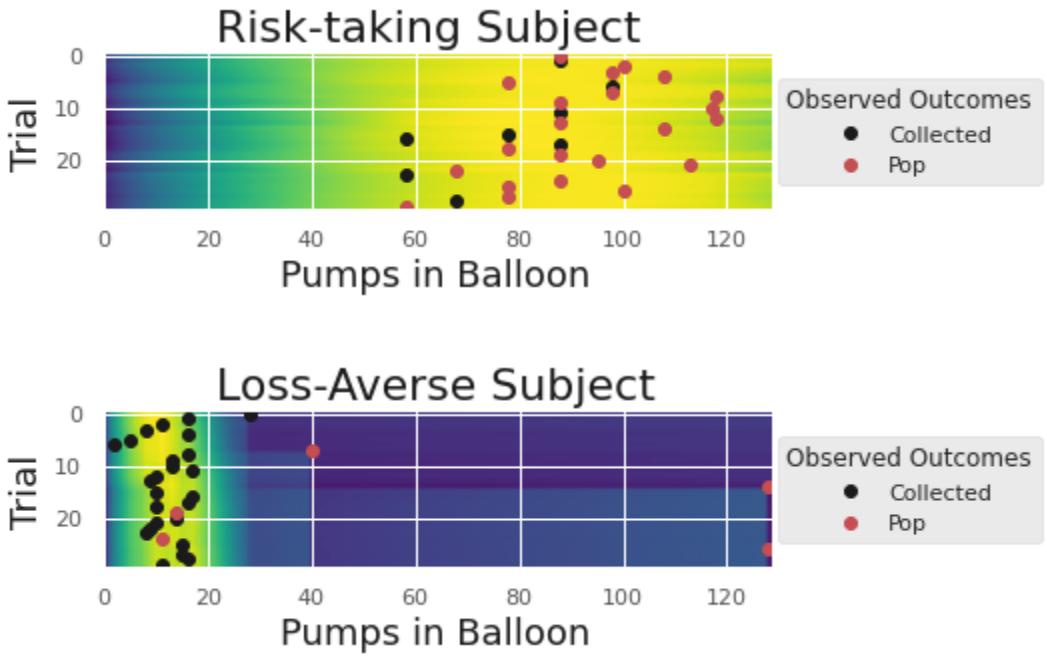


Figure 21: Single Simulation. Depicted here are the model-based likelihoods of any of the 128 choices being selected for two different subjects. The top figure shows the choice likelihood estimates for a participant that took greater risks. The bottom figure shows the results of a participant that performed conservatively. The black and red circles indicate the actual choices made by the participant. Red circles indicate outcomes where the balloon popped, while black

circles indicate collections. Lighter, yellow regions indicate higher model-based choice likelihoods, while darker regions indicate lower choice likelihoods.

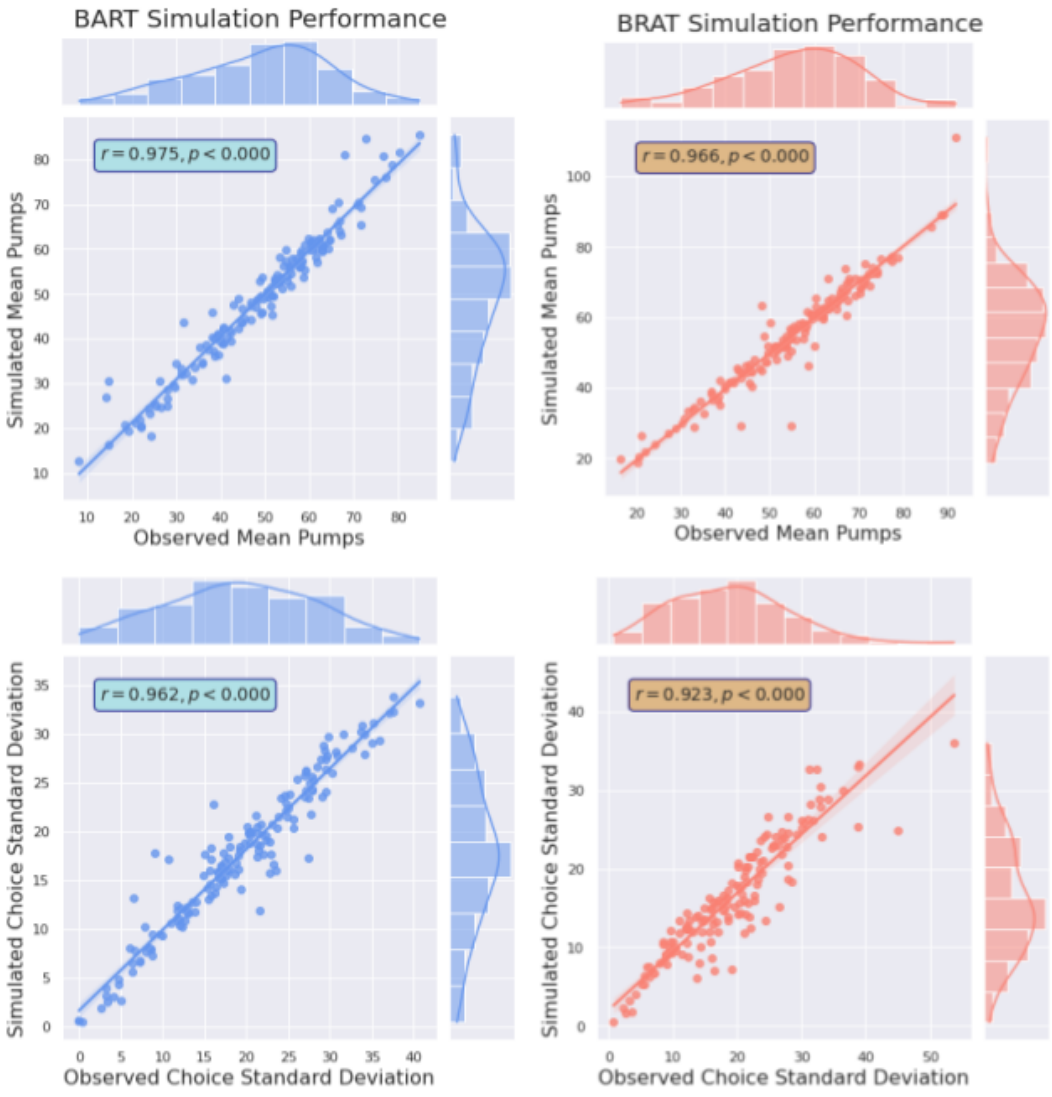


Figure 22: Simulation Results. Simulations were conducted using best-fitting parameters for all subjects. 1000 simulations were performed for each subject. Average simulation performance is correlated with observed performance.

4.4.3 Relationships Between Parameters & Behavior

Our proposed model is able to accurately generate observed behavior. Thus, we can examine interactions between the model parameters themselves and the observed behavior. For all statistical comparisons, see Table 4. Here we will report the significant statistics. h , the box kernel width, is not significantly correlated with mean pumps in the BART and is weakly inversely correlated with mean pumps in the BRAT. It is also not significantly correlated between BART and BRAT. λ, the learning rate, is not significantly correlated to performance in both BART and BRAT, nor is λ correlated with itself between the two task fits. μ, the initial mean of the pop representation, is correlated with BART performance (r=0.53, $p < 0.001$) and correlated with BRAT performance(r=0.40, $p < 0.001$). It is also correlated between BART and BRAT (r=0.23, $p < 0.05$). σ, the initial standard deviation of the pop representation, is weakly correlated both BART and BRAT performance (r=0.23, $p < 0.05$; r=0.23, $p < 0.05$ respectively), and not correlated with itself between the two tasks (r=-0.01, $p = 0.88$). γ modulates the reward value. γ is correlated with mean pumps in the BART (r=0.19, $p < 0.05$) and correlated with performance in the BRAT (r=0.17, $p < 0.05$). However, γ in the BART is correlated with γ in the BRAT (r=0.534, $p < 0.001$), implying that the participant's subjective perception of reward value may be similar in both tasks. τ, the perceived inter-choice difference, was weakly negatively correlated with performance in both BART and BRAT (r=-0.26, $p < 0.001$ and r=-0.38, $p < 0.001$, respectively). τ was correlated with itself between the two tasks, indicating that participants viewed the differences between pump values differently between the two tasks (r=0.29, $p < 0.001$). θ, the loss aversion parameter, was inversely correlated with performance in both BART (r = -0.67, $p < 0.001$) and BRAT (r =-0.64, $p < 0.001$). This is to be expected, as a greater loss aversion would upscale the value of the

potential loss when determining the expected value for a given choice, thereby inducing more conservative decision making (eq. 30). Interestingly enough, loss aversion between BART and BRAT was not strongly correlated (r=0.32, $p < 0.001$), implying that loss aversion between BART and BRAT can differ greatly within the same participant. This provides evidence that the perception of moving toward risk in the BART and moving away from risk in the BRAT are viewed through non-identical lenses. The fact that the γs for BART and BRAT were moderately correlated with one another provides further evidence of this, as participants' perception of point/monetary value did not drastically change between the tasks.

Parameter	BART Mean Pumps	BRAT Mean Pumps	BART X BRAT
h	r = 0.16, $p < 0.05$*	r = 0.37, $p < 0.001$**	r = 0.15, $p = 0.06$
λ	r = -0.14, $p = 0.08$	r = -0.02, $p = 0.79$	r = 0.04, $p = 0.62$
γ	r = 0.19, $p < 0.05$*	r = 0.17, p <0.05*	r = 0.54, $p < 0.001$**
τ	r = -0.26, $p < 0.001$**	r = -0.38, $p < 0.001$**	r = 0.29, $p < 0.001$**
μ	r = 0.53, $p < 0.001$**	r = 0.4, $p < 0.001$**	r = 0.23, $p < 0.05$*
σ	r = 0.23, $p < 0.05$*	r = 0.23, p <0.05*	r = -0.01, $p = 0.88$
θ	r = -0.67, $p < 0.001$**	r = -0.64, $p < 0.001$**	r = 0.32, $p < 0.001$**

Table 4: Correlating Independently-Fit Model Parameters with Performance: '*' indicates a correlation *p*-value below 0.05. '**' indicates a *p*-value below 0.001.

