Evaluating and Extending Theories of Choice Under Risk*

Drew Fudenberg and Indira Puri

Department of Economics, MIT

First posted: February 2, 2021
This version: May 25, 2023

Abstract

Prospect theory (Kahneman and Tversky (1979)) and cumulative prospect theory (Tversky and Kahneman (1992)) model agents who overweight small probabilities, while simplicity theory (Puri (2022) models agents who prefer lotteries with fewer outcomes. We evaluate the predictive performance of these theories, and of hybrid models that combine them, on lotteries with varying numbers of outcomes. A heterogeneous-agent model that combines simplicity theory with cumulative prospect theory has the highest outsample predictive accuracy, and comes close to machine learning performance. We also study the relationship between probability weighting and simplicity, and analyze observable determinants of membership in each behavioral group.

Keywords: probability weighting, simplicity, cumulative prospect theory, complexity, risk, heterogeneous agents, mixture models
JEL: D81, D90, C49

*We thank Monica Agrawal, Yoram Halevy, Gustav Karreskog, Annie Liang, and Charles Sprenger for helpful comments, and NSF Grant 1951056, the NSF Graduate Fellowship, and the Paul and Daisy Soros Fellowship for New Americans for financial support.

1 Introduction

There is abundant evidence that people's risk-taking behavior in laboratory experiments depends nonlinearly on the stated probabilities of the various outcomes, and in particular that people overweight small probabilities and underweight large ones. For that reason, the most commonly used alternatives to expected utility, namely prospect theory (PT, Kahneman and Tversky (1979)) and cumulative prospect theory (CPT, Tversky and Kahneman (1992)) both allow for for non-linear probability weighting. More recently, simplicity theory (Puri (2020)) posits that people assign a utility premium to lotteries with fewer possible outcomes. This is consistent with, but stronger than, a preference for deterministic payments over stochastic ones, and can explain some behavior that PT and CPT cannot. Conversely, because simplicity theory reduces to expected utility theory when comparing lotteries with the same support size, it cannot explain some behavior that is consistent with non-linear probability weights. This leads us to compare the predictive performance of these theories with hybrid models that we term PT-Simplicity and CPT-Simplicity.[1]

We evaluate the standard functional forms for utility functions and probability weighting, and a functional form for simplicity theory that we introduce here. For each model, we allow for heterogeneous agents with different model parameters. We find that a heterogeneous-agent model combining simplicity theory with CPT (the CPT-Simplicity model) does best, and comes close to matching machine-learning performance, which suggests that it captures most of the empirical regularities in the data. The other models all perform less well.

The CPT-Simplicity model uncovers three behavioral types in the population. Group 1, which comprises about 42% of our data, is complexity averse and distorts probabilities mildly. Group 2, comprising 36% of our data, is less strongly complexity averse, and also distorts probabilities mildly. Group 3, comprising 21%, weighs probabilities heavily and is complexity averse.

We also test for whether participants prefer simplicity over and above any preference for certainty. We do so by calculating the residuals from the three-group CPT model, and regressing these on the number of outcomes. We find that the coefficient on number of outcomes is statistically significant for two of three groups, which in-

[1]By predictive performance we mean that we train our models on one dataset, and test them on another, a standard technique to avoid overfitting (Tulchinsky, 2020).

dicates these groups have a preference for simplicity over and above a preference for certainty.

We then test whether it is possible to predict a person's behavioral type from their observable characteristics. Education, employment status, age, and race do not predict group membership in a statistically significant way. Financial literacy is predictive: a lower financial literacy score increases the probability of belonging to Group 3, which weights probabilities strongly, and decreases the probability of belonging to Groups 1 and 2, which weight probabilities less. However, all three groups have substantial shares of both financially literate and financially illiterate individuals. Gender and income are also predictive.

As tests of external validity, we apply the parameters and groups found by the heterogeneous CPT-Simplicity model to three other datasets. First, we show that the model accurately predicts the amount of Allais-type behavior in prior studies. Second, we show that our estimated model approximately replicates the event-splitting findings of Bernheim and Sprenger (2020), which compares two-outcome lotteries with three-outcome lotteries formed by adding a mean-preserving spread to what had been the more likely outcome.

Our estimation technique for heterogeneous economic models improves on earlier work in several ways, which allows it to more consistently find parameters used to generate synthetic data. For example, we use a validation set to choose the number of groups and other hyperparameters.

We focus on pure economic models, rather than on hybrids of machine learning and economic theory, because pure economic models are easier to interpret and also easier to port to other domains to evaluate comparative statics and policy implications.

2 Related Work

Testing the Independence Axiom The literature that directly tests (and typically rejects) the independence axiom of expected utility theory includes studies of the Allais paradox (including Conlisk (1989), Fan (2002), Huck and Müller (2012), Mongin (2019)) and also tests of the independence axiom with various combinations of probabilities and support sizes (Harless and Camerer (1994)). Fehr-Duda and Epper (2012) points out that probability weighting has been repeatedly observed in the lab, though evidence outside the lab is mixed; for example, Hajimoladarvish (2017) finds underweighting of 0.01 and 0.1, and Etchart-Vincent (2009) finds that probability weighting changes with the magnitude of the outcomes.

Representative-agent parametric estimates Most of the work that estimates CPT uses a representative-agent model and restricts attention to lotteries with two outcomes (Tversky and Kahneman (1992), Tversky and Fox (1995), Wu and Gonzalez (1996), Abdellaoui (2000), Bleichrodt and Pinto (2000), Booij, van Praag and van de Kuilen (2010), Tanaka, Camerer and Nguyen (2010)). Similarly, in a survey of work on probability weighting, Fehr-Duda and Epper (2012) only consider two-outcome lotteries, as does the Bruhin, Epper and Fehr-Duda (2010), which estimates a mixture model.

While many papers find evidence of PT or CPT behavior, parametric values seem to vary widely from study to study (Neilson and Stowe, 2002). Fehr-Duda and Epper (2012) also surveys work reporting that parametric values vary from population to population, for example, on student versus representative versus professional trading samples.

Although PT and CPT incorporate probability weighting, and help explain many experimental results, their validity is challenged in several recent studies (Etchart-Vincent (2009), Andreoni and Sprenger (2011), Bernheim and Sprenger (2020)). Bernheim and Sprenger (2020) tests and rejects the rank-dependence feature of CPT, and also find evidence of a form of stochastically dominated choices when events are "split" that violates both PT and CPT.

Mixture Models Conte, Hey and Moffatt (2011) estimates a mixture model on lotteries with 2 and 3 outcomes. It assumes there are two groups in the population, one of which has expected utility preferences and the other CPT with a one-parameter family of probability weighting functions. Their main conclusion is that there is indeed heterogeneity in the population. Bruhin, Epper and Fehr-Duda (2010) estimates a mixture model for PT/CPT on binary lotteries, where these two theories are equivalent, and uses training data to determine the optimal number of groups. They find that there are three distinct groups, with two of them (about 80% of the population) exhibiting significant probability weighting. Fudenberg et al. (2020) also estimates CPT with three types of agents on binary lotteries, and finds that it does a good job of predicting the average certainty equivalents.[2]

Certainty preference and complexity aversion Some part of the observed violations of PT and CPT seems to reflect a preference for certainty; as summarized by Harless and Camerer (1994), "in the triangular interior [e.g for lotteries with three outcomes], however, EU manages a miraculous recovery." Moffatt, Sitzia and Zizzo (2015) and Sonsino, Benzion and Mador (2002) find that, holding fixed the expected value, people prefer lotteries with fewer outcomes even when those lotteries have higher variance. Dillenberger (2010) captures a preference for certainty with the axiom of negative certainty independence, and Cerreia-Vioglio, Dillenberger and Ortoleva (2015) uses this axiom to provide a representation theorem for "cautious expected utility."

Puri (2020) introduces and axiomatizes a "simplicity theory," which posits that agents incur a weakly increasing disutility from increases in a lottery's support. The empirical evidence so far suggests that a preference for simplicity may explain regularities in behavior that probability weighting cannot: Bernheim and Sprenger (2020) finds an effect of event splitting that is consistent with simplicity preference but not with PT or CPT, and Goodman and Puri (2022) and Puri (2022) find evidence of behavior that cannot be rationalized by PT, or CPT but is predicted by simplicity theory. Finally Fudenberg and Puri (2022) used estimated parameters from an older version of this paper (Fudenberg and Puri, 2021) to speculatively predict take up of prize-linked savings accounts in South Africa.[3]

Machine Learning and Risk Preferences Erev et al. (2010), Erev et al. (2017), Plonsky et al. (2016), and Plonsky et al. (2017) run prediction contests in which participants were invited to submit machine learning algorithms to predict people's choices over lotteries. Plonsky et al. (2016) and Plonsky et al. (2017) find that a combination (representative agent) ML-theory algorithm performs better than other contest entries. Ke et al. (2020) similarly introduces a combination ML-theory algo-

[2]Andersen, Harrison and Rutström (2006), Harrison and Rutström (2009), and Harrison, Humphrey and Verschoor (2010) use finite mixture models, but rather than a mixture of individuals, these papers use a mixture model on choices.

[3]The parameter estimates have changed, and while the qualitative result that CPT predicts too much takeup is similar, CPT-simplicity now substantially underpredicts the actual take up of 63%: CPT now predicts that all groups should take up at 90-100%; and CPT-Simplicity now predicts that Group 2 has 46% takeup while the other groups have 14 and 7% respectively.

rithm. Our focus is instead on whether any economic theory can predict choices well; we do not introduce or evaluate combination ML-theory algorithms. Also, the models in these papers are implicitly representative agent: the same algorithm is applied to everyone. In contrast, we allow for heterogeneous agent models. Finally, when these papers use non-binary lotteries, they are lotteries with either a binomial distribution, or a left- or right-skewed binomial distribution. Our multi-outcome lotteries are generated uniformly at random, which gives these lotteries more variety both in the outcomes and in the probabilities assigned to those outcomes.

Peysakhovich and Naecker (2017) uses ridge regression to predict choices on binary lotteries, and compare its performance to CPT. We use a larger range of machine learning models, test a larger variety of economic models, use a dataset with more variance in the number of outcomes, and apply our findings to external datasets. We contribute to this literature by using unsupervised learning techniques, which we find perform best. We also use validation sets, rather than train or test sets, to select hyperparameters.

Other complexity measures There are other ways that lotteries can be complex beyond support size. For example, a uniform distribution over 20 outcomes may be more appealing that a lottery over 19 outcomes that has higher mean and lower variance if the distribution of the 19-outcome lottery is very irregular. Prior papers have hypothesized that insensitivity to probabilities may be due to cognitive limitations (Viscusi, 1989; Wakker, 2010), and Enke and Graeber (2021) considers a model that, in the risk context, implies probabilities are biased towards a uniform prior. Puri (2022) proposes a relatively abstract way to model complexity that can accommodate linguistic complexity and ease-of-computation effects. While our measure considers only one measure of simplicity, it is striking that this measure performs favorably relative to CPT, arguably the most prominent behavioral model over the past several decides.

Comparing the Performance of Theories Fudenberg et al. (2020) proposes that the "completeness" of a theory should be measured by comparing its percentage improvement over a naive model to that of a table-lookup algorithm. We modify this to a *machine learning completeness score* that uses a machine learning benchmark in place of a table-lookup benchmark.[4]

3 Experimental Design

3.1 Participants

To gather our data we ran a survey on Amazon Mechanical Turk (MT) from May - June 2022. We implemented four steps to help ensure clean data collection. First, we implemented a practice round and three comprehension questions. Second, we paid well above MT average earnings, with our participants earning, on average, \$6.81 for 19 minutes of survey completion time, for an hourly rate of \$20.43, three times over the typical hourly wage on MT (Hara et al., 2017). Third, we restricted participation to workers who had at least 100 prior tasks and had successfully completed at least 95% of them, using the auxiliary service CloudResearch, as suggested by Moss et al. (2021) and Peer et al. (2021). Fourth, we included a free response question at the end of the survey, asking participants whether or not they felt they understood the task.

We collected 201 responses.[5] Participants who completed the survey indicated that they understood it, with most responses to the question of whether they understood the task being some variant of 'I understood'; 'Yes'; or 'I understood from the beginning.'

We had collected an earlier dataset, and reported those results in an earlier version of the paper (Fudenberg and Puri, 2021). The 2022 dataset is superior in several ways. First, far more of the participants passed the comprehension test. For example, our free response question ('Do you felt you understood the task at the start of the experiment? If not, when did you start understanding it?') allows participants to indicate if they had any trouble understanding the certainty equivalent task. Almost all participants report understanding the task. Second, we introduce a practice round on certainty equivalents to further help participants understand the task. Third, our attrition rate in the 2022 sample is much lower than that in the earlier sample. Fourth, we present lotteries so that a single click autofills rows of the multiple price list. Fifth,

[4] This is in the spirit of Peysakhovich and Naecker (2017) and Bodoh-Creed, Boehnke and Hickman (2018), which compare the performance of machine learning algorithms with that of models of risk preference.

