Varieties of Risk Elicitation *

by Daniel Friedman, Sameh Habib, Duncan James, and Sean Crockett†

January 31, 2018

Abstract

We explore a variety of risk preference elicitation procedures that involve direct choice from a set of lotteries, including budget lines (BL) and binary choice lists (HL). We find statistically significant violations of the expected utility hypothesis (EUH) consistent with disappointment aversion, and also find violations of first order stochastic dominance, but both sorts of violations are mostly small and only slightly impair the predictive power of a parametric implementation of EUH. The estimated coefficient of relative risk aversion, gamma, varies widely across individual subjects (consistent with EUH) and also across elicitation tasks (inconsistent with direct implementation of EUH). An alternative nonparametric measure of risk preferences displays similar patterns. The two risk preference measures are highly correlated with each other for each elicitation task. Each separate measure varies widely across individual subjects and across elicitation tasks, with low to nil correlation between BL tasks and HL tasks. Some of the variation across tasks can be explained by attributes such as graphical vs text representation that have no role in decision theory.

Keywords: Risk Aversion, Experiment, Elicitation, Multiple Price List

JEL Classifications: C91, D81, D89

*We are grateful for financial support from the National Science Foundation under grant SES-1357867, and for programming assistance from Matt Jee and also from Logan Collingwood, Emily Hockel and Joshua Pena. We are grateful to Brett Williams for research assistance and for supplying a proof of Proposition 1. For helpful comments we thank Roel van Veldhuizen, Frank Heinemann, Dirk Engelmann, Amma Panin, Jessie Li, and seminar participants at WZB (April 25, 2017), and ESA (San Diego, May 23, 2017).

†Friedman and Habib: University of California, Santa Cruz, Santa Cruz, CA 95064, dan@ucsc.edu and sfhabib@ucsc.edu; James: Economics Department, Fordham University dujames@fordham.edu; Crockett: Baruch College, New York, NY 10010, sean.crockett@baruch.cuny.edu.

1 Introduction

Over the last six decades, economists have proposed a wide variety of methods to elicit individual subjects' risk preferences. The methods, of which there are now many dozens, share the same ultimate scientific goal: to predict out-of-sample risky choice behavior at the population level and, if possible, at the individual level (e.g., Smith, 1989, Friedman, Isaac, James, and Sunder, 2014).

In this paper we pursue a key intermediate goal: to compare subject behavior across elicitation methods. After all, if one can not predict choices in a different laboratory elicitation task, there is little hope of predicting risky choice behavior elsewhere. We will compare population distributions of risk preference estimates across elicitation tasks, and will examine within-subject consistency.

Our focus is direct choice methods. That is, in every elicitation task we consider, individual subjects directly choose a particular lottery from some given feasible set of lotteries; there is no bidding, asking, or strategizing at any point in any procedure. To promote comparability and consistency, we construct feasible sets that hold constant across elicitation tasks the key decision-theoretic variables such as price and probability. Of course, the elicitation tasks also differ in the ways that they present these variables. For example, some tasks convey price and probability information via text, while other tasks convey the same information via spatial displays, such as a budget line.

Section 2 below surveys some relevant literature and positions our contribution. Section 3 obtains theoretical predictions, some of them slightly novel. Our point of departure is the expected utility hypothesis (EUH): a subject's risk preferences can be represented via the expectation of a personal Bernoulli function that remains stable across time and contexts. We derive implications when feasible lotteries lie on a budget line defined for two Arrow securities. We show that a scalar variable, which we call L, is a sufficient statistic for the prices and probabilities of the Arrow securities, and that the coefficient γ of relative risk aversion is an attractive index of a subject's risk preferences. After noting stronger implications for the special case of constant absolute risk aversion, we derive implications for two specific alternatives to EUH, Disappointment Aversion and Prospect Theory. We also obtain results on stochastic dominance that apply to these and many other alternatives to EUH.

Section 4 lays out the design of our experiment. It presents our budget line (BL) screen display, as well as alternative non-spatial displays called budget jars (BJ and BJn) that offer precisely the same sets of feasible lotteries at the same sets of Arrow prices and probabilities. It also shows how the Eckel-Grossman task (Eckel and Grossman, 2002, 2008) can be displayed spatially as a set of six discrete points on a budget line, and how each line from a Holt-Laury multiple price list (Holt and Laury, 2002) can either be displayed as two points on a budget line or as a text line with two radio buttons. The section then lays out the way we structure a set of 56 trials for each subject. Each of our 142 subjects completes 5 of the 6 elicitation tasks with either prices or probabilities appearing in each block of trials in monotone or random order, in a balanced manner.