4.4.4 Validity evidence for BRAT

We conducted analyses to assess the validity of the BRAT and BART computational models as a measure of risk-taking. In terms of the BRAT, correlation analyses indicated that the BRAT parameter of loss aversion (θ) was positively associated with drug use (r = .19, $p < .05$), alcohol use (r = .19, p < .05), and fighting (r = .22, $p < .01$). Youths' standard deviation of the pop representation (σ) was positively related to drug use (r = .26, p < .01), marijuana use (r = .17, $p < .05$), alcohol use (r = .21, $p < .01$), and fighting (r = .24, $p < .01$). The BRAT

parameter of monetary value (γ) was also positively associated with drug use (r = .19, $p < .05$). In regard to the BART model, the loss aversion parameter (θ) was only positively related to fighting (r = .18, $p < .05$). Average number of pumps on the BRAT and BART were unrelated to any of youths' risk-taking behaviors.

4.5 BART vs BRAT: Investigating Cognitive Differences

The model proposed is able to re-create observed behavior reliably, and we have model parameters that are predictive of performance for both BART and BRAT while being only weakly correlated with each other across the two tasks. This provides evidence that participants are initially perceiving the popping probability spaces differently (via μ and σ) and weighing potential losses differently between the two tasks (via θ) whilst having a consistent perception of relative point value across both tasks (via γ). Since the model and data for both tasks is identical, we can fit the model to both tasks simultaneously. To do so, we create two sets of model parameters - one for each task. This doubles the number of model parameters. For example, the model has both γ_{BART} and γ_{BRAT}. We then systematically reduce parameter pairs into one parameter, testing every combination of task-exclusive/shared parameters (128 combinations in total). We followed the same procedure of model fitting as when fitting the model to BART and BRAT independently. We then calculated the Bayesian Predictive Information Criterion (BPIC) for each model variant for each subject. The more parameters a model has, the more it is penalized, making it a conservative assessment of goodness-of-fit (Ando, 2007). We calculated the within-subject mean-centered BPIC score for each model fit and ranked each model from lowest (best) average BPIC across all participants to highest mean BPIC.

The top 10 model variations (Figure 18) were not significantly different from one another, but patterns emerge in terms of which parameters are shared. All 10 models shared h and λ, with 7 sharing σ. γ was shared 6 times, with the top 4 models sharing γ. μ and τ were shared 5 times. θ was shared 4 times. Upon closer inspection, 2 out of the 4 times θ is shared, γ is not shared, implying that if θ is fixed between BART and BRAT, γ then takes up part of the role in differentiating value between the tasks. The model with the lowest average BPIC shared all parameters except for τ and θ. The second and third place models also did not share τ and θ, but then shared σ and μ respectively. The model that fit best for the most subjects was the model that shared all parameters between BART and BRAT except for loss aversion, θ.

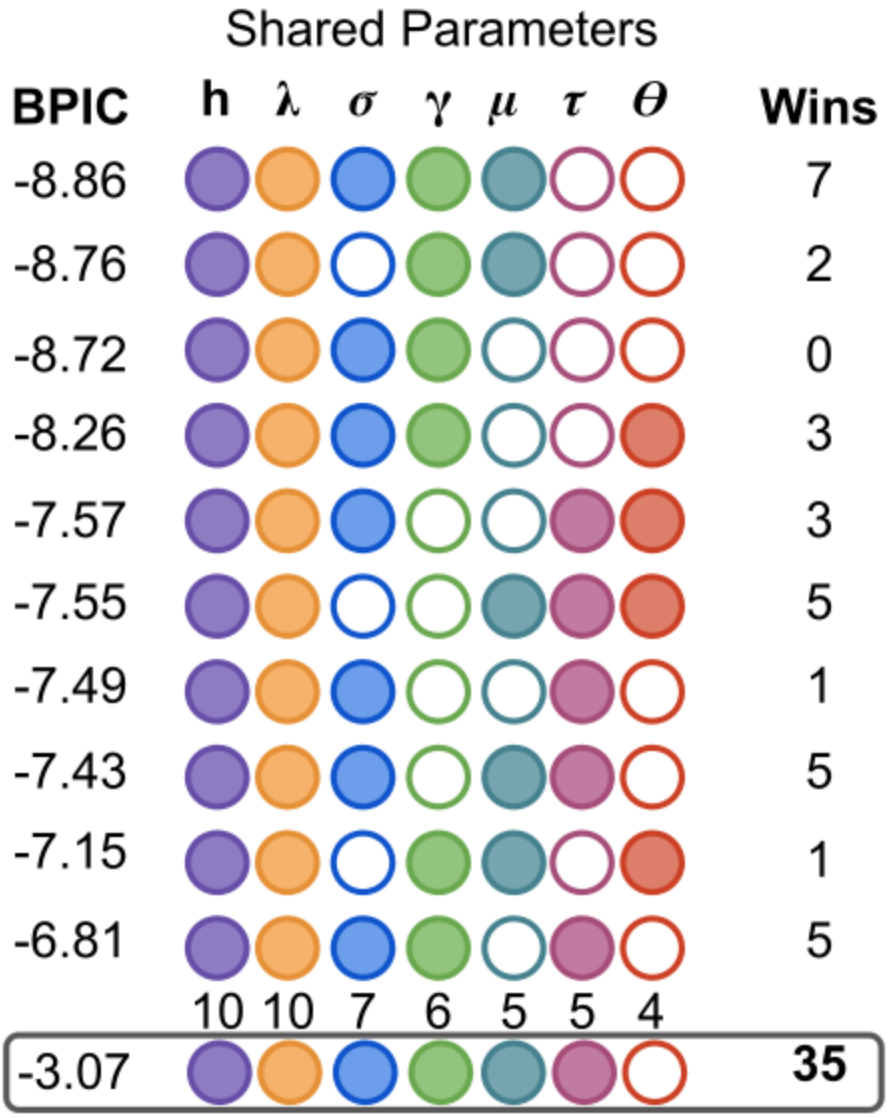


Figure 23: Combined Model Comparisons. Filled in circles indicate a parameter was shared while empty circles indicate the presence of two independent parameters (one for BART, one for BRAT). The first 10 models are listed, rank ordered from best (lowest BPIC score) to worse (higher BPIC score). The column on the right labeled "Wins" indicates how many subjects a given model fit best for. The model that fit the most participants the best is included at the bottom within the box.

These patterns of parameters being shared/not shared provides evidence that loss aversion θ and inter-choice differences τ were not the same within individuals during BART and BRAT. For the best-fitting model that shared all parameters except loss aversion and inter-choice

difference, θ and τ were significantly different from each other between the two tasks (two-sample t-tests: $t(DF) = 5.75$, $p < 0.001$ for θ and $t(DF) = -3.03$, $p = 0.003$ for τ). Loss aversion was significantly higher in the BRAT. At first glance, this may seem to contradict with the behavioral measures of performance, as participants left more air in the balloons in the BRAT than they added to the balloon in the BART (see figure 15). However, perceived inter-choice differences were greater in BRAT than in BART for most participants. We theorize there are two possible explanations for this: 1) Participants either perceive the distance needed to move to relative "safety" is smaller in the face of certain failure (ie, a false sense of security), or 2) when forced into a state of maximal risk, deflating the balloon in BRAT artificially gives the impression of "losing points", thereby making the balloon's value more salient and thus loss aversion is higher in BRAT than in BART.

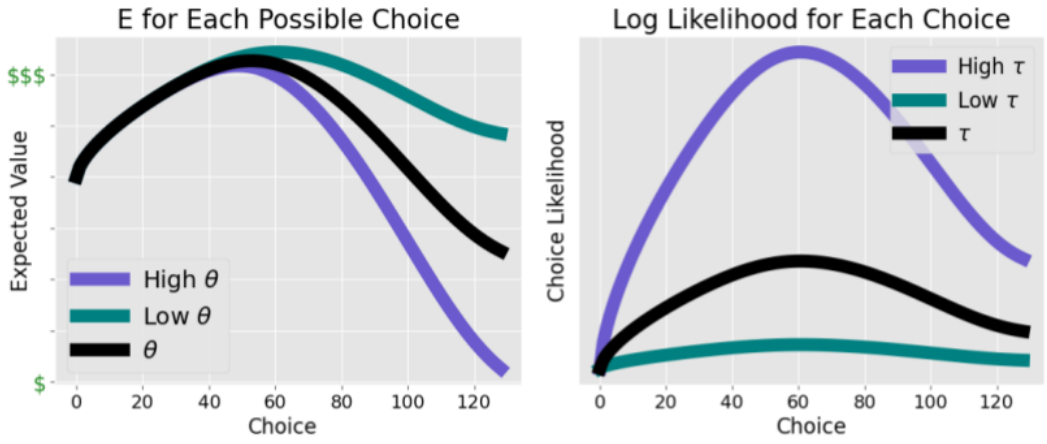


Figure 24: Impacts of Loss Aversion and Temperature Parameters on Decision Making. The left plot shows the effect higher and lower values of θ has on the perceived expected value. θ increases or decreases the magnitude of the expected value. The right plot shows how different values of τ can affect the overall likelihood of which choice will be selected, with lower values of τ flattening the likelihoods between options.

4.6 Examining Relationships Between Model Parameters and Behavioral Reports via Canonical Correlation Analysis.

The creators of the BRAT found that BART and BRAT performance related to different aspects of risk-related behaviors. BRAT performance was found to be related to risk-avoidant behaviors such as. With our model-based analysis, we fit the model to both BART & BRAT data independently and were able to reliably recreate observed participant responses. To examine the relationships between model parameters and independent measures (such as sleep, anxiety, etc.), we could perform a large number of pairwise correlations. However, not only would we have to account for a very large number of multiple comparisons (as we have parameter fits for two tasks), we would be ignoring the relationships between the independent measures as well as the model parameters at multiple levels.