[5] A further 55 people were kicked out after they failed a comprehension check.

the earlier experiment had payoffs and probabilities on a finer grid. We now generate all lotteries so that their probabilities are multiples of 0.05, and their payoffs are multiples of 0.25, which are 'clean' numbers participants may find easier to grapple with.

The two samples statistically come from different distributions (a Kolmogorov-Smirnoff test for differences in the distribution of relative risk premia yields a statistically significant p-value of < 0.01), so we do not combine the datasets, but only use the newer one.

Participants' task was to provide their certainty equivalents for 30 different lotteries: 10 each of 2, 4, and 6 outcome lotteries, with lottery presentation in random order. To elicit certainty equivalents, we used the standard multiple price list procedure with single switching, described in further detail below.

In addition to having participants provide certainly equivalents for lotteries, we ask participants about their income, age, sex, education, and employment. We also ask financial literacy questions from Lusardi and Mitchelli (2007). There are three financial literacy questions; the first two are simple percentage calculations, and the third is a compound interest calculation (the exact questions are in Appendix A.2). We impose a time limit of 45 seconds to answer the first question, one minute to answer the second question, and five minutes to answer the third question. Following Lusardi and Mitchelli (2007), the participant receives a score of 2 if they answer the first or second question correctly and the third question correctly; they receive a score of 1 if they answer the first or second question, but not the third question, correctly; and they receive a score of 0 if they answer no questions correctly.

The summary statistics for each demographic and financial literacy question are shown in Table 4 in Appendix A.3. In our sample, 69% of participants receive the highest financial literacy score, 57% have at least a four-year college degree, and 72% are under the age of 40.

3.2 Random Lottery Generation

There are 30 lotteries, 10 for each of two, four, and six outcomes. We do not want there to be structural differences between lotteries with different numbers of outcomes, so we generate random lotteries by matching on moments, e.g. we generate a set of four outcome lotteries, and match the two and six outcome lotteries to those of the four outcome lotteries. Probabilities range from 0 to 1 and occur in increments of 0.05. Payoffs range from \$1 to \$10 and occur in increments of 0.25. The full random lottery generation algorithm may be found in Appendix A.1, which also shows that the mean, variance, and skewness of payoffs are not significantly correlated with the number of outcomes. Each subject provided certainty equivalents for all lotteries.

3.3 Obtaining Certainty Equivalents

We use the multiple choice list method, as in Bruhin, Epper and Fehr-Duda (2010), Bernheim and Sprenger (2020), and Chapman et al. (2017), among others.[6] To ensure that participants do not have to spend excessive time clicking, we impose single switching, as in Andersen et al. (2006) and Tanaka, Camerer and Nguyen (2010).

Following Bruhin, Epper and Fehr-Duda (2010), for each lottery, each question has 20 evenly-spaced choices. The choices range from \$0.50 cents to \$10. This allows participants to violate dominance if they choose, by picking a certainty equivalent below the lowest outcome in the lottery. Participants were asked to provide certainty equivalents for all 30 lotteries. Figure 1 shows a sample question.

4 Preference Models

We consider monetary lotteries p with outcomes $x_1 > x_2 > ... \leq x_{|support(p)|} > 0$, where the support size $|support(p)|$ ranges from 2 to 6. Our basic preference models are expected utility theory, PT, CPT, and simplicity theory.

- Expected Utility Theory:

$$u(p) = \sum u(x)p(x)$$

- PT:

$$u(p) = \sum_i u(x_i)\pi(p(x_i))$$

[6]This is a standard way to determine risk preferences, but there is some evidence that other procedures yield different behavior, see e.g. Andreoni and Sprenger (2011), Freeman, Halevy and Kneeland (2019).

The LOTTERY for this question is:

Probability	Outcome
15%	$9.75
25%	$8.25
40%	$4.50
20%	$1.25

In each row, please decide whether you prefer Option Safe or Option Lottery.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00		
Option Safe = $9.50		
Option Safe = $9.00		
Option Safe = $8.50		
Option Safe = $8.00		
Option Safe = $7.50		
Option Safe = $7.00		
Option Safe = $6.50		
Option Safe = $6.00		
Option Safe = $5.50		
Option Safe = $5.00		
Option Safe = $4.50		
Option Safe = $4.00		
Option Safe = $3.50		
Option Safe = $3.00		
Option Safe = $2.50		
Option Safe = $2.00		
Option Safe = $1.50		
Option Safe = $1.00		
Option Safe = $0.50		

FIGURE 1: Sample Question

- CPT:

$$u(p) = \sum_i u(x_i) \left[\pi \left(\sum_{k=1}^{i} p_k \right) - \pi \left(\sum_{k=1}^{i-1} p_k \right) \right],$$

  where outcomes are in decreasing order, and the sum from 1 to 0 is defined to be 0.[7]

- Simplicity Theory:

$$u(p) = \sum_i u(x_i)p(x_i) - C(|\text{support}(p)|)$$

  We also consider hybrid models that combine simplicity theory with either PT or CPT.

- PT-Simplicity:

$$u(p) = \sum_i u(x_i)\pi(p(x_i)) - C(|\text{support}(p)|)$$

- CPT-Simplicity:

$$u(p) = \sum_i u(x_i) \left[\pi \left(\sum_{k=1}^{i} p_k \right) - \pi \left(\sum_{k=1}^{i-1} p_k \right) \right] - C(|\text{support}(p)|)$$

These hybrid models are proposed (but not axiomatized) in Puri (2020). We use the standard CRRA specification $u(x) = x^\alpha$, $\alpha > 0$,and probability weighting function $\pi(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}$, $\gamma > 0$ from Kahneman and Tversky (1979) and subsequent work.[8]

Simplicity theory is relatively new, so there is no standard functional form to use. We specify a three-parameter family of sigmoid functions, with $C(1) = 0$ as the theory requires: $C(x) = \frac{\iota}{1+e^{\kappa(x-\rho)}} - \frac{\iota}{1+e^{\kappa(1-\rho)}}$. In this specification, ι represents the height of the function, and ρ the midpoint of the rise, before normalization. The parameter κ represents the slope, with higher κ corresponding to a steeper slope. Because we normalize C so that $C(1) = 0$, larger values of κ and ρ also increase the height of the

[7] We will only consider the gains domain, and use a fixed reference point of 0, so this is equivalent to Rank-Dependent Expected Utility (Quiggin, 1991).

[8] Bruhin, Epper and Fehr-Duda (2010) uses a more flexible specification with an additional parameter; Fudenberg, Gao and Liang (2022) argues that this extra parameter adds more flexibility to the specification than is justified by the amount it improves predictions.

function, though to a lesser extent than increases in ι. [9] In fitting the function, we restrict each parameter to be weakly positive.[10]

When combining the models, we assume that the simplicity cost enters additively with the PT or CPT evaluation. There is no particular a priori reason that this should be the case, but as we will see this specification is very successful at predicting the certainty equivalents.

5 Econometric and Machine Learning Models

Instead of estimating the preferences of a representative agent with each preference model, we use mixture models, as in Bruhin, Epper and Fehr-Duda (2010), and allow the model's preference parameters to vary across individuals, with each individual classified into one of k groups. This allows some groups to exhibit more probability weighting than others, and also to have different risk preferences and responses to complexity. [11] Our use of outsample evaluation for these models helps reduce the risk of overfitting; models with more parameters can have higher cross-validated prediction errors.

5.1 Evaluation

Each model maps parameter vectors to predicted certainty equivalents for each group and lottery. We evaluate models based on their mean-squared error (MSE), which is the average of the mean-squared error over all observations. We use cross validation to split our data into train, validation, and test sections.[12] The validation sets are used to set hyperparameters such as the number of groups, while the test sets are used to evaluate model performance on unseen data. Because we evaluate on outsample error, our results should generalize more broadly than if we evaluated on insample error, and models with more parameters do not necessarily have an

[9] Because our data only has three different values for $|\text{support}(p)$ we could have instead estimated a separate cost for each support size, but the sigmoid function lets us extrapolate to lotteries with different support sizes.

[10] We have also tried specifications with ι allowed to take on any value, which allows for complexity-loving behavior. Even allowing for that flexibility, we found $\iota > 0$ for all groups, indicating that all groups are complexity averse.

[11] We do not, however, allow some groups to have PT preferences and others to have CPT.

[12] Validation sets are used in machine learning to tune hyperparameters. A hyperparameter is a value that contributes to the description of the model, such as a regularization weight in Lasso.

advantage.

Unless otherwise stated, for all models below, we use the following cross validation procedure. Our data consists of 30 lotteries total, which we randomly divide evenly into s subsets, numbered $A_1, A_2, .., A_s$. We use A_1 as test data, and B_2 as validation data. The models are trained on all other groups. The model hyperparameters are chosen to be those under which the validation MSE is lowest given the parameters estimated on the train data. Once the model parameters and hyperparameters are determined, the model is evaluated on the test data. We use $s = 5$, so that our data is split into 60% train, 20% validation, and 20% test (e.g, 20 train lotteries, 6 validation, and 6 test). To ensure that each fold has a balance along complexity levels, we split so that each fold has two 2 outcome lotteries, two 4 outcome lotteries, and two 6 outcome lotteries.

5.1.1 Econometric Specification

We wish to estimate parameters θ for model m. Denote by $\hat{ce}_l(\theta)$ the predicted certainty equivalent for lottery l when using model m with parameters θ. The observed certainty equivalent for individual i on lottery l is $ce_{i,l} = \hat{ce}_l(\theta) + \epsilon_{i,l}$, where $\epsilon_{i,l} \sim N(0, \sigma_i)$, $\sigma_i = \xi_i$, and each individual i has their own variance ξ_i in lottery selection.

Let $c = 1, 2, ..., C$ denote the different groups. Let θ_c be the parameters associated with group c. The probability density of individual i's choices under parameters θ_c is:

$$f(ce_i; \theta_c, \{\sigma_i\}_{l \in L}) = \prod_{l=1}^{L} \frac{1}{\sigma_{i,l}} \phi\left(\frac{ce_{i,l} - \hat{ce}_l}{\sigma_i}\right),$$

where ϕ denotes the density of the standard normal distribution.

Let π_c denote the probability of group membership of type c. The log-likelihood of the finite mixture model is:

$$\sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f(ce_i; \theta_c, \sigma_i),$$

where the first sum is over individuals and the second sum is over groups. The parameters to be estimated are $(\theta_1, ..\theta_C, \pi_1, .., \pi_C, \xi_1, .., \xi_N)$.

5.1.2 Estimation

We estimate the model using an expectation-maximization algorithm (Dempster, Laird and Rubin, 1977). This algorithm proceeds iteratively, first calculating the log-likelihood of the model, and then estimating the parameters of the model using maximum likelihood. We repeat these two steps until the process converges.

We estimate parameters for $k = 1, 2$, and 3 groups. Because the number of groups is a hyperparameter, we pick the best number of groups based on validation set performance. Our estimation procedure builds on that of Bruhin, Epper and Fehr-Duda (2010), with some changes: we evaluate based on outsample performance, and use validation performance to choose the number of groups for any given model, using the two-step procedure detailed in Section 5.1.2 rather than averaging over initializations. In combining cross validation with bootstrap, we first split the data into folds for cross validation. Then, for 1000 different random seeds, we bootstrap and select which folds correspond to test and validation sets. Also, because we have two levels of noise (the cross-validation procedure and the bootstrap) as opposed to one (just the bootstrap), we add several fixes for numerical stability: We impose a lower bound on each π_k of (number of groups)/100, and impose a lower bound on σ_i of 1/100 and an upper bound of 100. These fixes for numerical stability allow the mixture model to more consistently find parameters used to generate synthetic data. We also use a canonical standard error correction for cross-validation.[13]

5.2 Machine Learning Algorithms

We predicted certainty equivalents using three different machine learning algorithms: neural networks, k-means, and gradient boosting trees. Here we describe how we implemented these algorithms.