Section 5 presents the results. We find statistically significant violations of the EUH consistent with disappointment aversion, and also find violations of first order stochastic dominance, but both sorts of violations are mostly small and only slightly impair the predictive power of CRRA relative to DA. Both γ and RRP, a new nonparametric measure of individual risk aversion, vary widely across individual subjects (consistent with EUH) and across elicitation tasks (inconsistent with direct implementation of EUH). The risk aversion measures RRP and γ are highly correlated with each other within each elicitation task; each separate measure shows fairly high correlations for repetitions across the same task and across isomorphic tasks. Other correlations are low, and are low to nil between BL tasks and HL tasks. Some of the variation across tasks can be explained by attributes such as graphical vs text representation that have no role in decision theory.

2 Relevant Literature

A principal goal of our paper is to assess how design attributes of elicitation tasks influence risk preference estimates, including attributes central to decision theory (such as prices and probabilities) as well as other attributes (such as the format for presenting the task) deemed irrelevant by decision theory. We now review prior work with that goal in mind.

Auction Bids. Subjects' bids in First Price Sealed Bid (FPSB) auction and knowledge of their induced values imply, through the lens of a particular bidding model (such as the Constant Relative Risk Averse Model), specific values of their CRRA risk parameter γ.[1] The research program including Cox, Roberson, and Smith (1982) and Cox, Smith, and Walker (1988) presents theory and empirical estimates. FPSB tends to generate γ estimates suggestive of risk aversion.

Alternatively, one can make inferences about subjects' risk preferences from their bids in a second price sealed bid auction (SPSB) or, in perhaps more familiar language, from variants of the BDM procedure. The original version in Becker, DeGroot, and Marschak (1964) amounts to setting a reserve price for selling the lottery in a second price auction with automated bidders; it tends to generate gamma estimates on the risk-seeking side. A minor variant called buying-BDM

[1] The logic is similar to that of solving for implied volatility, given market prices and option contract specifications, using a particular option pricing model such as (Black and Scholes, 1973).

tends to generate gamma estimates on the risk-averse side; both of the preceding are documented by Kachelmeier and Shehata (1992). Dual-to-selling and dual-to-buying versions can also be constructed, as by James (2011); dual-to-selling responses tend towards risk aversion, while dual-to-buying responses tend towards risk-seeking.

Direct Choice. Another class of elicitation methods directly asks the subject to choose one lottery from a menu of lotteries. Binary choice (Hey and Orme, 1994) and choice from longer menus of lotteries (as in Binswanger, 1980, Eckel and Grossman, 2002, 2008) are examples of direct choice methods. Another example is choice from a budget line consisting of all affordable combinations of Arrow-Debreu state dependent securities, as in Choi, Fisman, Gale, and Kariv (2007), or Andreoni and Harbaugh (2009) or Andreoni, Kuhn, and Sprenger (2015). Although it does not explicitly display a budget line, the investment game of Gneezy and Potters (1997) also falls into this category, as noted in Section 4.1 below. Taken one row at a time, the well-known multiple price list of Holt and Laury (2002) is an instance of binary choice, but overall — with all rows present at the same time and with random selection of a single row for payment — it also fits another category; we shall return to this shortly.

Inconsistency across elicitation tasks. Lichtenstein and Slovic (1971, 1973) is an early attempt to cross-reference risky choice behavior across different elicitation tasks. They compared subject behavior in (a) direct choice between two lotteries (i.e. binary choice) and (b) numerical valuation of those lotteries by means of selling-BDM. They found a great deal of inconsistency — self-contradiction, in fact — in subject behavior across the two different procedures. Subsequent study and attempted resolution of those inconsistencies since then is known as the literature on preference reversals. Collins and James (2015) demonstrate that replacing selling-BDM with dual-to-selling-BDM eliminates a majority of reversals, and that those remaining fit the pattern predicted due to the model of noise in response behavior model of Blavatskyy (2014).

Isaac and James (2000) provide another striking example of apparently inconsistent individual behavior across elicitation tasks. Subjects in their study who appear risk averse in the FPSB task often seem risk seeking in the selling-BDM task, and vice versa. This result — the scrambling of subjects' gamma estimates across different procedures — remains an intriguing result, and one which has received much subsequent support (e.g. Berg, Dickhaut, and McCabe, 2005, Loomes and Pogrebna, 2014, Sprenger, 2015, Pedroni, Frey, Bruhin, Dutilh, Hertwig, and Rieskamp, 2017).