In order to examine relationships between fit model parameters and behavioral measures, we must first examine the relationships within these two sets of data. Due to the nature of PLUM (like most other cognitive models) and the fitting process, the model parameters have various relationships with one another and can correlate with each other. Behavioral measures are also likely to correlate with one another, such as amount of sleep and anxiety. When attempting to find relationships between two datasets with within-set relationships, standard pairwise regressions and multiple regression approaches are not adequate. The within-set relationships violate the underlying assumptions about the data in order to perform multiple regression. In such situations, a possible solution is canonical correlation analysis (CCA). CCA finds the linear combinations of elements from both sets such that the correlation between the two sets is maximized in order to examine the relationships between the sets (Hotelling, 1936; Hardoon et al., 2004; Parkhomenko et al., 2009). The comparisons in question are as follows: 1) BART

parameters vs independent measures 2) BRAT parameters vs independent measures, and 3) Combined model BART & BRAT parameters vs independent measures.

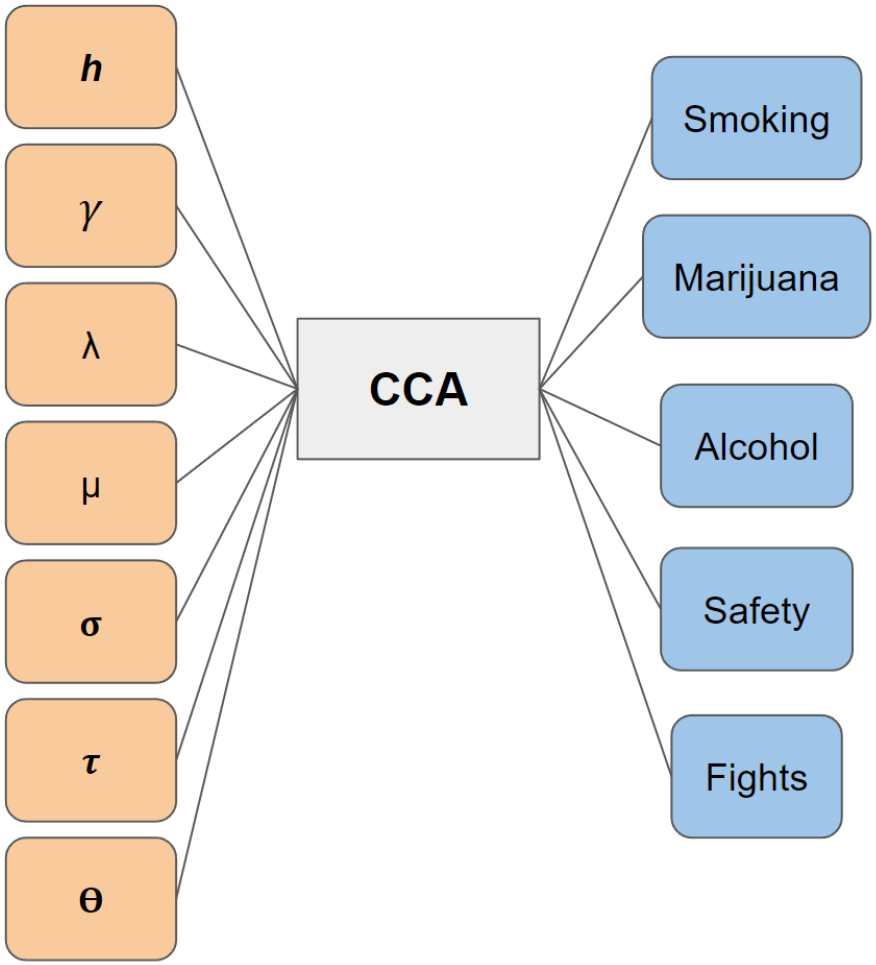


Figure 25: Examining canonical correlations between model parameters and behavioral measures.

Comparison 1 & 2 will provide insight as to whether the proposed cognitive mechanisms in the model are weighted differently between the two tasks whilst taking into consideration the within-task parameter correlations. The results of this CCA can then be compared to the pairwise correlations presented previously, which show how some parameters correlate between the two

tasks while others do not. We can also compare these CCA results to the results of the combined model BPIC comparisons, in which we found the best model to be the one in which all parameter values are shared between the two tasks except for loss aversion and perception of interchoice distance.

To conduct this analysis, we utilized the Python library CCA-Zoo (Chapman & Wang, 2021). After fitting a CCA, we used permutation testing to assess the significance of each model by randomly shuffling both sets of data 1000 times, generating a distribution of canonical correlations (components). Then using a null hypothesis test, we assess whether or not each component is significant where the null hypothesis is that there is no relationship between the model parameters and independent measures.

4.6.1 CCA Results

Significant results will be reported here. For all CCA results, please see appendix S3. We did not find any significant relationships between the model parameters and the independent measures when the model was fit to BART and BRAT independently. CCA does show a significant relationship between the combined model best-fitting parameters and independent measures. After permutation testing the 9^{th} component of this CCA model shows a statistically significant relationship between the two sets ($r = 0.14$, $p = 0.027$) (figure X). The figure in

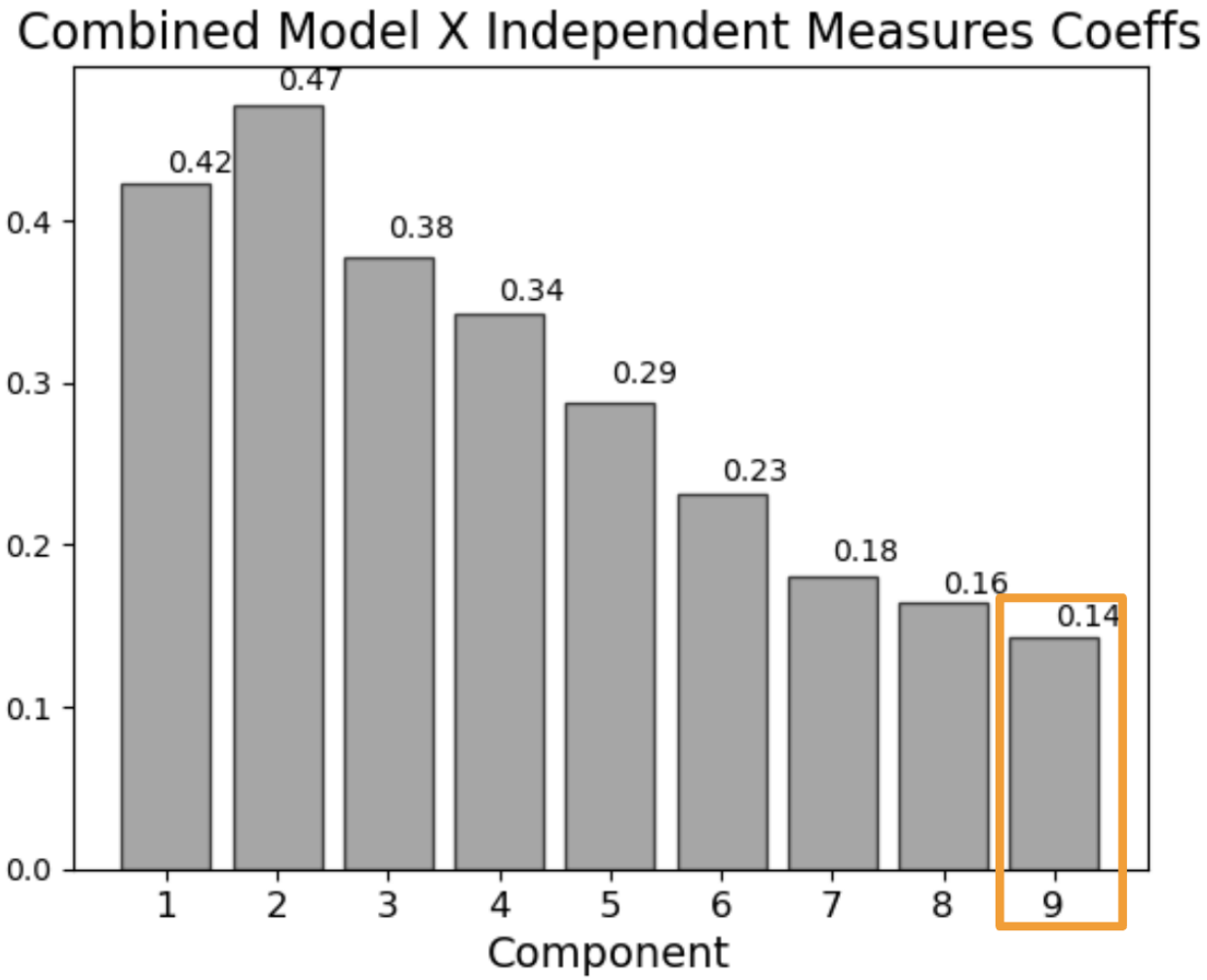


Figure 26: Combined Model Best-Fitting Parameters and Independent Measures. This barplot shows the correlation coefficients of each CCA model component. These coefficients represent the correlation between the combined model best-fitting parameters and the independent measures as a whole. Of the 9 components in this model, only one component was significant (outlined in orange).