5.2.1 Neural Networks

Our output variable for the neural network is the observed certainty equivalent for each lottery and individual. To obtain stronger performance, we allow the neural network to use the following sets of input variables: {the set of outcomes, the set of

[13]Hansen (2022) writes that an approximation for standard error takes the standard deviation over cross-validation estimates and divides by the square root of the number of folds, based on the idea that the cross-validation estimates are approximately uncorrelated across folds.

probabilities, and an indicator for each individual}; {expected value, variance, number of outcomes, and an indicator for each individual}; {expected value, variance, and an indicator for each individual}; and each of these without the individual-specific indicators. Similarly, we allow for any of the following architectures: 2-layer, 2-hidden unit; 2-layer, 3-hidden unit; and 3-layer, 2 hidden units per layer, choosing the best variable set by validation set performance. Having chosen the best input variable set and architecture using validation performance, we report test performance in the results section (Section 6). We use a cross-validation and bootstrap procedure similar to that described in Section 5.1 and evaluate each input variable set - architecture pair using 10 draws across 30 different initializations.

5.2.2 K-Means

K-Means is an unsupervised algorithm that clusters similar groups together. To apply it, we use a cross-validation and bootstrap procedure similar to that described in Section 5.1. Lotteries are split into train, validation, and test sets. We group individuals into k groups using the train data. On the validation data, we predict individual i's certainty equivalent for lottery g to be the average of the certainty equivalents for lottery g picked by everyone else in individual i's group (excluding individual i).[14] We split our data into 60% train, 20% validation, and 20% test. We attempt up to three groups, and pick the best group number and initialization based on validation set performance. For each group, we evaluate using 10 draws across 250 different initializations.

5.2.3 Gradient Boosting Trees

A gradient boosting tree uses multiple splits of the data to predict certainty equivalents. We split our data into 60% train, 20% validation, and 20% test in each bootstrap iteration. The input variable set is {expected value, variance, number of outcomes, and an indicator for each individual}, and the hyperparameter we focus on is the maximum depth of the tree. We use the default maximum depth of up to 3, choosing the best depth and initialization by performance on the validation set. For each maximum depth, we evaluate using 10 draws across 250 different initializations.

[14] To ensure the validation and test predictions are well defined, we impose the constraint that the minimum group size is two. For this, we use the standard Python implementation, which is based on Bradley, Bennett and Demiriz (2000).

6 Analysis

6.1 Relative Scores

Table 1 compares test performance of the economic theories to that of the best-performing machine learning algorithm, which turned out to be k-means; with the exception of EU, every model performs best with three groups.[15] We compute a score for each model, similar to Fudenberg et al. (2020). That paper suggests calculating a completeness score by using the performance of a naive algorithm and a benchmark hold-one-out evaluation method. Here, we use expected value as our naive algorithm, but as a benchmark we use machine learning, so we report the *machine learning completeness score*:

$$\text{Score (model)} = \frac{MSE_{\text{naive}} - MSE_{\text{model}}}{MSE_{\text{naive}} - MSE_{\text{ML}}}$$

We emphasize that we measure completeness with respect to this dataset and our machine learning algorithms. A larger dataset or one with more features could improve the machine learning benchmark, as could advances in machine learning algorithms. Conversely, we might expect larger prediction errors for both machine learning and for the CPT-Simplicity model on a dataset with more variation in the magnitudes of the prizes.

The combination model CPT-Simplicity achieves a 93% ML completeness score when used with three groups, which turns out to be optimal.[16] The next best models, in order, are: Simplicity, CPT, PT-Simplicity, PT, and EU. Simplicity on its own outperforms both CPT and PT. CPT, while performing better than PT, still misses regularities in behavior. [17]Because the test set consists of only six lotteries, the standard errors are relatively large, and the performance gap between CPT-Simplicity and the next three models is not statistically significant.

[15]See (Table 6 in Online Appendix C). The best performance of each machine learning algorithm, as measured by validation set MSE, is reported in Table 7 in Online Appendix C). The fact that the best performing machine learning algorithm is k-means may reflect the size of our dataset— with millions of observations, we might expect a neural network to perform better (D'Souza, Huang and Yeh, 2020).

[16]On the validation set, three groups gives a 20% improvement in MSE relative to using one group.

[17]Because of our more complex set of lotteries, CPT fits less well here than in the binary lotteries studied by Bruhin, Epper and Fehr-Duda (2010) and Fudenberg et al. (2020), both in an absolute sense and compared to ML performance.

Table 2 provides parameter estimates for the CPT-Simplicity model. Figure 2 plots the probability weighting functions for each group. Probability weighting for the majority groups (Group 1 and Group 2, which together constitute about 80% of the sample) is mild. Probability weighting for the Group 3 is much stronger.

In addition to estimating the parameters, the bootstrap method allows us to directly estimate complexity costs $C(2), C(4), C(6)$ as a function of the simplicity parameters found in each bootstrap iteration. Figure 3 plots the simplicity point estimates for each group, as reported in Table 2. All three groups display complexity aversion. The CRRA parameter for all groups is around between 0.4 - 0.9, consistent with α estimates in prior experimental work using prospect theory (Kahneman and Tversky, 1979; Bruhin, Epper and Fehr-Duda, 2010; Camerer and Ho, 1994). The probability weighting parameters, at 0.81 and 0.88 for the two larger groups, resemble that of Wu and Gonzalez (1996) who find $\gamma = 0.74$ in a representative agent context, and at 0.51 for the minority group resembles Camerer and Ho (1994) who find $\gamma = 0.56$ in a representative agent context. Thus probability weighting matters even when simplicity cost is included. This fits with the fact that it is helpful when considering only binary lotteries, where simplicity is irrelevant.

The minority group displays higher complexity aversion and more probability weighting than the other two groups. The two majority groups display some probability weighting and display complexity aversion. These results together suggest that people who display strong probability weighting may also display strong complexity aversion, and that even those who display mild probability weighting are often complexity averse.

Table 1: Outsample Performance

(a) Cross-Validation and Bootstrap

Model	Test MSE	ML Completeness Score
Expected Value	4.284	0%
	(0.271)	
EU	3.083	62%
	(0.190)	
PT	2.814	76%
	(0.245)	
PT-Simplicity	2.616	86%
	(0.262)	
CPT	2.514	91%
	(0.240)	
Simplicity	2.486	92%
	(0.251)	
CPT-Simplicity	2.477	93%
	(0.257)	
Machine Learning Benchmark	2.337	100%
	(0.113)	

Standard error in parentheses.

(b) Cross-Validation Only

Model	Test MSE	ML Completeness Score
Expected Value	4.284	0%
	(0.271)	
CPT	2.416	89%
	(0.089)	
Simplicity	2.385	90%
	(0.085)	
CPT-Simplicity	2.374	91%
	(0.062)	
Machine Learning Benchmark	2.189	100%
	(0.232)	

Standard error in parentheses.

Table 2: CPT-Simplicity Group Parameters

α	0.784	0.649	0.503
	(0.098)	(0.139)	(0.143)
γ	0.811	0.884	0.510
	(0.071)	(0.080)	(0.117)
ι	0.351	0.334	0.573
	(0.211)	(0.202)	(0.289)
κ	4.526	4.216	5.240
	(2.209)	(1.772)	(1.504)
ρ	1.554	1.714	0.917
	(1.261)	(1.222)	(0.508)
$C(2)$	0.062	0.051	0.107
	(0.044)	(0.039)	(061)
$C(4)$	0.088	0.074	0.118
	(0.055)	(0.045)	(0.062)
$C(6)$	0.096	0.083	0.122
	(0.055)	(0.044)	(0.061)
Number	91.867	70.900	38.233
	(5.461)	(4.855)	(6.000)

Standard error in parentheses.

Figure 2: Probability Weighting Functions By Group

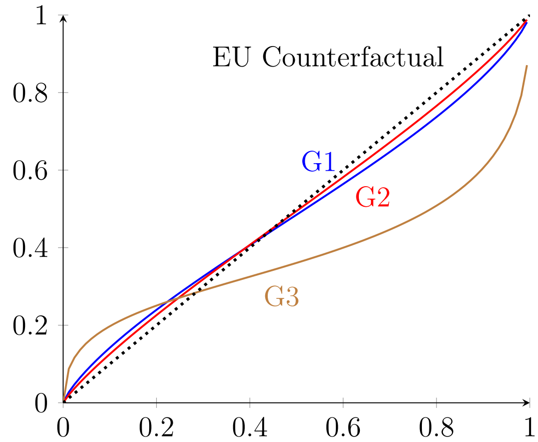
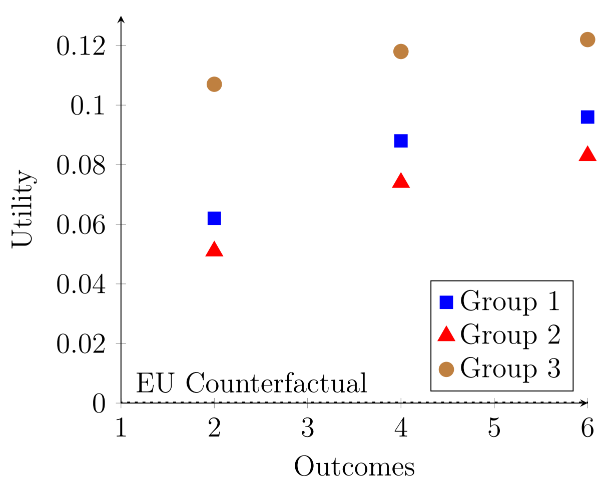


Figure 3: Simplicity Functions By Group



To further supplement our analysis, in Online Appendix D.2, we re-estimate the CPT-Simplicity and CPT models on the 14 lotteries that are not dominated by any other lottery in our dataset. Simplicity allows for dominance violations while CPT does not, so this procedure allows us to check whether Simplicity is adding value over and above allowing for such behavior. We find that CPT-Simplicity continues to outperform CPT outsample, which shows that simplicity's contribution extends beyond explaining dominance violations.

6.2 Constraints on Generality

We start by speaking to the extent to which we would expect these results to hold in other experiments with the same lotteries and same subject population, and then speak to the likely generalizability of our results to other lotteries or individuals.

Because cross-validation is known to have high standard errors (Braga-Neto and Dougherty, 2004; Efron and Tibshirani, 1997), even for data with thousands of observations, computer scientists often use the 'one standard error rule,' which says that one picks the most interpretable model within one standard error of the best performer (Piegorsch, 2015). Following the one standard error rule would imply that *no* model beats machine learning, as none of them are within one standard deviation of ML performance. However, the closest performer is CPT-Simplicity.

It is worth pointing out that typical ML also does not use bootstrap, while we do, introducing even more variance into our analysis. Table 1b recomputes results for the three-group CPT-Simplicity, Simplicity, and CPT models, using cross-validation but not bootstrap, in line with more standard CS procedure. Here again, the "one standard error" rule suggests choosing CPT-Simplicity. In addition, CPT-Simplicity is about one standard deviation better than CPT, and while it is only marginally better than Simplicity alone our priors and the weight of past evidence both suggest that probability weighting is important.

One may also ask whether these results would hold on other datasets, or with other study participants. We developed our algorithm to generate lotteries uniformly at random, keeping mean, variance, and skewness similar across complexity levels. A similar dataset, with lotteries generated in this way, may produce similar results. However, our results may or may not generalize to special distributions such as normals, exponentials, etc. which we do not study here.

Our study population is taken from Amazon Mechanical Turk, which many have argued (Berinsky, Huber and Lenz, 2012) may be at least as representative of the population as studies using a student population.

6.3 Group Composition

The *propensity score* of individual i for group c is the average of the probability that individual i belongs to group c, taken over all iterations.[18] Uniformly random assignment would correspond to a propensity score of 33%. Figure 5 in Appendix B.1 shows that the propensity scores are near zero or above 50% for many individuals, especially in Groups 1 and 3. We now statistically test whether certain characteristics predict group membership. In each iteration of the mixture model, we calculated the probability that an individual belonged to a particular group. For each group, we regress the probability that an individual belongs to group c in iteration b on the matrix of financial literacy and demographic characteristics for that individual. We cluster standard errors at the individual level.[19] Table 5 in Appendix B.2 shows that age, employment status, and education do not predict group membership at the 10% significance level. Few variables are predictive, and even those that are explain little of the variation in group composition. The only variables that have predictive content are financial literacy score, gender, and income. Men are 7% more likely to be in Group 1, and women are 6% more likely to be in Group 3; recall also that the former group shows less risk aversion than the latter. This result is consistent with experimental literature showing that women are more risk averse than men (Eckel and Grossman, 2008), and are more prone to probability weighting (Rieger, Wang and Hens, 2017). A $1,000 increase in income increases the probability of being in Group 1 by 0.07%. Those who are less financially literate are 5% more likely to be in Group 3. Greater financial literacy is associated with being in Groups 1 or 2: a one point increase in the financial literacy score increases the probability of being in Group 1 by 3%, and Group 2 by 2%. Our results seem to indicate that financial literacy matters for probability weighting, but that even financially literate people can have simplicity preferences.