Equivalent Procedures. From a decision-theoretic standpoint, the Holt-Laury multiple price list procedure is equivalent to the dual-to-selling version of BDM (James, 2011), which is in turn related to Grether (1981) if the latter is applied to comparison between two lotteries, each with unknown probabilities (as is done by Crockett and Crockett, 2018). This equivalence was anticipated, though not specifically identified, by Freeman, Halevy, and Kneeland (2016), who pointed out that paying one row of Holt-Laury, randomly selected, rendered Holt-Laury subject to the same non-EUT critique applied to selling-BDM by Karni and Safra (1987). In other words, Holt-Laury implemented with pay-one-randomly (POR) payment protocol (Cox et al., 2015) is a form of BDM.

The various balloon (Lejuez, Read, Kahler, Richards, Ramsey, Stuart, Strong, and Brown, 2002) or bomb elicitation (Crossetto and Filippin, 2013) tasks, experiencing some popularity at present, are equivalent to an N=2 FPSB auction. This is because the task of uncovering a zero-payoff absorbing state while being paid in a linearly increasing manner for exploration, up to but short of discovering that state, is payoff equivalent to pitting the human subject against a random draw from a uniform density [Vlower, Vupper], where the latter is actually Nash play in an N=2 FPSB. (The robot bidders used in Cox, Smith, and Walker (1988) did just this, albeit with N>2.)

Supposed Irrelevancies. Decision theory, for example using the formalism of mechanism design, can be used to classify the different elicitation procedures (e.g., Hurwicz, 1972, Smith, 1976). For example, FPSB and SPSB (and thus BDM procedures) have different cost rules and induce different strategies. (Of course, if subjects have stable preferences, these should still be recoverable from either and consistent across both procedures.) Conversely, two institutions identical in cost and allocations rules, and thus in the predicted mapping from endowments and preferences to behavior, may have differences in practical implementation. For example, FPSB and the Dutch auction have identical cost and allocation rules, but differ in transition rules, and empirically generate different behavior (Cox, Roberson, and Smith, 1982) despite predicted revenue equivalence (at least given agents who correctly perform Bayesian updating).

Yet more subtle differences in implementation exist, and may affect behavior (again, despite theoretical predictions to the contrary). For example, the response mode (see Section 4.1) afforded the subject by the experimenter may shape the subject's responses. Or the manner in which information is presented to the subject may affect behavior, even when holding response mode constant. For example, Habib, Friedman, Crockett, and James (2017) find that spatial (volume) representation of payoff and probability information in the Holt-Laury task supports different (less risk averse) behavior than text representation of the same information; this is despite algebraic and response mode (direct choice via radio button) equivalence across the two versions of the task.

Even the pattern (or lack thereof) in the way in which the exogenous parameters of the experiment are run past the decision-making agent might change behavior. Lévy-Garboua, Maafi, Masclet, and Terracol (2012) find that behavior in Holt-Laury is different depending on whether the rows, and thus state probabilities, of the Holt-Laury task are presented in a monotone order (i.e., $P_h = 0.1$ is followed by $P_h = 0.2, 0.3, ...$) or in a random sequence. Specifically, behavior is closer to risk neutrality when the rows of the task are presented in monotone order. Our experiment will revisit these and other effects deemed irrelevant by decision theory.

Positioning the present paper. There is a vast literature on testing the implications of particular generalizations of the expected utility hypothesis, and another vast literature on searching for some "best" elicitation method. Neither of these is an appropriate context for what we attempting in this paper. Rather, we seek regularities in cross-method (in)consistency. Our comparisons of individual risk preference parameters elicited across different procedures use only the simplest possible cost and allocation rule — direct choice from a given set of lotteries — in order to isolate the effects of the presentation format.

3 Theoretical Predictions

In all risk preference elicitation tasks that we consider, a subject chooses an allocation (x, y) from a compact feasible set F of Arrow securities. As the notation suggests, we assume two mutually exclusive possible states, X and Y, with known probabilities $\pi_X > 0$ and $\pi_Y > 0$ with $\pi_X + \pi_Y = 1$; a chosen allocation (x, y) pays x points in state X and y points in state Y.