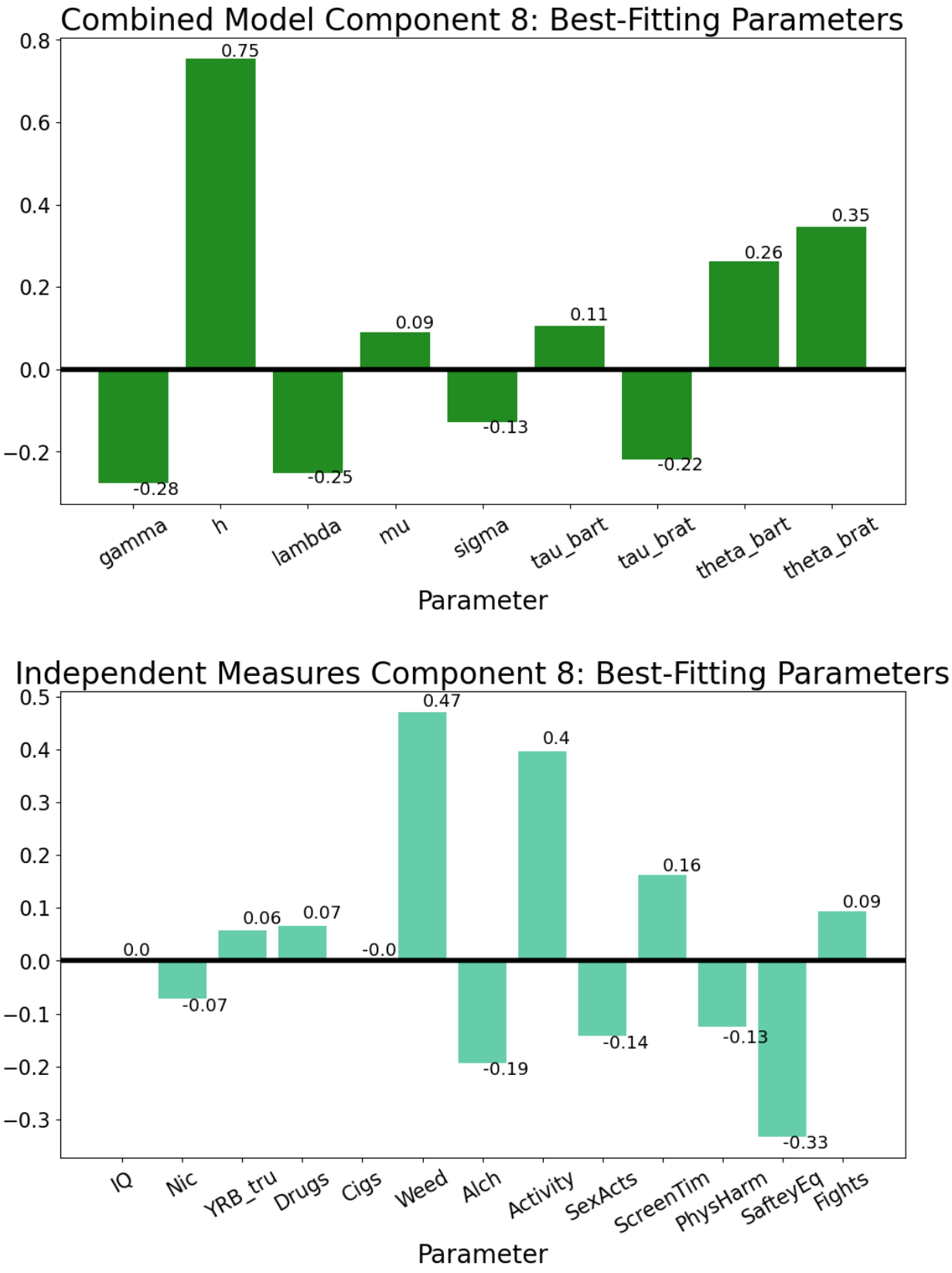


Figure 27: Features of the significant component. This figure shows the weights of each feature

for both the best-fitting combined model parameters and independent measure for the significant component.

The magnitude (absolute value) of the feature weights indicates the strength of the relationship between each feature and the linear combination (also known the canonical variate) it contributes to. The higher the absolute value of a weight, the stronger it is associated with the linear combination between sets. The direction (positive or negative) indicates if a feature is positively or negatively associated with the canonical variate. In this instance, for the model parameters, *h*, θ_{BRAT}, γ, and θ_{BART} had the greatest weight magnitudes (0.75, 0.35, -0.28, and, 0.26 respectively). In the model comparison, the top performing model was the one that shared all parameters except for τ and θ. This provides some evidence that these two parameters of inter-choice difference and loss aversion are different between the two tasks, and that loss aversion in the BRAT may map onto manifested behaviors differently than loss aversion in BART. Marijuanna usage had the most significant influence on the canonical correlations between the sets (0.47), followed by risky activity (0.40) and sense of safety (-0.33).

Discussion

The context of risk-ambiguous environments greatly impacts a participant's evaluation process. BART and BRAT are essentially the same task. Participants select a certain number of pumps of air to be in the balloon, observe the result, and update their underlying representations of the outcomes. The only thing that differs between BART from BRAT is the way the task is framed. In BART, the participant moves from a position of low risk and no reward (0 pumps of air in the balloon at onset) into a state of greater reward and greater risk. In the BRAT, the participant is moving from a position of certain failure and no reward to a position of lower risk

and variable reward. Superficially, it makes sense to think that an agent would behave identically in both tasks due to the essence of both tasks being to select a number of pumps to be in a balloon and add those points to a score. However, the results presented here demonstrate that risk-taking and risk avoidant behaviors may not have a one-to-on relationship. Though these two behaviors may utilize the same/similar cognitive mechanisms, these mechanisms may be weighted differently.

This asymmetry between these behaviors creates an opportunity for more research to be conducted that investigates risk-avoidant behaviors both independently as well as in conjunction with examining risk-taking behaviors. A problem with studying risk-related behaviors is that the behavior is often observed and quantified in a vacuum, ignoring the context driving the behavior. This could lead to misidentifying risk-avoidant behaviors as risk-taking ones and vice versa. Behaviors deemed to be risk-taking can actually be the manifestation of avoiding some other stimulus (Kashdan et al., 2008). For example, a youth could abuse drugs not because they are willing to take the risk of possible bodily harm so that they can experience the drug's effects or to gain standing with some group, but because they are avoiding being ostracized from said group. It is possible that this was in part why CCA analysis of the independently fit model to BART and BRAT was unsuccessful in finding any significant linear combinations between best-fitting parameters while the combined model CCA found a significant component. Perhaps more risk-avoidance and risk-taking tasks should be administered in tandem or fit in a united model of risky decision making that can accommodate both paradigms. Herein lies an opportunity for the dissection of risk-related behaviors via generative modeling. A generative model can provide a more detailed insight into the machinations of the cognitive processes that manifested outward behaviors and be useful for accurately labeling them as risk-taking or risk-avoidant. Given the

scant scientific work investigating risk avoidance in general, it is likely that there is much to discover in this regard.

The model presented here performs admirably, but there is still room for improvement. Currently, the model is able to capture behavior when a simulated agent consistently makes the same, relatively low risk number of pumps several times in a row, then suddenly switches to making consistent high-risk choices. However, the model is inflexible in its ability to account for the opposite response pattern of moving from consistent high-risk choices to low-risk choices. Figure 21 demonstrates this inconsistency. A possible method for improving the model is to instantiate an initial cumulative pop probability distribution as the prior perceived likelihood of the balloon popping at each possible choice p_0. We would remove μ and σ parameters, and instead have both of the observation probability mass functions be initially uniform distributions. The initial pop probability would utilize a beta distribution given shape via free parameter ρ. After each choice, the observation distributions *O* would be updated via kernel K, but the degree to which *O* would be updated is modulated by a mixture mechanism with p_0. In effect, this mixture mechanism is suggesting that the participant is combining their initial bias of the pop probability with observed outcomes. The degree to which the participant relies upon or abandons this initial bias would be parameterized and impacted by the number trials, with the initial bias having less impact as more observations have been made.

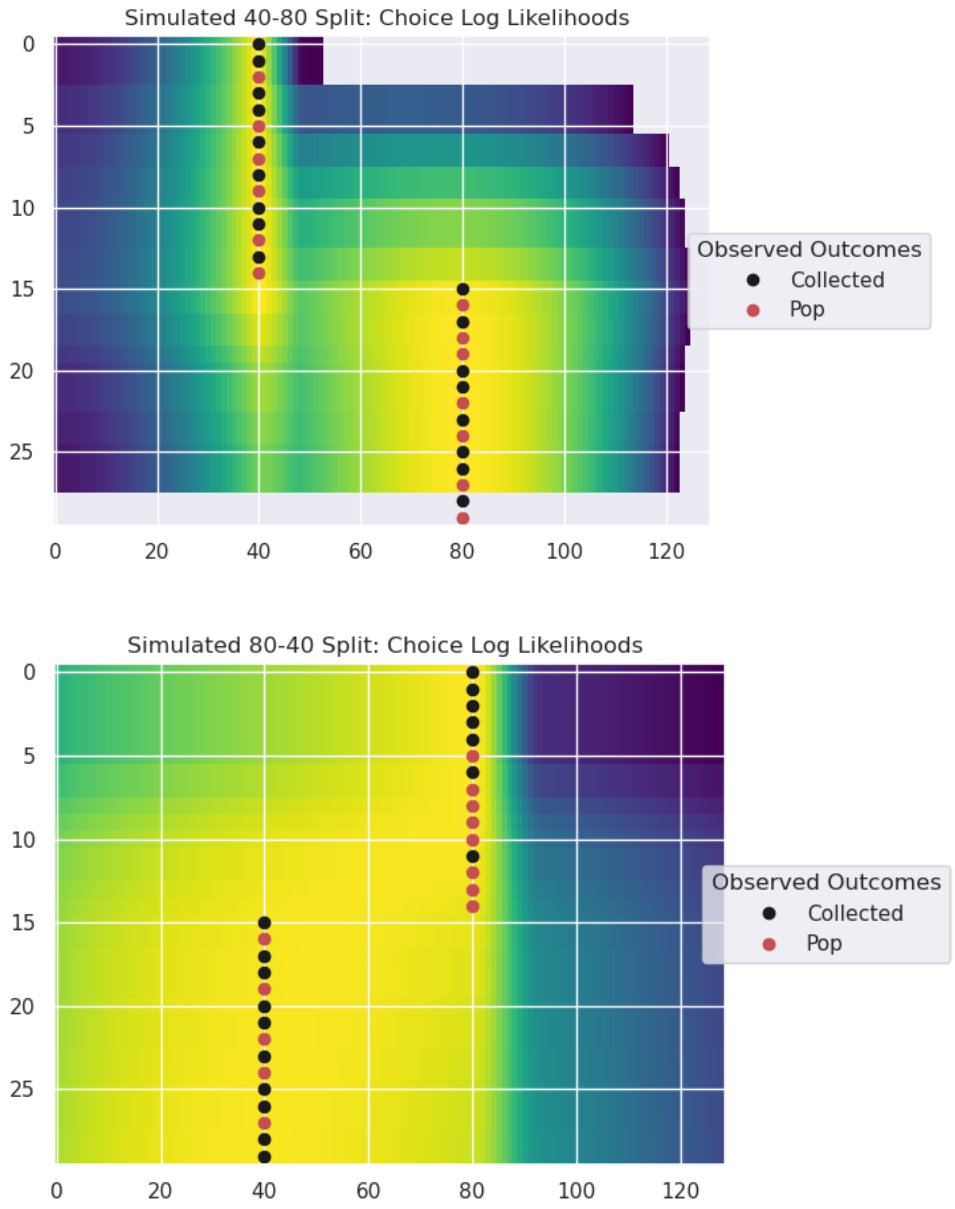


Figure 28: Simulated Bifurcating Agents. The top panel shows a simulated agent selecting 40 pumps to be in the balloon for the first half of the session, followed by 80 pumps for the remaining 15 balloons. We see that the model-estimated choice likelihoods are able to reasonably capture the simulated behavior. However, when reversing the simulation (80 pumps followed by 40 pumps), the model is unable to adjust to this behavior.