[18] For ease of computation, for this and other sections involving group membership analysis, we rerun the model on all individuals without bootstrapping. This gives a probability that the individual belongs to each group for each iteration, whereas with bootstrap, there may be some iterations where the individual is not drawn. There remains variation from which folds are assigned as validation, train, or test. The resulting parameter estimates for this procedure are numerically similar to those with bootstrap.

[19] Precisely, the specification is $p_{i,c,b} = \beta X_i + \epsilon_{i,c}$, where $p_{i,c,b}$ is the probability that individual i belongs to group c in bootstrap iteration b and x_i are individual characteristics.

7 Non-Parametric Analysis

In this section we examine risk-taking behavior, probability weighting, and complexity aversion without imposing a parametric structure on the simplicity cost. To examine probability weighting, we follow Bruhin, Epper and Fehr-Duda (2010) and plot the median *relative risk premium* $(EV - CE)/EV$.[20] Probability weighting predicts that the relative risk premium is weakly increasing in the probability of the better outcome. We also plot the risk premium against the variance of the lotteries to examine attitudes towards risk.

To examine complexity aversion, we use the *CPT residuals*, which are the differences between the actual certainty equivalents and those predicted by CPT with our estimated parameters. If CPT is the true model this residual should be unrelated to the number of outcomes, whereas complexity aversion predicts the residual would be increasing in the number of outcomes. In all regressions that follow, we include an intercept term, so that any relationship between number of outcomes and CPT residuals is relative to the canonical CPT model. One interpretation of including an intercept is what CPT plus a certainty premium captures, with respect to number of outcomes. Regardless of whether or not we include an intercept term, we obtain similar results (e.g, statistical significance on the number of outcomes coefficient is similar regardless of whether an intercept is included).

7.1 Overall

Figure 6 in Online Appendix C.2 shows that individuals respond strongly to both expected value and risk, with the risk premium increasing with variance. There is evidence of mild probability weighting: the median risk premium is negative for probabilities below 0.2, and mostly positive afterwards. The relative risk premium first increases, then decreases in the probability of the better outcome. There is evidence that complexity matters: a regression of the CPT residual on the number of outcomes, clustering standard errors by individuals, yields a statistically significant coefficient of 0.03 (p-value < 0.03).

A representative agent here would be complexity averse and display some prob-

[20]While we follow the literature in calling the object $EV - CE$ the risk premium against the probability of obtaining the better outcome in two-outcome lotteries (Fehr-Duda and Epper, 2012). Puri (2020) pointed out that in the presence of a simplicity preference, the risk premium is better thought of as a "risk-complexity premium."

ability weighting. We had instead fit a heterogeneous agent model. The next few subsections consider the unconditional data by group, checking whether the data without parametric assumptions maps back to the parametric behavior we found. Breaking the data by group allows more nuanced patterns to emerge.

7.2 By Group

We show that the non-parametric behavior described for each group is qualitatively similar to the parametric behavior predicted for that group. We assign an individual to a group if their propensity score for that group is at least 0.5. This procedure successfully assigns 161 of the 201 individuals (about 80%) to groups.

In additional to the graphical analysis described earlier, we also examine complexity aversion analytically, by regressing the CPT residual against the number of outcomes, clustering standard errors at the individual level. To construct a CPT residual for each group, we use the γ and α estimates for that group as found using a representative agent CPT model for that group.

Figure 7 shows the probability weighting, variance, and expected value plots for Group 1 (all figures referred to in this section are in Online Appendix C.2). The group shows mild evidence of probability weighting: the median relative risk premium increases, then decreases in the probability of the higher outcome for two-outcome lotteries (Figure 7c). The group shows strong evidence of risk aversion, with median risk premium increasing linearly in the variance of the lottery (Figure 7b).

This group shows evidence of complexity aversion: regressing the CPT residual on the number of outcomes, clustering standard errors by individual, yields a coefficient of -0.03, with a p-value of 0.04.

Group 2 likewise displays evidence of complexity aversion, and mild probability weighting. For complexity aversion, regressing the CPT residual on number of outcomes, clustering standard errors at the individual level, yields a coefficient of -0.05, with p-value < 0.01.

In Figure 8 the median relative risk premium first increases, then remains flat or decreases, as the probability of the better outcome increases for two-outcome lotteries (Figure 8c), which is mild evidence of probability weighting (see Fehr-Duda and Epper (2012) and Bruhin, Epper and Fehr-Duda (2010)). This group also responds strongly to expected value, and to variance (compare Figures 8b and 7b).

Group 3 displays evidence for probability weighting (Figure 9c), as their relative risk premia tends to increase in the probability of the better outcome. The group also displays evidence of complexity aversion, and mild probability weighting. For complexity aversion, regressing the CPT residual on number of outcomes, clustering standard errors at the individual level, yields a coefficient of -0.05, with p-value 0.04.

Like the other two groups, this group responds strongly to the expected value (Figure 9a). This group also responds more strongly to variance than the first two groups (Figure 9b).

7.3 Testing for a Simplicity Preference

Section 6.3 looked at the CPT residual, with the groups as found by the CPT-Simplicity model. This helped us understand behavior for each of those groups. To test for a simplicity preference, we now use the groups found by the CPT model[21], and construct CPT residuals for each of these groups. The null hypothesis behind this test is that the true model within each group is CPT, and that the CPT model correctly assigns people to groups.

Table 3 shows the coefficient and p-value after regressing the CPT residual on number of outcomes, clustering errors at the individual level. These regressions have a statistically significant slope for the second and third group. Separate regressions also indicate complexity aversion over and above what CPT with a certainty preference can capture; in CPT with a certainty preference, we would expect the intercept to be positive and the slope flat, which is not what we find.[22] An alternative explanation for complexity aversion is that the probability weighting parameter γ is not fixed, but varies with the cardinality of the lottery's support. To test this alternative theory, we hold α fixed at the value found for each group on two outcomes only, and separately estimate a CPT model for each number of outcomes for each group. Online Appendix D.1 shows that, for Groups B and C (for which the CPT residuals test showed a need for simplicity), γ's do not vary statistically by number of outcomes.

[21]The parameters found by the heterogeneous agent CPT model are reported in Appendix C.3.

[22]Regressions with an intercept term yield a negative and statistically significant slope for the largest group, and statistically insignificant slopes for the other two groups.

Table 3: Regress CPT Residual on Number of Outcomes; Groups Found by Three-Group CPT Model

Group	Coefficient on N. Outcomes	p-value of Coefficient
Group A	-0.01	0.67
Group B	-0.05	0.00
Group C	-0.07	0.01

8 External Applications

In this section we examine the predictions of our estimated CPT-Simplicity model on data we did not collect. To make predictions for each group we draw parameters for that group from a normal distribution whose mean is the group's point estimate for that parameter and whose standard deviation is the group's standard deviation for that parameter as reported in 2.[23] We draw these parameters 1000 times. Some of the applications involve payoffs beyond the range of our experiment. In these cases we linearly decrease the payoffs to fall within the range that we examined.

The Allais paradox For the Allais paradox, the behavior our model predicts is close to those found in incentivized experiments. Conlisk (1989) reports that 6% of people violate EU in the Allais paradox; Huck and Müller (2012) finds 8% of people display Allais behavior; and Harrison (1994) finds 15% of people displaying Allais behavior. Using the the payoffs of Conlisk (1989) (which are similar to those in Huck and Müller (2012)), our model predicts that Group 1 and Group 2 display Allais behavior 0.2% and 12% of the time, respectively while the minority Group 3 displays Allais behavior 62% of the time. Taking a population weighted average, our simulations predict that 18% of the time, the population displays Allais behavior when faced with real, small stakes, in line with proportion of Allais behavior found by studies using these stakes.[24]

[23]This is standard Monte Carlo (Brandimarte, 2014) on group behavior, where the uncertainty arises from uncertainty over the parameter estimates. Here, individual heterogeneity is already accounted for by splitting the population into three groups.

[24]There is a vast literature on Allais with large and hypothetical stakes. We compare our predictions to the real stakes results because they are incentivized. The real stakes used in experiments are also close to those used in our experiment.

Neilson and Stowe (2002) notes that there is a large range of probability weighting estimates for CPT models. In particular, Kahneman and Tversky (1979) finds a CPT parameter of $\gamma = 0.61$; Camerer and Ho (1994) finds $\gamma = 0.56$; and Wu and Gonzalez (1996) finds $\gamma = 0.74$. Our heterogeneous agent model can accommodate the range of previous parameters found.[25]

Event splitting Our model also predicts the event splitting behavior found by Bernheim and Sprenger (2020), where spitting the probability of an event leads to a drop of the certainty equivalent that is not consistent with CPT. Paralleling their "split-low" case, we simulate the utility for each group for the lotteries $\{\$2, 0.6; \$3, 0.4\}$ and $\{\$2 - \epsilon, 0.3; \$2 + \epsilon, 0.3; \$3, 0.4\}$, where $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$.[26] Bernheim and Sprenger (2020) finds that experimental participants provide a lower certainty equivalent from the three-outcome lotteries, over and above what CPT would predict; and further the certainty equivalent is higher for the two-outcome lottery than for the three-outcome lottery with the smallest value of ϵ.

Simulating behavior for each group, we find that Groups 2 and 3 incur a utility penalty over and above what CPT would predict for the three-outcome lotteries relative to the two-outcome lotteries (see Figure 12 in Online Appendix E.1), roughly replicating the empirical pattern found in Bernheim and Sprenger (2020). Group 1 incurs less of a utility penalty than CPT predicts. Taking a population-weighted average yields that the simulated certainty equivalent decreases by \$0.15. Figure 13 shows that we have analogous findings in the "split-high" case where the two-outcome lottery $\{\$2, 0.4; \$3, 0.6\}$ is compared to $\{\$2, 0.4; \$3 - \epsilon, 0.3; \$3 + \epsilon, 0.3\}$.

9 Conclusion

We found that the CPT-simplicity model comes close to matching machine learning algorithms in making outsample predictions and outperforms simplicity theory, CPT, PT-Simplicity, and PT. Our results show that simplicity matters for prediction;

[25] The lowest γ we found in the three-group CPT-Simplicity model is around the lowest representative agent γ in the papers discussed in this section; similarly, the highest γ in the three groups roughly corresponds to the highest γ from the representative agent literature.

[26] The payoffs they used are 10 times as large, and they bound their multiple price list at a minimum of zero and a maximum of the highest outcome in the lottery; we impose the same restrictions on our predicted certainty equivalents.

this is confirmed in the Goodman and Puri (2022) study of the preferences of traders in binary option markets. Also, while we use a particular function for simplicity cost as a function of support size; it would be interesting to consider other functional forms, and to incorporate other aspects of simplicity into the cost function. Finally, we found that financial literacy, gender, and income help predict group membership, i.e. the best fitting form of risk preferences. It would be interesting to better understand the link between these characteristics and and tolerance for complexity.

References

Abdellaoui, Mohammed. 2000. "Parameter-Free Elicitation of Utility and Probability Weighting Functions." *Managment Science*, 46: 1497–1512.

Andersen, Steffen, Glenn W Harrison, and E Elisabet Rutström. 2006. "Choice Behavior, Asset Integration and Natural Reference Points." *Working Paper*.

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström. 2006. "Elicitation Using Multiple Price List Formats." *Experimental Economics*, 9: 383–405.

Andreoni, James, and Charles Sprenger. 2011. "Uncertainty Equivalents: Testing the Limits of the Independence Axiom." *NBER*. https://www.nber.org/papers/w17342.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis*, 20: 351–368. Publisher: Cambridge University Press.

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica*, 88: 1363–1409.

Bleichrodt, Han, and Jose Luis Pinto. 2000. "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis." *Management science*, 46: 1485–1496.

Bodoh-Creed, Aaron, Jörn Boehnke, and Brent Hickman. 2018. "Using Machine Learning to Predict Price Dispersion." *Working paper*.

Booij, Adam S., Bernard M. S. van Praag, and Gijs van de Kuilen. 2010. "A Parametric Analysis of Prospect Theory's Functionals for the General Population." *Theory and Decision*, 68: 115–148.

Bradley, P.S, K.P Bennett, and A Demiriz. 2000. "Constrained K-Means Clustering." *Technical Report MSR-TR-2000-65, Microsoft Research, Redmond, WA*.

Braga-Neto, U. M., and E. R. Dougherty. 2004. "Is cross-validation valid for small-sample microarray classification?" *Bioinformatics*, 20: 374–380.

Brandimarte, P. 2014. *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics.* Wiley.

Bruhin, Adrian, Thomas Epper, and Helga Fehr-Duda. 2010. "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica*, 78: 1375–1412.