5. Conclusion

In this dissertation, I have presented the development and evaluation of cognitive generative models of risk-related decision making under ambiguity using the BART paradigm. With prospect theory as a backbone, we have created and tested several risky decision models geared to two different kinds of BART: One in which participants make each pump choice individually, the other when participants select the number of pumps ahead of time. Though this distinction may seem arbitrary, the implications it carries when developing computational models cannot be understated. Both of these models are able to reliably recreate observed behavior, providing evidence that they are tapping into possible latent cognitive mechanisms. In a preliminary analysis, we've demonstrated how to use model-based fMRI analysis to assess a novel model of the BART (PLUM) and investigate brain regions that may be involved in the risky decision-making process. Testing models against neural data lead to development of better models and better task designs for specific hypotheses.We developed a model of risk taking that is able to capture both risk-taking and risk-avoidance behavior in a task in which very little information about the likelihoods of failure is gleaned. With this model, we were able to uncover asymmetries in risk-related behavior that would normally not be seen via general performance metrics. Generative computational models provide the means to provide a window into cognition. They are only limited by one's imagination and available computing power.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics, 8. https://www.frontiersin.org/articles/10.3389/fninf.2014.00014

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

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. Frontiers in Psychology, 5. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00849

Aklin, W. M., Lejuez, C. W., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2005). Evaluation of behavioral measures of risk taking propensity with inner city adolescents. Behaviour Research and Therapy, 43(2), 215–228. https://doi.org/10.1016/j.brat.2003.12.007

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

Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., & Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. Human Brain Mapping, 36(2), 449–462. https://doi.org/10.1002/hbm.22639

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

Brener, N. D., Simon, T. R., Anderson, M., Barrios, L. C., & Small, M. L. (2002). Effect of the incident at Columbine on students' violence- and suicide-related behaviors. American Journal of Preventive Medicine, 22(3), 146–150. https://doi.org/10.1016/S0749-3797(01)00433-0

Brown, J. W., & Alexander, W. H. (2017). Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making. Journal of Cognitive Neuroscience, 29(10), 1656–1673. https://doi.org/10.1162/jocn_a_01140

Brown, J. W., & Braver, T. S. (2005). Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Brown, J. W., & Braver, T. S. (2008). A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. Brain Research, 1202, 99–108. https://doi.org/10.1016/j.brainres.2007.06.080

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

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. Journal of Risk and Uncertainty, 5(4), 325–370. https://doi.org/10.1007/BF00122575

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://www.frontiersin.org/articles/10.3389/fnbeh.2018.00111

Chapman, J., & Wang, H.-T. (2021). CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework. Journal of Open Source Software, 6(68), 3823. https://doi.org/10.21105/joss.03823

Coffey, S. F., Schumacher, J. A., Baschnagel, J. S., Hawk, L. W., & Holloman, G. (2011). Impulsivity and risk-taking in borderline personality disorder with and without substance use disorders. Personality Disorders: Theory, Research, and Treatment, 2, 128–141. https://doi.org/10.1037/a0020574

Collado, A., Felton, J. W., MacPherson, L., & Lejuez, C. W. (2014). Longitudinal trajectories of sensation seeking, risk taking propensity, and impulsivity across early to middle adolescence. Addictive Behaviors, 39(11), 1580–1588. https://doi.org/10.1016/j.addbeh.2014.01.024

Coon, J., & Lee, M. D. (2022). A Bayesian method for measuring risk propensity in the Balloon Analogue Risk Task. Behavior Research Methods, 54(2), 1010–1026. https://doi.org/10.3758/s13428-021-01634-1

Cox, R. W. (1996a). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, an International Journal, 29(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W. (1996b). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. Computers and Biomedical Research, 29(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017a). FMRI clustering and false-positive rates. Proceedings of the National Academy of Sciences of the United States of America, 114(17), E3370–E3371. https://doi.org/10.1073/pnas.1614961114

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017b). FMRI Clustering in AFNI: False-Positive Rates Redux. Brain Connectivity, 7(3), 152–171. https://doi.org/10.1089/brain.2016.0475

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. NMR in Biomedicine, 10(4–5), 171–178. https://doi.org/10.1002/(sici)1099-1492(199706/08)10:4/5<171::aid-nbm453>3.0.co;2-l

Crowley, M. J., van Noordt, S. J. R., Castagna, P. J., Vaca, F. E., Wu, J., Lejuez, C. W., & Mayes, L. C. (2022). Avoidance in Adolescence: The Balloon Risk Avoidance Task (BRAT). Journal of Psychopathology and Behavioral Assessment, 44(2), 297–311. https://doi.org/10.1007/s10862-021-09928-6

de Groot, K. (2020). Burst Beliefs – Methodological Problems in the Balloon Analogue Risk Task and Implications for Its Use. Journal of Trial and Error, 1(1). https://doi.org/10.36850/mr1

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, Biases, and Rational Decision-Making in the Human Brain. Science, 313(5787), 684–687. https://doi.org/10.1126/science.1128356

DeWitt, S. J., Aslan, S., & Filbey, F. M. (2014). Adolescent risk-taking and resting state functional connectivity. *Psychiatry Research: Neuroimaging*, *222*(3), 157–164. https://doi.org/10.1016/j.pscychresns.2014.03.009

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

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., McManus, T., Chyen, D., Whittle, L., Lim, C., & Wechsler, H. (2012). Youth Risk Behavior Surveillance—United States, 2011. Morbidity and Mortality Weekly Report: Surveillance Summaries, 61(4), 1–162.

Eckstrand, K. L., Choukas-Bradley, S., Mohanty, A., Cross, M., Allen, N. B., Silk, J. S., Jones, N. P., & Forbes, E. E. (2017). Heightened activity in social reward networks is associated with adolescents' risky sexual behaviors. *Developmental Cognitive Neuroscience*, *27*, 1–9. https://doi.org/10.1016/j.dcn.2017.07.004

Figner, B., Mackinlay, R. J., Wilkening, F., Weber, E. U., & Figner (corresponding, B. (2008). Affective and Deliberative 1 Running Head: AFFECTIVE AND DELIBERATIVE RISKY DECISION MAKING Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task.

Fiore, V. G., & Gu, X. (2021). Similar network compositions, but distinct neural dynamics underlying belief updating in environments with and without explicit outcomes (p. 794669). bioRxiv. https://doi.org/10.1101/794669

Fukunaga, R., Brown, J. W., & Bogg, T. (2012). Decision making in the Balloon Analogue Risk Task (BART): Anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. Cognitive, Affective, & Behavioral Neuroscience, 12(3), 479–490. https://doi.org/10.3758/s13415-012-0102-1

Gallistel, C., Krishan, M., Liu, Y., Miller, R., & Latham, P. (2014). The Perception of Probability. Psychological Review, 121, 96–123. https://doi.org/10.1037/a0035232

Gärling, T., Kirchler, E., Lewis, A., & van Raaij, F. (2009). Psychology, Financial Decision Making, and Financial Crises. Psychological Science in the Public Interest, 10(1), 1–47. https://doi.org/10.1177/1529100610378437

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

Gläscher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. WIREs Cognitive Science, 1(4), 501–510. https://doi.org/10.1002/wcs.57

Grinblatt, M., & Han, B. (2005). Prospect theory, mental accounting, and momentum. Journal of Financial Economics, 78(2), 311–339. https://doi.org/10.1016/j.jfineco.2004.10.006

Gu, R., Zhang, D., Luo, Y., Wang, H., & Broster, L. S. (2018). Predicting risk decisions in a modified Balloon Analogue Risk Task: Conventional and single-trial ERP analyses.

Cognitive, Affective, & Behavioral Neuroscience, 18(1), 99–116. https://doi.org/10.3758/s13415-017-0555-3

Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical Correlation Analysis: An Overview with Application to Learning Methods. Neural Computation, 16(12), 2639–2664. https://doi.org/10.1162/0899766042321814

Helion, C., Krueger, S. M., & Ochsner, K. N. (2019). Chapter 14—Emotion regulation across the life span. In M. D'Esposito & J. H. Grafman (Eds.), Handbook of Clinical Neurology (Vol. 163, pp. 257–280). Elsevier. https://doi.org/10.1016/B978-0-12-804281-6.00014-8

Hiser, J., & Koenigs, M. (2018). The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. Biological Psychiatry, 83(8), 638–647. https://doi.org/10.1016/j.biopsych.2017.10.030

Hotelling, H. (1992). Relations Between Two Sets of Variates. In S. Kotz & N. L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution (pp. 162–190). Springer. https://doi.org/10.1007/978-1-4612-4380-9_14

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under Uncertainty: Probabilistic Context Influences Activation of Prefrontal and Parietal Cortices. The Journal of Neuroscience, 25(13), 3304–3311. https://doi.org/10.1523/JNEUROSCI.5070-04.2005

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

Jones, H. A., & Lejuez, C. W. (2005). Personality Correlates of Caffeine Dependence: The Role of Sensation Seeking, Impulsivity, and Risk Taking. Experimental and Clinical Psychopharmacology, 13, 259–266. https://doi.org/10.1037/1064-1297.13.3.259

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

Kashdan, T. B., Elhai, J. D., & Breen, W. E. (2008). Social anxiety and disinhibition: An analysis of curiosity and social rank appraisals, approach–avoidance conflicts, and disruptive risk-taking behavior. Journal of Anxiety Disorders, 22(6), 925–939. https://doi.org/10.1016/j.janxdis.2007.09.009

Kessler, L., Hewig, J., Weichold, K., Silbereisen, R. K., & Miltner, W. H. R. (2017). Feedback negativity and decision-making behavior in the Balloon Analogue Risk Task (BART) in adolescents is modulated by peer presence. Psychophysiology, 54(2), 260–269. https://doi.org/10.1111/psyp.12783

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: A review of recent literature. Current Opinion in Psychiatry, 20(4), 359. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Kóbor, A., Kardos, Z., Takács, Á., Éltető, N., Janacsek, K., Tóth-Fáber, E., Csépe, V., & Nemeth, D. (2021). Adaptation to recent outcomes attenuates the lasting effect of initial experience on risky decisions. Scientific Reports, 11(1), Article 1. https://doi.org/10.1038/s41598-021-89456-1

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

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. Journal of Behavioral Decision Making, 27(1), 20–36. https://doi.org/10.1002/bdm.1784

Lee, T. M. C., Chan, C. C. H., Leung, A. W. S., Fox, P. T., & Gao, J.-H. (2009). Sex-Related Differences in Neural Activity during Risk Taking: An fMRI Study. Cerebral Cortex, 19(6), 1303–1312. https://doi.org/10.1093/cercor/bhn172

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. https://doi.org/10.1016/S0140-1971(03)00036-8

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

Leslie, M., Leppanen, J., Paloyelis, Y., Nazar, B. P., & Treasure, J. (2019). The influence of oxytocin on risk-taking in the balloon analogue risk task among women with bulimia nervosa and binge eating disorder. Journal of Neuroendocrinology, 31(8), e12771. https://doi.org/10.1111/jne.12771

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural Representation of Subjective Value Under Risk and Ambiguity. Journal of Neurophysiology, 103(2), 1036–1047. https://doi.org/10.1152/jn.00853.2009

Li, X., Pan, Y., Fang, Z., Lei, H., Zhang, X., Shi, H., Ma, N., Raine, P., Wetherill, R., Kim, J. J., Wan, Y., & Rao, H. (2020). Test-retest reliability of brain responses to risk-taking during the balloon analogue risk task. NeuroImage, 209, 116495. https://doi.org/10.1016/j.neuroimage.2019.116495

Lorian, C. N., & Grisham, J. R. (2011). Clinical implications of risk aversion: An online study of risk-avoidance and treatment utilization in pathological anxiety. Journal of Anxiety Disorders, 25(6), 840–848. https://doi.org/10.1016/j.janxdis.2011.04.008

Molins, F., Paz, M., Rozman, L., Ben Hassen, N., & Serrano, M. Á. (2022). Stressed individuals exhibit pessimistic bursting beliefs and a lower risk preference in the balloon analogue risk task. Physiology & Behavior, 256, 113953. https://doi.org/10.1016/j.physbeh.2022.113953

Monti, M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. Frontiers in Human Neuroscience, 5. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00028

Moore, M. N., Salk, R. H., Van Hulle, C. A., Abramson, L. Y., Hyde, J. S., Lemery-Chalfant, K., & Goldsmith, H. H. (2013). Genetic and Environmental Influences on Rumination, Distraction, and Depressed Mood in Adolescence. Clinical Psychological Science, 1(3), 316–322. https://doi.org/10.1177/2167702612472884

Park, H., Yang, J., Vassileva, J., & Ahn, W.-Y. (2021). Development of a novel computational model for the Balloon Analogue Risk Task: The exponential-weight

mean–variance model. Journal of Mathematical Psychology, 102, 102532. https://doi.org/10.1016/j.jmp.2021.102532

Parkhomenko, E., Tritchler, D., & Beyene, J. (2009). Sparse Canonical Correlation Analysis with Application to Genomic Data Integration. Statistical Applications in Genetics and Molecular Biology, 8(1). https://doi.org/10.2202/1544-6115.1406

Paulus, M. P., Hozack, N., Zauscher, B., McDowell, J. E., Frank, L., Brown, G. G., & Braff, D. L. (2001). Prefrontal, Parietal, and Temporal Cortex Networks Underlie Decision-Making in the Presence of Uncertainty. NeuroImage, 13(1), 91–100. https://doi.org/10.1006/nimg.2000.0667

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. Journal of Experimental Psychology – Learning Memory and Cognition, 167–185.

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

Pleskac, T. J., & Wershbale, A. (2014). Making assessments while taking repeated risks: A pattern of multiple response pathways. Journal of Experimental Psychology: General, 143, 142–162. https://doi.org/10.1037/a0031106

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. Neuron, 51(3), 381–390. https://doi.org/10.1016/j.neuron.2006.06.024

Rao, H., Korczykowski, M., Pluta, J., Hoang, A., & Detre, J. A. (2008). Neural correlates of voluntary and involuntary risk taking in the human brain: An fMRI Study of the

Balloon Analog Risk Task (BART). NeuroImage, 42(2), 902–910. https://doi.org/10.1016/j.neuroimage.2008.05.046

Rao, L.-L., Zhou, Y., Zheng, D., Yang, L.-Q., & Li, S. (2018). Genetic Contribution to Variation in Risk Taking: A Functional MRI Twin Study of the Balloon Analogue Risk Task. Psychological Science, 29(10), 1679–1691. https://doi.org/10.1177/0956797618779961

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

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), Article 2. https://doi.org/10.1038/npp.2013.218

Rudorf, S., Preuschoff, K., & Weber, B. (2012). Neural Correlates of Anticipation Risk Reflect Risk Preferences. Journal of Neuroscience, 32(47), 16683–16692. https://doi.org/10.1523/JNEUROSCI.4235-11.2012

Schonberg, T., Fox, C., Mumford, J., Congdon, E., Trepel, C., & Poldrack, R. (2012). Decreasing Ventromedial Prefrontal Cortex Activity During Sequential Risk-Taking: An fMRI Investigation of the Balloon Analog Risk Task. Frontiers in Neuroscience, 6. https://www.frontiersin.org/articles/10.3389/fnins.2012.00080

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

Shoham, R., Sonuga-Barke, E. J. S., Aloni, H., Yaniv, I., & Pollak, Y. (2016). ADHD-associated risk taking is linked to exaggerated views of the benefits of positive outcomes. Scientific Reports, 6(1), Article 1. https://doi.org/10.1038/srep34833

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive Coding of Reward Value by Dopamine Neurons. Science, 307(5715), 1642–1645. https://doi.org/10.1126/science.1105370

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

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. Psychological Review, 118(4), 583–613. https://doi.org/10.1037/a0025191

Tversky, A., & Kahneman, D. (n.d.). Advances in prospect theory: Cumulative representation of uncertainty. 27.

van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2011). Cognitive model decomposition of the BART: Assessment and application. Journal of Mathematical Psychology, 55(1), 94–105. https://doi.org/10.1016/j.jmp.2010.08.010

Wallsten, T. S., & Budescu, D. V. (1983). State of the Art—Encoding Subjective Probabilities: A Psychological and Psychometric Review. Management Science, 29(2), 151–173. https://doi.org/10.1287/mnsc.29.2.151

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

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005b). 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

Wang, M., Zhang, S., Suo, T., Mao, T., Wang, F., Deng, Y., Eickhoff, S., Pan, Y., Jiang, C., & Rao, H. (2022). Risk-taking in the human brain: An activation likelihood estimation meta-analysis of the balloon analog risk task (BART). Human Brain Mapping, 43(18), 5643–5657. https://doi.org/10.1002/hbm.26041

Weber, R., Mangus, J. M., & Huskey, R. (2015). Brain Imaging in Communication Research: A Practical Guide to Understanding and Evaluating fMRI Studies.

Communication Methods and Measures, 9(1–2), 5–29. https://doi.org/10.1080/19312458.2014.999754

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence—Second Edition. https://doi.org/10.1037/t15171-000

Weichart, E. R., Darby, K. P., Fenton, A. W., Jacques, B. G., Kirkpatrick, R. P., Turner, B. M., & Sederberg, P. B. (2021). Quantifying mechanisms of cognition with an experiment and modeling ecosystem. Behavior Research Methods, 53(5), 1833–1856. https://doi.org/10.3758/s13428-020-01534-w

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

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. NeuroImage, 91, 412–419. https://doi.org/10.1016/j.neuroimage.2013.12.058

Wright, R. J., & Rakow, T. (2017). Don't sweat it: Re-examining the somatic marker hypothesis using variants of the Balloon Analogue Risk Task. Decision, 4(1), 52–65. https://doi.org/10.1037/dec0000055

Wu, S., Sun, S., Camilleri, J. A., Eickhoff, S. B., & Yu, R. (2021). Better the devil you know than the devil you don't: Neural processing of risk and ambiguity. NeuroImage, 236, 118109. https://doi.org/10.1016/j.neuroimage.2021.118109

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

Zhang, J., Bogacz, R., & Holmes, P. (2009). A comparison of bounded diffusion models for choice in time controlled tasks. Journal of Mathematical Psychology, 53(4), 231–241. https://doi.org/10.1016/j.jmp.2009.03.001

Zhou, R., Myung, J. I., & Pitt, M. A. (2021). The scaled target learning model: Revisiting learning in the balloon analogue risk task. Cognitive Psychology, 128, 101407. https://doi.org/10.1016/j.cogpsych.2021.101407

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

S2.1: fMRI Cluster Tables

Significant clusters from the resulting neural analysis from chapter 3. Keys: *L = Left, R = Right, Inf = Inferior, Sup = Superior, Mid = Middle, Ant = Anterior*