Camerer, Colin F., and Teck-Hua Ho. 1994. "Violations of the Betweenness Axiom and Nonlinearity in Probability." *Journal of Risk and Uncertainty*, 8: 167–196.

Cerreia-Vioglio, Simone, David Dillenberger, and Pietro Ortoleva. 2015. "Cautious Expected Utility and the Certainty Effect." *Econometrica*, 83: 693–728.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2017. "Willingness to Pay and Willingness to Accept are Probably Less Correlated than You Think." *NBER*. https://www.nber.org/papers/w23954.

Conlisk, John. 1989. "Three Variants on the Allais Example." *The American Economic Review*, 79: 392–407.

Conte, Anna, John D. Hey, and Peter G. Moffatt. 2011. "Mixture Models of Choice under Risk." *Journal of Econometrics*, 162: 79–88.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the *EM* Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)*, 39: 1–22.

Dillenberger, David. 2010. "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior." *Econometrica*, 78: 1973–2004.

D'Souza, R.N, P.Y Huang, and F.C Yeh. 2020. "Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size." *Scientific Reports*, 824.

Eckel, Catherine C., and Philip J. Grossman. 2008. "Chapter 113 Men, Women and Risk Aversion: Experimental Evidence." In *Handbook of Experimental Economics Results*. Vol. 1, 1061–1073. Elsevier.

Efron, Bradley, and Robert Tibshirani. 1997. "Improvements on Cross-Validation: The 632+ Bootstrap Method." *Journal of the American Statistical Association*, 92: 548–560.

Enke, Benjamin, and Thomas Graeber. 2021. "Cognitive Uncertainty." *NBER*. https://www.nber.org/papers/w26518.

Erev, Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan M. Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West, and Christian Lebiere. 2010. "A Choice Prediction Competition: Choices from Experience and from Description." *Journal of Behavioral Decision Making*, 23: 15–47.

Erev, Ido, Eyal Ert, Ori Plonsky, Doron Cohen, and Oded Cohen. 2017. "From Anomalies to Forecasts: Toward a Descriptive Model of Decisions under Risk, under Ambiguity, and from Experience." *Psychological Review*, 124: 369–409.

Etchart-Vincent, Nathalie. 2009. "Probability Weighting and the 'Level' and 'Spacing' of Outcomes: An Experimental Study over Losses." *Journal of Risk and Uncertainty*, 39: 45–63.

Fan, Chinn-Ping. 2002. "Allais Paradox in the Small." *Journal of Economic Behavior & Organization*, 49: 411–421.

Fehr-Duda, Helga, and Thomas Epper. 2012. "Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences." *Annual Review of Economics*, 4: 567–593.

Freeman, David J., Yoram Halevy, and Terri Kneeland. 2019. "Eliciting Risk Preferences Using Choice Lists." *Quantitative Economics*, 10: 217–237.

Fudenberg, Drew, and Indira Puri. 2021. "Evaluating and Extending Theories of Choice Under Risk." *Working Paper.*

Fudenberg, Drew, and Indira Puri. 2022. "Simplicity and Probability Weighting in Choice under Risk." Vol. 112, 421–25.

Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan. 2020. "Measuring the Completeness of Economic Models." *Working Paper*. http://economics.mit.edu/files/20972.

Fudenberg, Drew, Wayne Gao, and Annie Liang. 2022. "How Flexible is that Functional Form? Quantifying the Restrictiveness of Theories." *http://economics.mit.edu/files/20947.*

Goodman, Aaron, and Indira Puri. 2022. "Arbitrage in the Binary Option Market: Distinguishing Behavioral Biases." *Working Paper.*

Hajimoladarvish, Narges. 2017. "Very Low Probabilities in the Loss Domain." *The Geneva Risk and Insurance Review*, 42: 41–58.

Hansen, Bruce E. 2022. *Econometrics.* Princeton:Princeton University Press.

Hara, Kotaro, Abi Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey Bigham. 2017. "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk." Number: arXiv:1712.05796 arXiv:1712.05796.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, 62: 1251–1289.

Harrison, Glenn W. 1994. "Expected Utility Theory and the Experimentalists." In *Experimental Economics.* Springer-Verlag.

Harrison, Glenn W., and E. Elisabet Rutström. 2009. "Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral." *Experimental Economics*, 12: 133–158.

Harrison, Glenn W., Steven J. Humphrey, and Arjan Verschoor. 2010. "Choice under Uncertainty: Evidence from Ethiopia, India and Uganda." *The Economic Journal*, 120: 80–104.

Huck, Steffen, and Wieland Müller. 2012. "Allais for All: Revisiting the Paradox in a Large Representative Sample." *Journal of Risk and Uncertainty*, 44: 261–293.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47: 263–292.

Ke, Shaowei, Chen Zhao, Zhaoran Wang, and Sung-Lin Hsieh. 2020. "Behavioral Neural Networks." *SSRN*. https://ssrn.com/abstract=3633548.

Lusardi, Annamaria, and Olivia S Mitchelli. 2007. "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education." *Business Economics*, 42: 35–44.

Moffatt, Peter G., Stefania Sitzia, and Daniel John Zizzo. 2015. "Heterogeneity in Preferences towards Complexity." *Journal of Risk and Uncertainty*, 51: 147–170.

Mongin, Philippe. 2019. "The Allais Paradox: What It Became, What It Really Was, What It Now Suggests to Us." *Economics and Philosophy*, 35: 423–459.

Moss, Aaron J, Cheskie Rosenzweig, Shalom Noach Jaffe, Richa Gautam, Jonathan Robinson, and Leib Litman. 2021. "Bots or inattentive humans? Identifying sources of low-quality data in online platforms." *PsyArxiv*.

Neilson, William, and Jill Stowe. 2002. "A Further Examination of Cumulative Prospect Theory Parameterizations." *Journal of Risk and Uncertainty*, 24: 31–46.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2021. "Data quality of platforms and panels for online behavioral research." *Behavior Research Methods*.

Peysakhovich, Alexander, and Jeffrey Naecker. 2017. "Using Methods from Machine Learning to Evaluate Behavioral Models of Choice under Risk and Ambiguity." *Journal of Economic Behavior & Organization*, 133: 373–384.

Piegorsch, Walter W. 2015. *Statistical data analytics: foundations for data mining, informatics, and knowledge discovery.* Chichester, West Sussex:John Wiley & Sons Inc.

Plonsky, Ori, Apel Reut, Ido Erev, Eyal Ert, and Moshe Tennenholtz. 2017. "When and How Can Social Scientists Add Value to Data Scientists? A Choice Prediction Competition for Human Decision Making." *CPC White Paper*. https://cpc-18.com/wp-content/uploads/2018/03/cpc18-white-paper-march-update.pdf.

Plonsky, Ori, Ido Erev, Tamir Hazan, and Moshe Tennenholtz. 2016. "Psychological Forest: Predicting Human Behavior." *SSRN*. https://ssrn.com/abstract=2816450.

Puri, Indira. 2020. "Preference for Simplicity." *SSRN*. https://ssrn.com/abstract=3253494.

Puri, Indira. 2022. "Simplicity and Risk." *Working Paper*.

Quiggin, John. 1991. "Comparative statics for rank-dependent expected utility theory." *Journal of Risk and Uncertainty*, 4: 339–350.

Rieger, Marc Oliver, Mei Wang, and Thorsten Hens. 2017. "Estimating cumulative prospect theory parameters from an international survey." *Theory and Decision*, 82: 567–596.

Sonsino, Doron, Uri Benzion, and Galit Mador. 2002. "The Complexity Effects on Choice with Uncertainty – Experimental Evidence." *The Economic Journal*, 112: 936–965.

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review*, 100: 557–571.

Tulchinsky, Igor, ed. 2020. *Finding alphas: a quantitative approach to building trading strategies.* . Second edition ed., Chichester, West Sussex:Wiley.

Tversky, Amos, and Craig R Fox. 1995. "Weighing Risk and Uncertainty." *Psychological Review*, 102: 269.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and uncertainty*, 5: 297–323.

Viscusi, W Kip. 1989. ""Prospective reference theory: Toward an explanation of the paradoxes." *Journal of Risk and Uncertainty*.

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity.* Cambridge:Cambridge University Press.

Wu, George, and Richard Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science*, 42: 1676–1690.

Online Appendix A Data

A.1 Random Lottery Generation Procedure

There are 30 lotteries: 10 with 2 outcomes, 10 with 4, and 10 with 6. We generate these using a greedy algorithm:

1. Randomly generate 10 4-outcome lotteries.
2. Randomly generate 500,000 2-outcome and 500,000 6-outcome lotteries.
3. For each 4 outcome lottery $p^{(4)}$, in the entire set of n outcome lotteries $p^{(n)}$, $n \in \{2, 6\}$, find the n-outcome lottery that minimizes the loss function

$$\begin{aligned} &0.4|\text{Mean}(p^{(4)}) - \text{Mean}(p^{(n)})| + 0.1|\text{Var}(p^{(4)}) - \text{Var}(p^{(n)})| \\ &+0.1|\text{Skew}(p^{(4)}) - \text{Skew}(p^{(n)})| + 0.4|\text{Range}(p^{(4)}) - \text{Range}(p^{(n)})| \end{aligned}$$

By randomly generate, we mean that the following process was used:

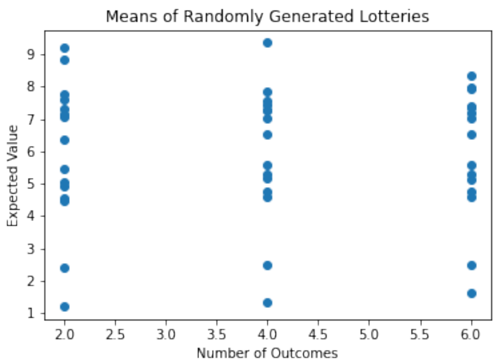
- Set remaining probability to 1, and the outcome set O to the empty set.
- Until the desired number of outcomes n is reached:
  - Uniformly at random draw a number between 1 and 10, which is a multiple of \$0.25. If that number is not in the outcome set, add it to the outcome set. If that number is in the outcome set already, then repeat this step until drawing an outcome not already in the outcome set.
  - Uniformly at random draw a probability between 0.05 and remaining probability$-$ $(n - |O|) * 0.05$, such that this probability is a multiple of 0.05.

The resulting lotteries do not differ significantly in mean, variance, or skewness by number of outcomes. Figure 4 plots each measure against the number of outcomes. Kolmogorov-Smirnoff tests do not reject that for each number of outcomes these moment distributions are identical to each other.

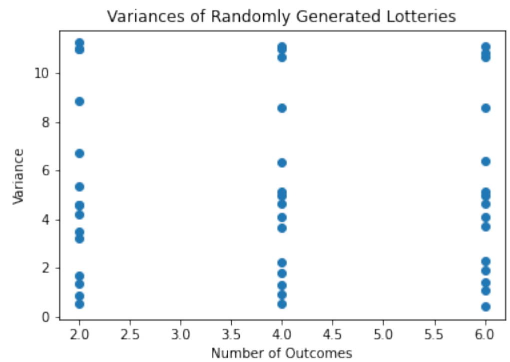
A.2 Financial Literacy Questions

The financial literacy questions asked are:

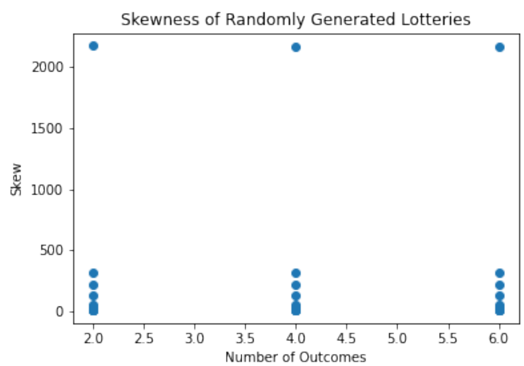
1. If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?
2. If 5 people all have the winning number in the lottery and the prize is 2 million dollars, how much will each of them get?



(A)



(B)



(C)

FIGURE 4: Moments of the Lottery and Number of Outcomes

3. Let's say you have 200 dollars in a savings account. The account earns 10 percent interest per year. How much would you have in the account at the end of two years?