Subjective Pop Probability

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	33.5	21.5	9.5	6.960913	5464	Frontal_Inf_Tri_R
1a	43.5	19.5	-4.5	6.188147		Insula_R
1b	31.5	25.5	-6.5	4.985929		Frontal_Inf_Orb_R
1c	45.5	21.5	9.5	4.751680		Frontal_Inf_Tri_R
2	31.5	-4.5	63.5	6.422814	22752	Frontal_Sup_R
2a	1.5	21.5	37.5	5.956784		Cingulum_Mid_R
2b	27.5	-4.5	59.5	5.936561		Frontal_Sup_R
2c	25.5	-6.5	51.5	5.886572		Precentral_R
3	-30.5	21.5	7.5	5.985740	2200	Insula_L
3a	-36.5	11.5	9.5	4.742115		Insula_L
3b	-40.5	17.5	-2.5	4.521940		Insula_L
3c	-30.5	23.5	-6.5	4.276510		Insula_L
4	-24.5	-62.5	59.5	5.732955	2352	Parietal_Sup_L
4a	-28.5	-54.5	65.5	5.025360		Parietal_Sup_L
4b	-16.5	-64.5	59.5	5.008266		Precuneus_L
4c	-20.5	-58.5	65.5	4.939154		Parietal_Sup_L
5	27.5	-56.5	67.5	5.676863	7712	Parietal_Sup_R
5a	41.5	-26.5	39.5	5.517349		Postcentral_R
5b	45.5	-40.5	57.5	5.256649		Parietal_Sup_R
5c	11.5	-56.5	63.5	5.253322		Parietal_Sup_R
6	35.5	47.5	29.5	5.629131	4688	Frontal_Mid_R
6a	33.5	43.5	37.5	5.347194		Frontal_Mid_R
6b	37.5	49.5	21.5	4.655511		Frontal_Mid_R
6c	45.5	49.5	13.5	4.600538		Frontal_Mid_R
7	-30.5	-8.5	51.5	5.443968	5904	Precentral_L
7a	-14.5	-10.5	71.5	5.295216		Supp_Motor_Area_L
7b	-22.5	-2.5	55.5	5.210890		Frontal_Sup_L
7c	-30.5	-16.5	61.5	4.913141		Precentral_L
8	-40.5	-42.5	61.5	5.282494	4296	Postcentral_L
8a	-48.5	-32.5	49.5	5.260949		Postcentral_L
8b	-46.5	-40.5	59.5	5.125327		Postcentral_L
8c	-40.5	-40.5	43.5	5.009447		Parietal_Inf_L
9	-6.5	7.5	-0.5	5.089802	880	Caudate_L
9a	-12.5	9.5	-8.5	4.678853		Putamen_L
9b	-8.5	3.5	7.5	4.248066		Caudate_L
9c	-16.5	11.5	-4.5	4.037984		Putamen_L
10	29.5	-76.5	37.5	5.053495	512	Occipital_Mid_R

Continued on next page:

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
11	-44.5	-50.5	-32.5	4.881114	1224	Cerebelum_Crus1_L
11a	-38.5	-56.5	-28.5	4.704007		Cerebelum_Crus1_L
11b	-34.5	-46.5	-36.5	4.620439		Cerebelum_Crus1_L
11c	-42.5	-58.5	-26.5	4.605147		Cerebelum_Crus1_L
12	-52.5	-82.5	-0.5	4.808173	520	undefined
13	19.5	15.5	-8.5	4.732319	760	Putamen_R
13a	9.5	17.5	7.5	4.325260		Caudate_R
13b	11.5	11.5	1.5	4.006930		Caudate_R
13c	7.5	5.5	5.5	3.984350		Caudate_R
14	-8.5	-76.5	-18.5	4.661213	208	Cerebelum_6_L
15	7.5	-8.5	9.5	4.635930	400	Thalamus_R
15a	3.5	-2.5	5.5	3.867853		undefined
16	55.5	-12.5	3.5	4.461817	120	Temporal_Sup_R
17	57.5	-28.5	41.5	4.453650	376	SupraMarginal_R
17a	63.5	-22.5	45.5	3.764379		undefined
18	-48.5	-74.5	9.5	4.414696	328	Occipital_Mid_L
18a	-42.5	-68.5	5.5	4.372835		Occipital_Mid_L
19	29.5	-78.5	19.5	4.385517	720	Occipital_Mid_R
19a	29.5	-84.5	29.5	3.961537		Occipital_Mid_R
19b	39.5	-84.5	31.5	3.881809		undefined
19c	27.5	-74.5	25.5	3.674534		Occipital_Mid_R
20	63.5	-16.5	35.5	4.374826	456	Postcentral_R
21	-22.5	-74.5	49.5	4.339939	304	Parietal_Sup_L
21a	-16.5	-68.5	57.5	4.302666		Parietal_Sup_L
22	-28.5	-48.5	43.5	4.331262	128	Parietal_Inf_L
23	41.5	-72.5	17.5	4.249827	208	Occipital_Mid_R
24	-8.5	-58.5	63.5	4.247816	280	Precuneus_L
24a	-16.5	-56.5	65.5	4.193982		Precuneus_L
25	47.5	3.5	1.5	4.231605	128	Rolandic_Oper_R
26	45.5	-60.5	11.5	4.215615	216	Temporal_Mid_R
27	-28.5	33.5	25.5	4.205963	208	Frontal_Mid_L
27a	-34.5	41.5	27.5	3.563049		Frontal_Mid_L
28	-0.5	-2.5	35.5	4.147858	160	Cingulum_Mid_L
29	-40.5	3.5	21.5	4.126805	408	Frontal_Inf_Oper_L
29a	-42.5	3.5	29.5	3.994030		Precentral_L
29b	-50.5	5.5	27.5	3.850725		Precentral_L
30	11.5	43.5	11.5	4.036452	112	Cingulum_Ant_R
31	-24.5	-88.5	17.5	4.004548	128	Occipital_Mid_L
32	-58.5	-2.5	39.5	3.903448	152	Postcentral_L
33	1.5	-28.5	-4.5	3.893854	96	undefined
34	-34.5	43.5	37.5	3.888006	104	Frontal_Mid_L
35	9.5	-42.5	51.5	3.832263	152	Precuneus_R
36	-28.5	-70.5	35.5	3.809221	192	Occipital_Mid_L
36a	-22.5	-68.5	41.5	3.496696		Parietal_Sup_L

Table S.2.1.1 Significant Clusters: Subjective Risk

Objective Risk

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	61.5	3.5	-0.5	-4.803749	168.0	Temporal_Pole_Sup_R
1a	53.5	-0.5	-2.5	-4.041432		Temporal_Sup_R
2	23.5	-16.5	-16.5	-3.810005	136.0	Hippocampus_R

S.2.1.2 Significant Clusters: Objective Risk

Subjective Balloon Total

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	-28.5	-94.5	13.5	5.107556	536	Occipital_Mid_L
2	-14.5	-92.5	-8.5	4.775488	1680	Calcarine_L
2a	-14.5	-94.5	-16.5	4.566446		Lingual_L
2b	-26.5	-86.5	-20.5	4.528158		Cerebelum_Crus1_L
2c	-20.5	-92.5	-18.5	4.341572		undefined
3	45.5	-70.5	3.5	4.775125	784	Temporal_Mid_R
3a	49.5	-66.5	-2.5	3.961978		Temporal_Inf_R
3b	55.5	-68.5	5.5	3.786144		undefined
4	17.5	-84.5	-12.5	4.670000	3256	Lingual_R
4a	31.5	-84.5	-16.5	4.565753		Lingual_R
4b	25.5	-60.5	-16.5	4.548874		Cerebelum_6_R
4c	31.5	-48.5	-6.5	4.532816		Fusiform_R
5	25.5	-86.5	23.5	4.609972	360	Occipital_Sup_R
5a	21.5	-92.5	27.5	4.121848		Occipital_Sup_R
6	21.5	-98.5	3.5	4.512460	432	Occipital_Sup_R
6a	27.5	-94.5	13.5	3.903283		Occipital_Mid_R
7	39.5	-80.5	5.5	4.432329	1160	Occipital_Mid_R
7a	35.5	-92.5	9.5	3.926808		undefined
7b	39.5	-74.5	13.5	3.885410		Occipital_Mid_R
7c	39.5	-88.5	-2.5	3.880806		Occipital_Inf_R
8	29.5	-44.5	-12.5	4.428687	184	Fusiform_R
9	-0.5	-84.5	-4.5	4.410237	184	Calcarine_L
10	-24.5	-100.5	3.5	4.350987	168	Occipital_Mid_L
11	-44.5	-80.5	-0.5	4.312437	168	Occipital_Mid_L
12	-6.5	-88.5	-6.5	4.253062	248	Calcarine_L
13	-38.5	-70.5	-16.5	4.236236	200	Fusiform_L
14	-10.5	-104.5	-4.5	4.115636	192	Calcarine_L
15	3.5	29.5	35.5	4.109365	144	Cingulum_Mid_R
16	-16.5	-92.5	17.5	4.045882	176	Occipital_Sup_L
17	11.5	-76.5	5.5	3.987839	96	Calcarine_R
18	11.5	43.5	19.5	3.759769	112	Cingulum_Ant_R

Table S.2.1.3 Significant Clusters: Subjective Balloon Total

Objective Balloon Total

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	-24.5	-0.5	69.5	4.795791	128	Frontal_Sup_L
2	-38.5	37.5	31.5	4.183302	312	Frontal_Mid_L
2a	-38.5	45.5	33.5	3.989016		Frontal_Mid_L
3	-42.5	-74.5	33.5	4.098896	112	Occipital_Mid_L
4	-50.5	-78.5	27.5	3.866819	144	undefined

Table S2.1.4 Significant Clusters: Objective Balloon Total

Subjective Pump Reward

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	-28.5	-94.5	13.5	5.107556	536	Occipital_Mid_L
2	-14.5	-92.5	-8.5	4.775488	1680	Calcarine_L
2a	-14.5	-94.5	-16.5	4.566446		Lingual_L
2b	-26.5	-86.5	-20.5	4.528158		Cerebelum_Crus1_L
2c	-20.5	-92.5	-18.5	4.341572		undefined
3	45.5	-70.5	3.5	4.775125	784	Temporal_Mid_R
3a	49.5	-66.5	-2.5	3.961978		Temporal_Inf_R
3b	55.5	-68.5	5.5	3.786144		undefined
4	17.5	-84.5	-12.5	4.670000	3256	Lingual_R
4a	31.5	-84.5	-16.5	4.565753		Lingual_R
4b	25.5	-60.5	-16.5	4.548874		Cerebelum_6_R
4c	31.5	-48.5	-6.5	4.532816		Fusiform_R
5	25.5	-86.5	23.5	4.609972	360	Occipital_Sup_R
5a	21.5	-92.5	27.5	4.121848		Occipital_Sup_R
6	21.5	-98.5	3.5	4.512460	432	Occipital_Sup_R
6a	27.5	-94.5	13.5	3.903283		Occipital_Mid_R
7	39.5	-80.5	5.5	4.432329	1160	Occipital_Mid_R
7a	35.5	-92.5	9.5	3.926808		undefined
7b	39.5	-74.5	13.5	3.885410		Occipital_Mid_R
7c	39.5	-88.5	-2.5	3.880806		Occipital_Inf_R
8	29.5	-44.5	-12.5	4.428687	184	Fusiform_R
9	-0.5	-84.5	-4.5	4.410237	184	Calcarine_L
10	-24.5	-100.5	3.5	4.350987	168	Occipital_Mid_L
11	-44.5	-80.5	-0.5	4.312437	168	Occipital_Mid_L
12	-6.5	-88.5	-6.5	4.253062	248	Calcarine_L
13	-38.5	-70.5	-16.5	4.236236	200	Fusiform_L
14	-10.5	-104.5	-4.5	4.115636	192	Calcarine_L
15	3.5	29.5	35.5	4.109365	144	Cingulum_Mid_R
16	-16.5	-92.5	17.5	4.045882	176	Occipital_Sup_L
17	11.5	-76.5	5.5	3.987839	96	Calcarine_R
18	11.5	43.5	19.5	3.759769	112	Cingulum_Ant_R