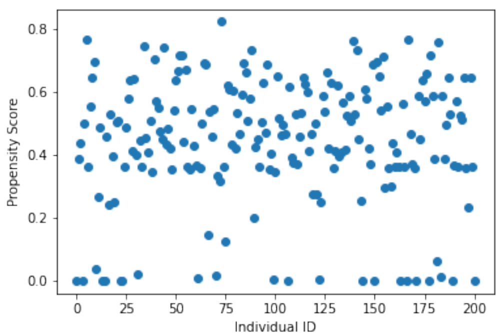
A.3 Demographic Summary

Table 4: Demographics and Financial Literacy, Overall Sample

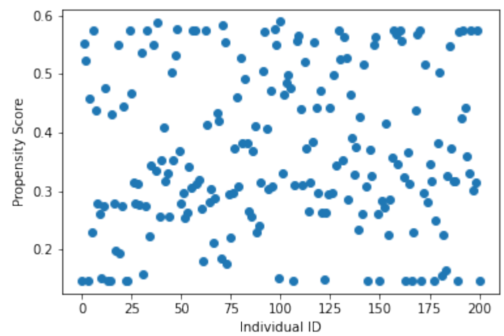
Age	
20-29	19%
30-39	37%
40-49	20%
50-59	14%
> 60	8%
Gender	
Male	58%
Female	42%
Education	
High school graduate	13%
Some college	18%
2-Year degree	11%
4-Year degree	47%
Master's degree	8%
Doctoral degree	1%
Professional degree (ex. JD, MD)	1%
Employment	
Paid employee	64%
Self-employed	26%
Not working	10%
Household Income	
< $10,000	4%
$10,000-$24,999	11%
$25,000-$49,999	33%
$50,000-$74,999	21%
$75,000-$99,999	13%
$100,000-$124,999	4%
$125,000-$149,999	4%
> $150,000	7%
Prefer not to answer	1%
Financial Literacy	
Score = 0	1%
Score = 1	44%
Score = 2	54%
Respondents	201

Online Appendix B Group Membership

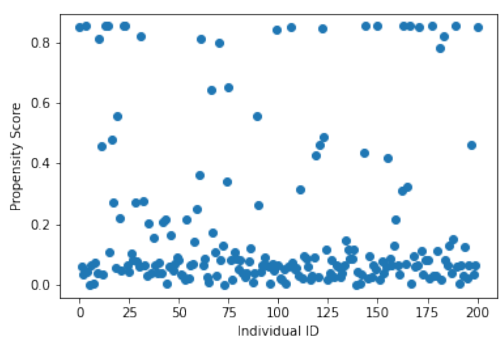
B.1 Propensity Scores



(A) Group 1



(B) Group 2



(C) Group 3

FIGURE 5: Propensity Scores by Group

B.2 Group Membership and Demographics

Table 5 shows the results of regressing the probability of belonging to each group, for each individual, on demographic characteristics, where probabilities of belonging are taken over all bootstrap iterations.

Table 5: Group Membership Regressions

	Complexity Averse	Complexity Averse (Stronger)	Prob. Weighers
Financial Literacy	0.0314*	0.0227*	-0.0541**
	(0.019)	(0.012)	(0.025)
Age	0.0005	0.0003	-0.0008
	(0.001)	(0.001)	(0.002)
Income	0.0007**	0.000	-0.0007*
	(0.000)	(0.000)	(0.000)
College-Educated	-0.0077	0.0094	-0.0017
	(0.031)	(0.020)	(0.038)
Male	0.0675**	-0.0017	-0.0658
	(0.031)	(0.020)	(0.040)
Employed	0.0350	0.0303	-0.0654
	(0.051)	(0.030)	(0.065)
N.Individuals	201	201	201
R^2	0.015	0.002	0.035

Standard errors clustered at the individual level. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, , *** $p < 0.01$

Online Appendix C Analysis

C.1 Validation Performance

Table 6: Validation Performance of Models by Number of Groups

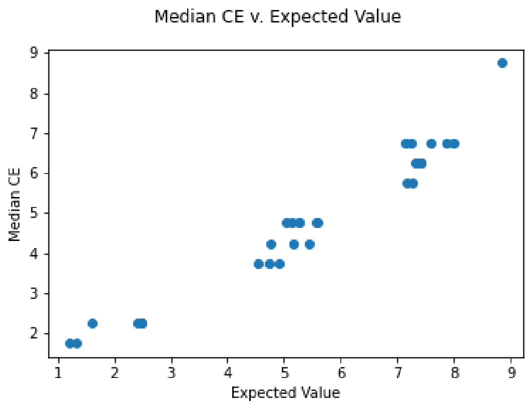
Model	**Groups**	**Validation MSE**
Expected Utility	1	3.281 (0.191)
	2	3.095 (0.190)
	3	3.146 (0.251)
Prospect Theory	1	4.404 (0.395)
	2	3.190 (0.389)
	3	2.823 (0.256)
Simplicity Theory	1	3.110 (0.181)
	2	2.536 (0.140)
	3	2.485 (0.253)
Cumulative Prospect Theory	1	3.048 (0.131)
	2	2.520 (0.116)
	3	2.511 (0.240)
PT-Simplicity	1	3.256 (0.535)
	2	2.625 (0.159)
	3	2.624 (0.262)
CPT-Simplicity	1	3.068 (0.141)
	2	2.507 (0.112)
	3	2.472 (0.256)

Standard error in parentheses.

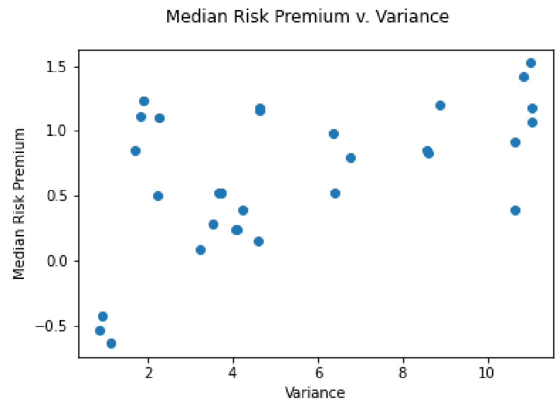
Table 7: Machine Learning Algorithms: Best Validation Performance

Model	**Validation MSE**
Gradient Boosting Tree	1.954
Neural Network	1.668
K-Means	1.532

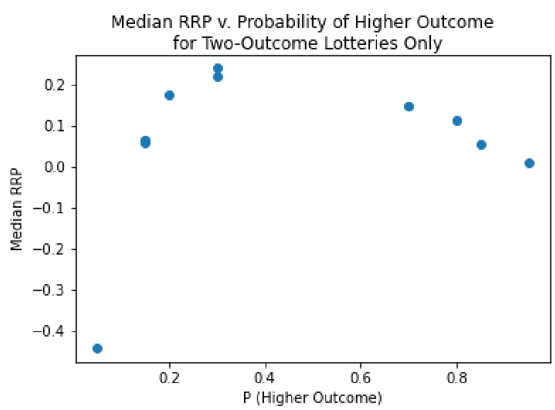
C.2 Data Visualization: Figures



(A) Expected Value

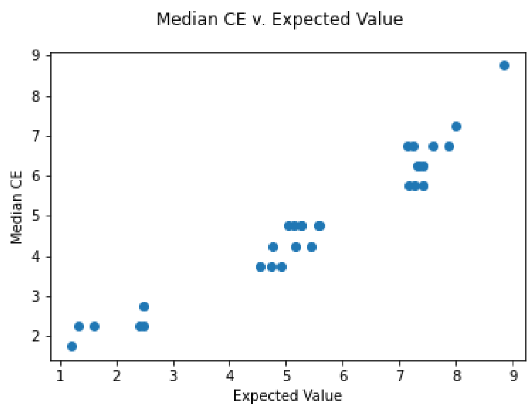


(B) Variance

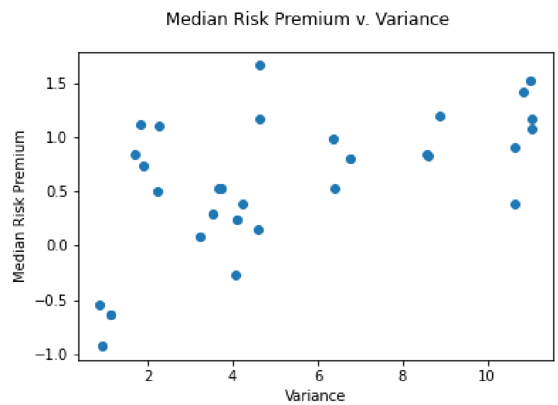


(C) Probability Weighting

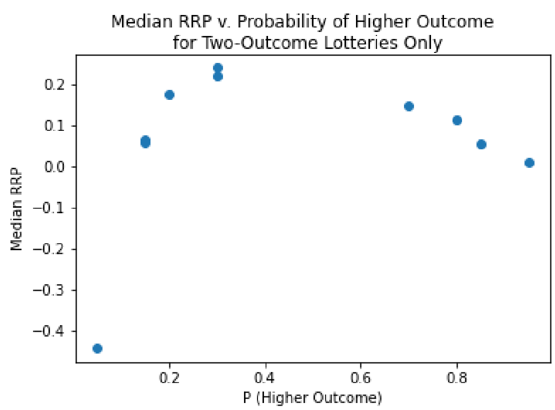
FIGURE 6: Data Overall: Non-Parametric Behavior



(A) Expected Value

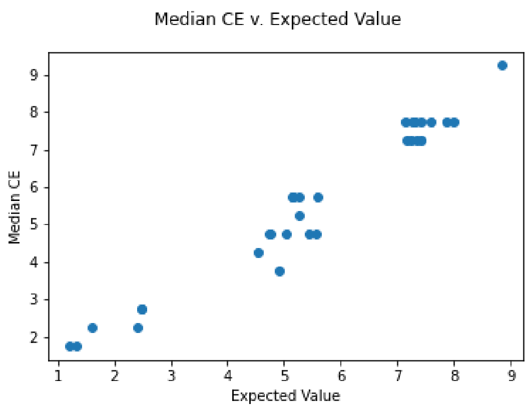


(B) Variance

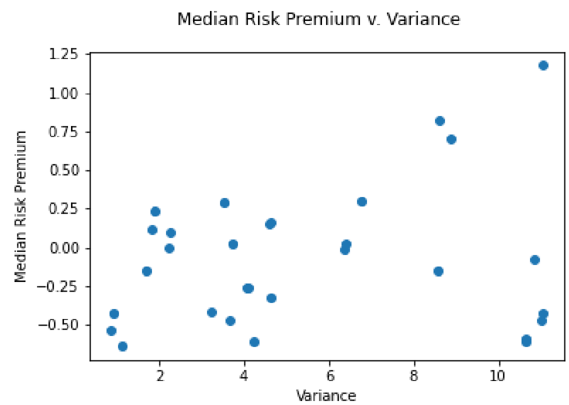


(C) Probability Weighting

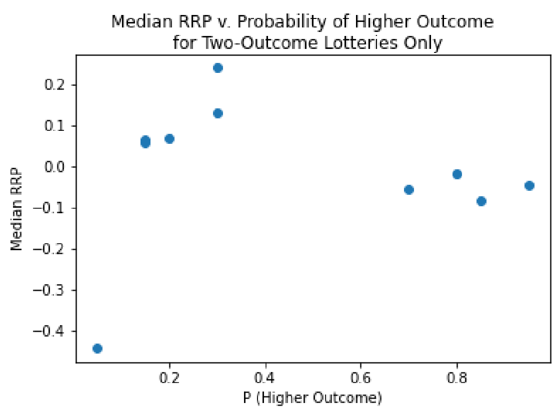
FIGURE 7: Group 1: Non-Parametric Behavior



(A) Expected Value

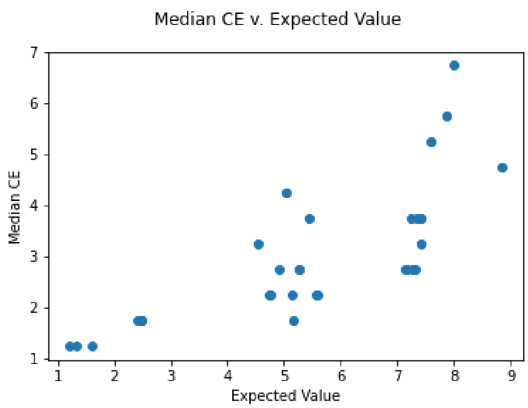


(B) Variance

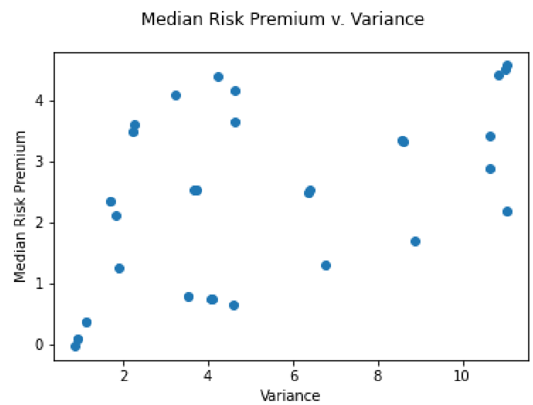


(C) Probability Weighting

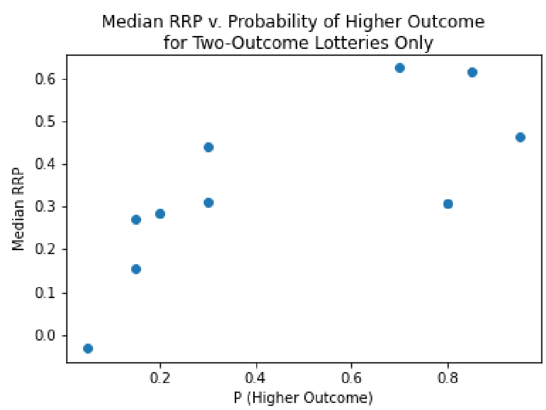
Figure 8: Group 2: Non-Parametric Behavior



(A) Expected Value



(B) Variance



(C) Probability Weighting

Figure 9: Group 3: Non-Parametric Behavior

C.3 CPT Parameter Estimates

Table 8: CPT Parameters Estimated on All Outcomes

γ	0.79	0.83	0.24
	(0.13)	(0.19)	(0.03)
α	0.59	0.90	1.35
	(0.33)	(0.31)	(0.17)
Number	95	77	30
	(10)	(7)	(7)

Standard deviation in parentheses.

Online Appendix D Supplemental Analysis

D.1 Varying γ by number of outcomes

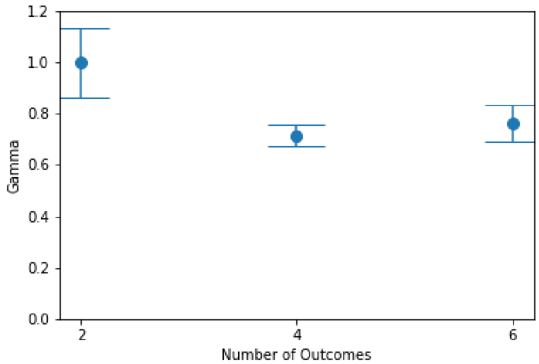
Here we allow γ to vary by number of outcomes, using the group assignments found by the three-group CPT model, where we assign a person to a group if their propensity score for that group exceeds 50%. Since this test is for varying γ and not varying αs[27], we hold α for each group fixed at the value when estimating a representative agent CPT model for that group. Figure 10 shows point estimates and 95% confidence intervals for each γ estimate. For Groups B and C, γ's do not vary statistically by number of outcomes, nor are they much different numerically. For Group A, γ at 2 outcomes is higher than γ at 4, but between 4 and 6 outcomes the γ's are statistically indistinguishable and numerically similar.

D.2 Dominance Relations

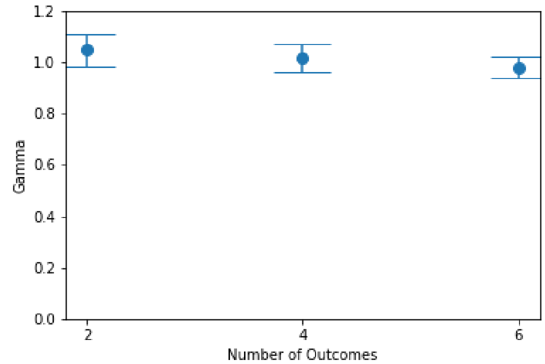
Because we have 30 lotteries, there are 435 pairwise comparisons, and in 177 of these one lottery first-order stochastically dominates the other. We say that define a dominance violation occurs for participant i in lottery pair p, q if lottery p dominates lottery q, but participant i provides at least as high a a certainty equivalent for q as for p.[28] Figure 11 plots frequency of dominance relations by subject, sorted in increasing order. The x-axis is an arbitrary subject number (e.g subject 1, subject 2,...), and the y axis is the number of times the participant exhibits pairwise dominance violations. On average, subjects have 26.95 pairwise dominance violations, corresponding to 3% of all pairwise comparisons and 15% of all possible dominance relations. On the subset of 135 pairs of lotteries with the same

[27] α's that vary by number of outcomes are evidence for complexity aversion, as the theory in Puri (2022) shows.

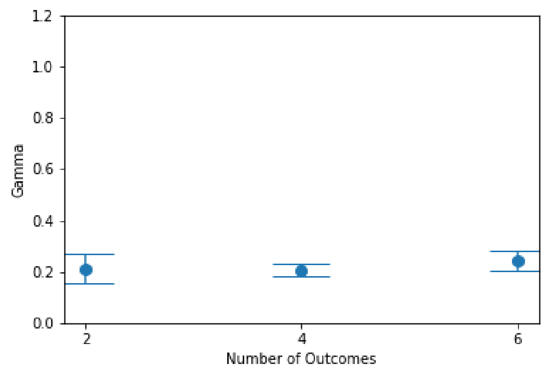
[28] Because participants are restricted to certainty equivalents on a finite grid, this is an upper bound on the number of actual dominance violations.



(A) CPT Group 1 γ's (99 people total)



(B) CPT Group 2 γ's (47 people total)



(C) CPT Group 3 γ's (25 people total)

Figure 10: Allowing γ to Vary by Number of Outcomes

supports, there are 61 pairs where one lottery dominates the other, and on these lotteries subjects have an average of 9.6 dominance violations. [29]

We now drop all lotteries that are dominated by some other lottery, and re-run our cross validation procedure on these for the three-group CPT and CPT-Simplicity models. Since this procedure uses a subset of lotteries, we drop the requirement that the cross-validation split has balanced complexity levels in each fold, but try to have at least some complexity-balanced folds. The resulting procedure shows that CPT-Simplicity continues to substantially outperform CPT, with the outsample error of CPT-Simplicity 10% less than that of CPT. Table 9 shows the outsample performance for the three-group CPT-Simplicity, CPT, Simplicity, and PT-Simplicity. The ranking of CPT-Simplicity first, then Simplicity, is similar to our overall results in Table 1. One change is that PT-Simplicity outperforms

[29] There are 18 pairs of two-outcome lotteries that are ranked by first-order dominance, and on these pairs there were on average 2.7 dominance violations per subject, or 15% of the possible violations.

CPT on this data.

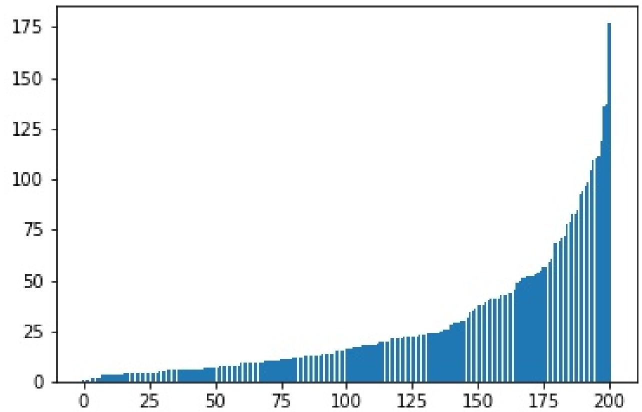


FIGURE 11: Number of Pairwise Dominance Violations by Subject

Online Appendix E Applications

E.1 Event Splitting

Figures 12 and 13 show the results of our simulations for the split low and split high tests of Bernheim and Sprenger (2020), respectively. Observe that the disutility incurred from splitting is over and above the CPT prediction for the complexity averse groups, while the utility added from splitting is over and above the CPT prediction for the complexity loving groups.

Online Appendix F Experimental Instructions

Screenshots of the experimental instructions and comprehension questions are shown in Figures 14 - 20. In addition, at the end of the experiment, we asked participants, in a free response question, 'Do you felt you understood the task at the start of the experiment? If not, when did you start understanding it?'

Table 10 lists the 30 randomly generated lotteries used.

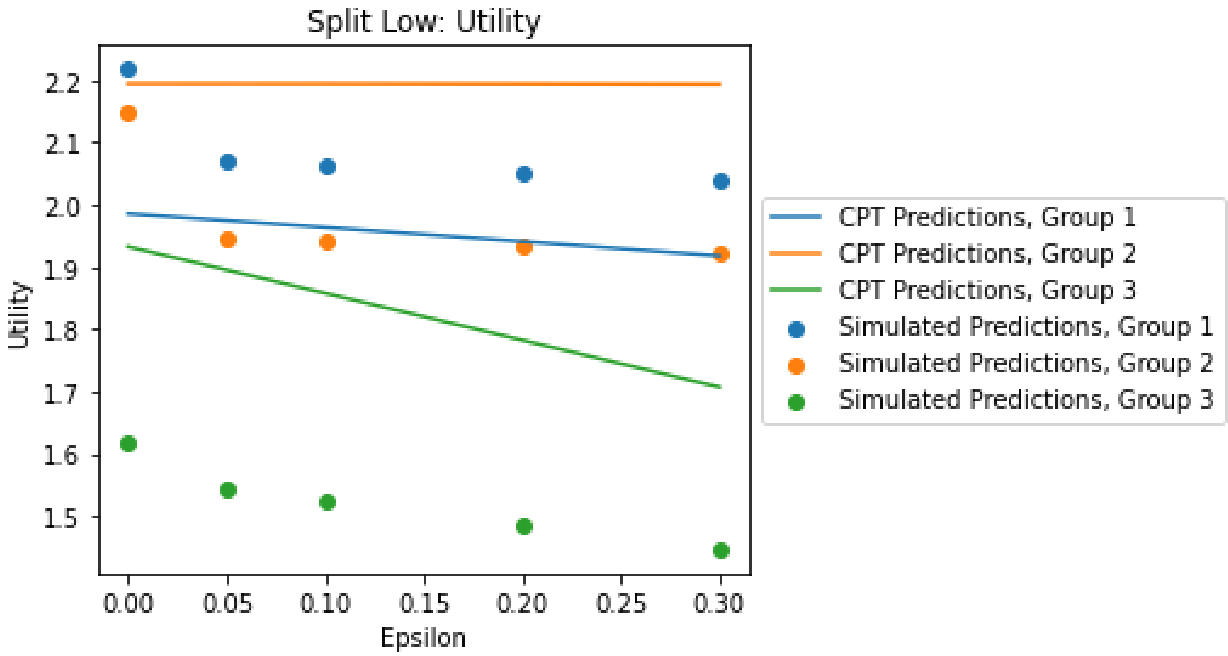


FIGURE 12: Split Low

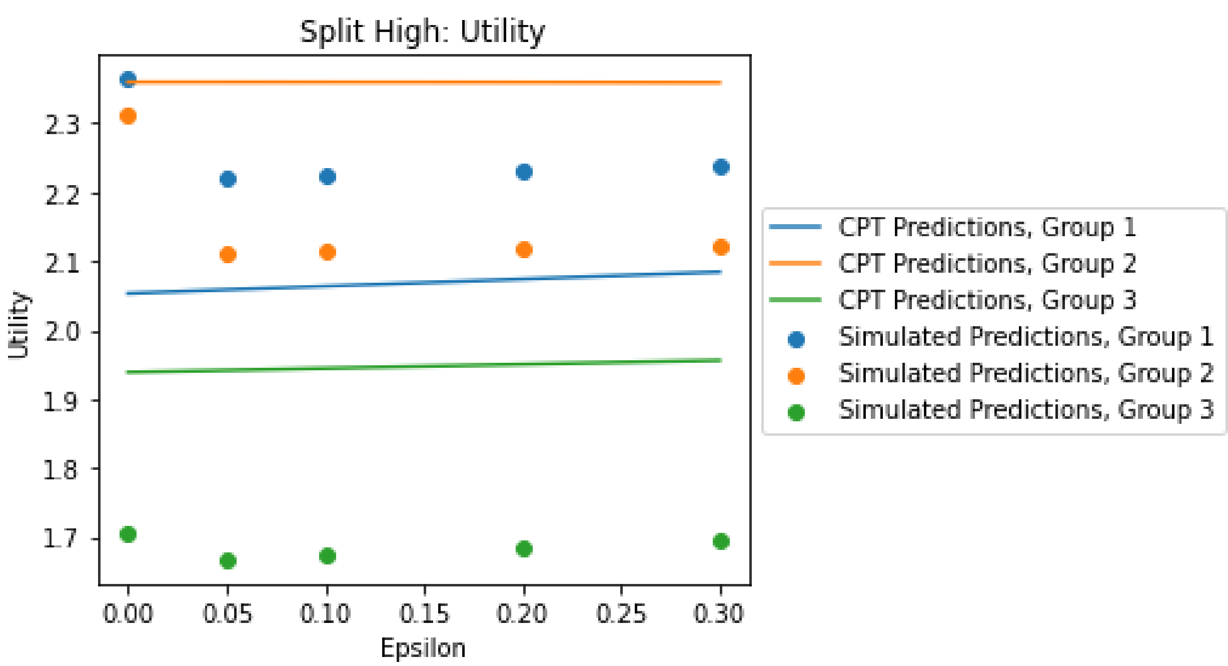


FIGURE 13: Split High

Model	Test MSE
CPT-Simplicity	2.365
	(0.200)
Simplicity	2.378
	(0.186)
PT-Simplicity	2.461
	(0.219)
CPT	2.594
	(0.274)
PT	3.253
	(0.510)
EU	3.845
	(0.316)