Table S2.1.5 Significant Clusters: Subjective Pump Reward

Objective Pump Reward

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	53.5	17.5	-4.5	4.883434	112	Temporal_Pole_Sup_R
2	33.5	41.5	23.5	4.769185	792	Frontal_Mid_R
2a	35.5	49.5	29.5	4.204031		Frontal_Mid_R
3	-34.5	35.5	31.5	4.658528	624	Frontal_Mid_L
3a	-40.5	41.5	27.5	3.830004		Frontal_Mid_L
4	-0.5	21.5	41.5	4.555008	280	Frontal_Sup_Medial_L
5	3.5	11.5	67.5	4.541164	536	Supp_Motor_Area_R
5a	5.5	5.5	55.5	4.155540		Supp_Motor_Area_R
5b	1.5	9.5	49.5	4.112620		Supp_Motor_Area_R
5c	7.5	3.5	59.5	3.983601		Supp_Motor_Area_R
6	-8.5	3.5	51.5	4.485950	184	Supp_Motor_Area_L
7	-38.5	-56.5	-58.5	4.343652	168	Cerebelum_8_L
8	5.5	9.5	39.5	4.177705	272	Cingulum_Mid_R
9	33.5	25.5	7.5	4.093515	136	Frontal_Inf_Tri_R
10	-44.5	-58.5	-30.5	4.057850	96	Cerebelum_Crus1_L
11	21.5	-0.5	69.5	3.921308	168	Frontal_Sup_R
12	33.5	17.5	9.5	3.878735	232	Insula_R

Table S2.1.6 Significant Clusters: Objective Pump Reward

Expected Value

Cluster ID	X	Y	Z	Peak Stat	Cluster Size (mm3)	Brain Region
1	-4.5	-56.5	23.5	5.140080	2904	Precuneus_L
1a	-8.5	-64.5	23.5	4.950534		Cuneus_L
1b	-0.5	-62.5	21.5	4.721457		Precuneus_L
1c	-4.5	-50.5	13.5	4.517141		Precuneus_L
2	-18.5	-50.5	5.5	4.922153	112	Calcarine_L
3	-60.5	-54.5	23.5	4.744176	3384	SupraMarginal_L
3a	-52.5	-74.5	33.5	4.575413		undefined
3b	-44.5	-74.5	31.5	4.540016		Occipital_Mid_L
3c	-42.5	-64.5	29.5	4.396492		Angular_L
4	-24.5	27.5	47.5	4.682723	184	Frontal_Mid_L
5	-54.5	1.5	-20.5	4.637917	192	Temporal_Mid_L
6	-64.5	-56.5	-0.5	4.418828	880	Temporal_Mid_L
6a	-66.5	-46.5	3.5	4.335637		Temporal_Mid_L
6b	-68.5	-46.5	-12.5	3.933416		Temporal_Mid_L
6c	-72.5	-36.5	-0.5	3.837472		undefined
7	-2.5	25.5	-10.5	4.345196	168	Frontal_Med_Orb_R
8	-22.5	27.5	37.5	4.322398	152	Frontal_Mid_L
9	13.5	-88.5	-38.5	4.143846	104	Cerebelum_Crus2_R
10	-66.5	-8.5	-10.5	4.106753	120	Temporal_Mid_L
11	1.5	55.5	-10.5	4.072276	112	Frontal_Med_Orb_R
12	49.5	-58.5	25.5	3.961827	160	Angular_R
13	-8.5	-44.5	1.5	3.934750	112	Lingual_L
14	5.5	-46.5	31.5	3.779486	192	Cingulum_Mid_R
14a	3.5	-44.5	23.5	3.555864		Cingulum_Post_R

Table S2.1.7 Significant Clusters: Expected Value

S2.2
fMRI Prep Boilerplate:
fMRI Data Collection

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.4 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.6.1 (K. Gorgolewski et al. (2011); K. J. Gorgolewski et al. (2018); RRID:SCR_002502).

Preprocessing of B0 inhomogeneity mappings

A total of 11 fieldmaps were found available within the input BIDS structure for this particular subject. A *B0*-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with topup (Andersson, Skare, and Ashburner (2003); FSL 6.0.5.1:57b01774).

Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset.The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.3.3 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152*

Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 11 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al. 2002). The estimated *fieldmap* was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.346s (0.5 of slice acquisition range 0s-0.693s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal

components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e.

head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Functional data preprocessing

For each of the 11 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al. 2002). The estimated *fieldmap* was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.348s (0.5 of slice acquisition range 0s-0.695s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the

definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a *preprocessed*

BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

S3: CCA Analysis: Combined Model Permutation Testing

CCA analysis exploring best-fitting parameters in the independently fit BART

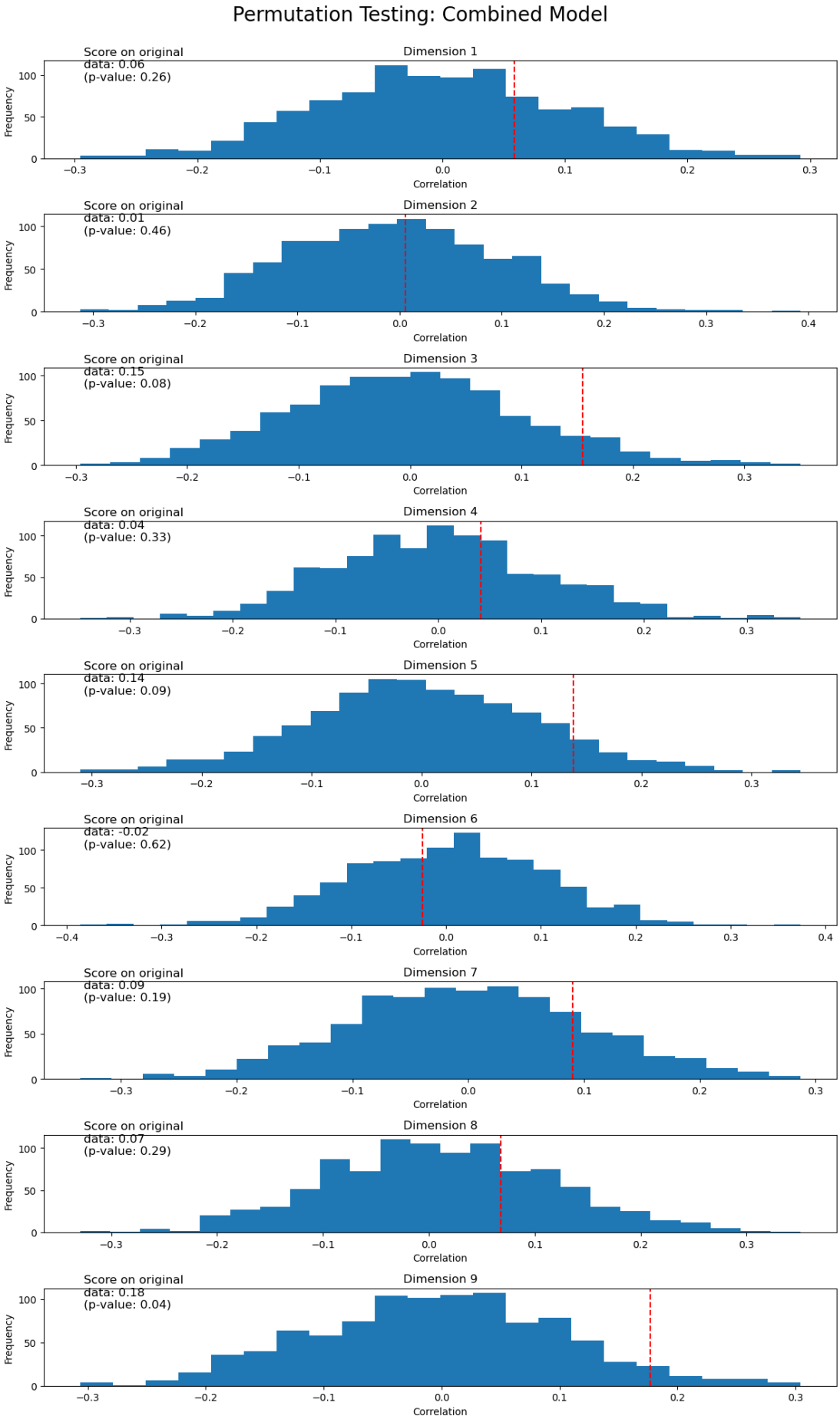

Permutation Testing: Combined Model
Score on original data: 0.06 (p-value: 0.26)
Dimension 1
Score on original data: 0.01 (p-value: 0.46)
Dimension 2
Score on original data: 0.15 (p-value: 0.08)
Dimension 3
Score on original data: 0.04 (p-value: 0.33)
Dimension 4
Score on original data: 0.14 (p-value: 0.09)
Dimension 5
Score on original data: -0.02 (p-value: 0.62)
Dimension 6
Score on original data: 0.09 (p-value: 0.19)
Dimension 7
Score on original data: 0.07 (p-value: 0.29)
Dimension 8
Score on original data: 0.18 (p-value: 0.04)
Dimension 9
Frequency
Correlation

Figure S3.3: CCA Permutation Testing. Permutation testing in CCA occurs by shuffling the two data sets and refitting numerous times to see if a stronger linear combination of component features can be found. Permutation testing was performed with 1000 permutations. The red dotted line represents the observed CCA score. The blue histograms show the scores of 1000 permutations. Dimension 9 is the only component that demonstrates the linear combination between the observed data sets is significant.