Table 9: Outsample Performance of Theories on Data that Excludes Dominance Violations

Table 10: Lotteries

N. Outcomes	O1	P1	O2	P2	O3	P3	O4	P4	O5	P5	O6	P6	EV	Var	Skew
4	8.75	0.4	8.5	0.35	5.0	0.1	3.0	0.15					7.42	4.63	50.1
4	7.75	0.15	7.0	0.35	2.5	0.4	1.25	0.1					4.74	6.37	12.21
4	5.25	0.05	2.5	0.05	1.25	0.15	1.0	0.75					1.32	0.92	10.28
4	9.5	0.05	7.75	0.55	7.0	0.35	1.25	0.05					7.25	2.21	127.33
4	9.75	0.05	8.0	0.4	6.5	0.15	1.25	0.4					5.16	10.66	8.42
4	9.75	0.15	8.25	0.25	4.5	0.4	1.25	0.2					5.57	8.61	12.52
4	7.75	0.15	6.25	0.45	5.5	0.1	2.5	0.3					5.27	3.66	28.69
4	9.0	0.35	8.75	0.25	7.25	0.15	5.75	0.25					7.86	1.82	214.46
4	7.5	0.05	6.75	0.1	4.75	0.05	1.5	0.8					2.49	4.11	7.19
4	9.75	0.6	9.0	0.05	3.5	0.15	2.25	0.2					7.28	11.01	16.45
6	9.0	0.6	7.5	0.05	6.75	0.05	5.25	0.15	3.5	0.1	3.25	0.05	7.41	4.63	50.3
6	7.5	0.4	4.75	0.25	2.5	0.05	1.75	0.2	1.25	0.05	1.0	0.05	4.77	6.39	12.22
6	5.25	0.05	2.75	0.15	2.25	0.05	1.75	0.05	1.25	0.15	1.0	0.55	1.61	1.1	10.39
6	10.0	0.05	8.25	0.15	7.75	0.35	7.25	0.35	5.5	0.05	1.75	0.05	7.35	2.27	128.29
6	9.5	0.1	7.5	0.4	7.0	0.1	2.0	0.05	1.25	0.15	1.0	0.2	5.14	10.67	8.32
6	9.75	0.05	8.0	0.5	3.75	0.15	2.5	0.1	1.75	0.1	1.25	0.1	5.6	8.59	12.41
6	7.75	0.3	6.75	0.1	4.5	0.3	3.5	0.15	2.75	0.1	2.5	0.05	5.28	3.73	28.73
6	9.75	0.25	9.5	0.05	9.0	0.05	8.5	0.15	7.5	0.1	6.5	0.4	7.99	1.89	213.59
6	7.0	0.15	4.25	0.05	2.5	0.05	1.75	0.4	1.5	0.1	1.0	0.25	2.49	4.08	7.14
6	10.0	0.5	9.25	0.05	7.25	0.1	3.0	0.2	2.75	0.05	2.5	0.1	7.18	10.86	16.37
2	8.0	0.85	2.25	0.15									7.14	4.22	50.48
2	9.75	0.2	3.25	0.8									4.55	6.76	12.11
2	5.25	0.05	1.0	0.95									1.21	0.86	10.3
2	9.25	0.95	1.0	0.05									8.84	3.23	129.35
2	10.0	0.3	2.75	0.7									4.92	11.04	8.58
2	10.0	0.3	3.5	0.7									5.45	8.87	12.49
2	9.5	0.15	4.25	0.85									5.04	3.51	29.43
2	8.25	0.8	5.0	0.2									7.6	1.69	215.85
2	7.5	0.15	1.5	0.85									2.4	4.59	6.73
2	9.5	0.7	2.25	0.3									7.32	11.04	16.46

Thank you for taking part in this survey. The base payment for completing the survey satisfactorily is $1.50. Please note that there are several attention and comprehension checks throughout the survey. **All such questions say 'CORRECT RESPONSE REQUIRED FOR PAYMENT' at the top. If you incorrectly answer an attention or comprehension check question, you will not be approved for payment.**

The attention and comprehension checks are not designed to trick or trap you: if you are reading the instructions and responding thoughtfully then these should not be exceedingly difficult. The code for MTurk will be provided at the **end** of the survey. Again, if you answer an attention or comprehension check question incorrectly, you will not be approved for payment.

Thank you for taking part in this survey. The base payment for completing the survey satisfactorily is $1.50. Please note that there are several attention and comprehension checks throughout the survey. **All such questions say 'CORRECT RESPONSE REQUIRED FOR PAYMENT' at the top. If you incorrectly answer an attention or comprehension check question, you will not be approved for payment.**

The attention and comprehension checks are not designed to trick or trap you: if you are reading the instructions and responding thoughtfully then these should not be exceedingly difficult. The code for MTurk will be provided at the **end** of the survey. Again, if you answer an attention or comprehension check question incorrectly, you will not be approved for payment.

FIGURE 14: Experimental Instructions

SAMPLE QUESTION

Recall that it is in your interest to answer each question honestly, as any question may be selected for payment. Here is an example of a question and bonus payment.

Suppose Bob answers the question below as shown, and the row of Option Safe of $5.50 vs Option Lottery, was selected for payment. Since Bob chose Option Safe in this row, he would receive a bonus of $5.50 .

Conversely, if the row of Option Safe of $2.50 vs Option Lottery were selected for payment, then since Bob chose Option Lottery in that row, he would receive the result of the simulated lottery. As described below, this lottery pays $7.00 with a 50 in 100 chance, $3 with a 50 in 100 chance.

The LOTTERY for this question is:

Probability	Outcome
50%	$7
50%	$3

In each row, please decide whether you prefer Option Safe or Option Lottery.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00	●	○
Option Safe = $9.50	●	○
Option Safe = $9.00	●	○
Option Safe = $8.50	●	○
Option Safe = $8.00	●	○
Option Safe = $7.50	●	○
Option Safe = $7.00	●	○
Option Safe = $6.50	●	○
Option Safe = $6.00	●	○
Option Safe = $5.50	●	○
Option Safe = $5.00	○	●
Option Safe = $4.50	○	●
Option Safe = $4.00	○	●
Option Safe = $3.50	○	●
Option Safe = $3.00	○	●
Option Safe = $2.50	○	●
Option Safe = $2.00	○	●
Option Safe = $1.50	○	●
Option Safe = $1.00	○	●
Option Safe = $0.50	○	●

FIGURE 15: Experimental Instructions Continued: Sample Question

On this page, you'll practice responding and understanding how bonus payments are made. Following this practice, we will ask you comprehension questions, for which correct responses are required for payment.

Here is the practice lottery (which you will **not** see in the actual task).

Probability	**Outcome**
33%	$10
33%	$5
33%	$1

In each row, please decide whether you prefer Option Safe or Option Lottery.
Once you click on an option, we'll fill in the rest of the choices for you so that they make sense (this will save you time!). However, you should keep clicking on options that you prefer until the choice in each line is what you would like.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00	○	○
Option Safe = $9.50	○	○
Option Safe = $9.00	○	○
Option Safe = $8.50	○	○
Option Safe = $8.00	○	○
Option Safe = $7.50	○	○
Option Safe = $7.00	○	○
Option Safe = $6.50	○	○
Option Safe = $6.00	○	○
Option Safe = $5.50	○	○
Option Safe = $5.00	○	○
Option Safe = $4.50	○	○
Option Safe = $4.00	○	○
Option Safe = $3.50	○	○
Option Safe = $3.00	○	○
Option Safe = $2.50	○	○
Option Safe = $2.00	○	○
Option Safe = $1.50	○	○
Option Safe = $1.00	○	○
Option Safe = $0.50	○	○

Figure 16: Experimental Instructions Continued: Practice Round

You indicated that, at row 10 , you switched from preferring the safe amount to preferring the lottery.

What this means for your bonus payment is the following:

Example Scenario #1
Suppose the computer selected row 1, *Option Safe of $10 v Option Lottery* for payment. You would have received a bonus of $10 since you indicated that you prefer Option Safe in this row.

Example Scenario #2
Suppose the computer selected row 20, *Option Safe of $0.50 v. Option Lottery* for payment. The computer would then have simulated the lottery, since you indicated that you prefer Option Lottery in this row. In this practice simulation, the outcome of the lottery was $1, which would have then been your bonus payment.

For your actual bonus payment, all lotteries and all rows are equally likely to be selected for payment.

Figure 17: Experimental Instructions Continued: Post Practice Round Sample

COMPREHENSION QUESTION

CORRECT RESPONSE REQUIRED FOR PAYMENT

Suppose Bob responds to the below question as follows.

The LOTTERY for this question is:

Probability	Outcome
50%	$7
50%	$3

In each row, please decide whether you prefer Option Safe or Option Lottery.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00	●	○
Option Safe = $9.50	●	○
Option Safe = $9.00	●	○
Option Safe = $8.50	●	○
Option Safe = $8.00	●	○
Option Safe = $7.50	●	○
Option Safe = $7.00	●	○
Option Safe = $6.50	●	○
Option Safe = $6.00	●	○
Option Safe = $5.50	●	○
Option Safe = $5.00	○	●
Option Safe = $4.50	○	●
Option Safe = $4.00	○	●
Option Safe = $3.50	○	●
Option Safe = $3.00	○	●
Option Safe = $2.50	○	●
Option Safe = $2.00	○	●
Option Safe = $1.50	○	●
Option Safe = $1.00	○	●
Option Safe = $0.50	○	●

Which of the following is true?

- Bob prefers $1 over taking the lottery
- Bob prefers taking the lottery over receiving $4
- Bob prefers taking the lottery over receiving $7

Figure 18: Experimental Instructions Continued: Comprehension Question 1

COMPREHENSION QUESTION

CORRECT RESPONSE REQUIRED FOR PAYMENT

Suppose Evan responds to the below question as follows.

The LOTTERY for this question is:

Probability	Outcome
50%	$7
50%	$3

In each row, please decide whether you prefer Option Safe or Option Lottery.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00	●	○
Option Safe = $9.50	●	○
Option Safe = $9.00	●	○
Option Safe = $8.50	●	○
Option Safe = $8.00	●	○
Option Safe = $7.50	●	○
Option Safe = $7.00	●	○
Option Safe = $6.50	○	●
Option Safe = $6.00	○	●
Option Safe = $5.50	○	●
Option Safe = $5.00	○	●
Option Safe = $4.50	○	●
Option Safe = $4.00	○	●
Option Safe = $3.50	○	●
Option Safe = $3.00	○	●
Option Safe = $2.50	○	●
Option Safe = $2.00	○	●
Option Safe = $1.50	○	●
Option Safe = $1.00	○	●
Option Safe = $0.50	○	●

Which of the following is true?

- Evan prefers taking the lottery to getting $6
- Even prefers $5 to taking the lottery
- Evan prefers $0.50 to taking the lottery

FIGURE 19: Experimental Instructions Continued: Comprehension Question 2

COMPREHENSION QUESTION

CORRECT RESPONSE REQUIRED FOR PAYMENT

Suppose Amanda responds to the below question as follows.

The LOTTERY for this question is:

Probability	Outcome
50%	$7
50%	$3

In each row, please decide whether you prefer Option Safe or Option Lottery.

	... I prefer Option Safe	... I prefer Option Lottery
Option Safe = $10.00	●	○
Option Safe = $9.50	●	○
Option Safe = $9.00	●	○
Option Safe = $8.50	●	○
Option Safe = $8.00	●	○
Option Safe = $7.50	●	○
Option Safe = $7.00	●	○
Option Safe = $6.50	●	○
Option Safe = $6.00	●	○
Option Safe = $5.50	●	○
Option Safe = $5.00	●	○
Option Safe = $4.50	●	○
Option Safe = $4.00	●	○
Option Safe = $3.50	●	○
Option Safe = $3.00	●	○
Option Safe = $2.50	○	●
Option Safe = $2.00	○	●
Option Safe = $1.50	○	●
Option Safe = $1.00	○	●
Option Safe = $0.50	○	●

Which of the following is true?

- Amanda prefers $6 to taking the lottery
- Amanda prefers taking the lottery to $4.50
- Amanda prefers $2 to taking the lottery

FIGURE 20: Experimental Instructions Continued: Comprehension Question 3