According to the Expected Utility Hypothesis (EUH), each human subject has her own fixed Bernoulli function, i.e., a smooth (twice differentiable) and strictly increasing function $u : \mathbb{R} \to \mathbb{R}$, defined up to a positive affine transformation. The EUH further states that the subject's choice (x^*, y^*) solves

$$\max_{(x,y)\in F} \pi_X u(x) + \pi_Y u(y). \tag{1}$$

The art of elicitation is for the experimenter to choose the feasible set F (or a sequence of F's) so that subjects' choices reveal key aspects of their Bernoulli functions u. For some elicitation tasks, the feasible set is a standard budget set: non-negative bundles that are affordable. Since $u' > 0$, there is then no further loss of generality in replacing F by the budget constraint

$$p_x x + p_y y = m, \tag{2}$$

where m is an (implicit or explicit) endowment of cash, and $p_x > 0$ and $p_y > 0$ are the prices of the two Arrow securities. In all elicitation tasks that we study, F is a subset (sometimes a finite subset) of points satisfying (2). We normalize prices so that $p_x + p_y = 1$; this jibes with the convention that a unit of cash is the portfolio (1,1).

The first order conditions for optimization problem (1)-(2) can be written out in terms of the Lagrange multiplier λ for (2) as

$$\lambda = \frac{\pi_Y}{p_y} u'(y) = \frac{\pi_X}{p_x} u'(x), \tag{3}$$

or as

$$MRS = \frac{u'(x)}{u'(y)} = \frac{\pi_Y}{\pi_X} \frac{p_x}{p_y}, \tag{4}$$

or as

$$\ln \frac{u'(x)}{u'(y)} = -[\ln \pi_X - \ln \pi_Y - \ln p_x + \ln p_y] \equiv -L \tag{5}$$

Thus, for whichever Bernoulli function u a subject may have, the EUH implies that

1. An interior choice (x, y) is determined by ratios of state prices and probabilities.

2. The composite variable $L = \ln \pi_X - \ln \pi_Y - \ln p_x + \ln p_y$ is a sufficient statistic for prices and probabilities. Equation (5) holds at interior solutions, and corner solutions are also defined by L: corner $(\frac{m}{p_x}, 0)$ is chosen if $\ln \frac{u'(\frac{m}{p_x})}{u'(0)} \geq -L$, while corner $(0, \frac{m}{p_y})$ is chosen if $\ln \frac{u'(0)}{u'(\frac{m}{p_y})} \leq -L$.

3. When regressing log marginal rate of substitution on log price ratio and log odds, the coefficients should be equal in magnitude with opposite signs.

We will soon see that some popular generalizations of EUH obey similar rules. But first we note that we can say more in important special cases.

3.1 Special cases

For a **risk neutral** agent we have $u'(x) = u'(y) = constant$, and (3) becomes

$$\frac{\pi_Y}{p_y} = \frac{\pi_X}{p_x}. \tag{6}$$

Equation (6) can only be satisfied if $L = 0$. Otherwise we'll get a corner solution where the risk neutral person spends her entire budget on the asset with higher probability/ price ratio, viz. $x^* = 0$ if $L < 0$ and $y^* = 0$ if $L > 0$.

CRRA, a widely used parametric family of Bernoulli functions, sets $u(c|\gamma) = \frac{c^{1-\gamma}}{1-\gamma}$ where the parameter $\gamma \geq 0$ is the coefficient of relative risk aversion. (For $\gamma = 1$ the function is log utility, as can be seen using L'Hospital's rule.) Here $u'(c) = c^{-\gamma}$, so MRS $= [\frac{x}{y}]^{-\gamma}$. Inserting this into (4) and taking logs yields

$$\ln \frac{x}{y} = \frac{-1}{\gamma}[\ln \pi_Y - \ln \pi_X - \ln p_y + \ln p_x] = \frac{1}{\gamma} L. \tag{7}$$

That is, regressing log-odds of the chosen allocation on L will directly reveal (as the inverse slope) the subject's coefficient γ of relative risk aversion. Moreover, as separate regressors, all four components of L (log prices and log probabilities) should have exactly the same coefficients, $\pm\gamma^{-1}$.

3.2 Generalizations of EUT

Gul (1991) presents a model with a free parameter $\beta \geq 0$ intended to capture disappointment aversion as a probability distortion in a two state world — people make choices as if maximizing expected utility that assigns extra weight (by a factor of $1+\beta$) to the probability of the less favorable state. In our notation, the unnumbered equations near the top of Gul (1991, p. 678) say that the indifference curve segments have slope

$$-\frac{dy}{dx} = B\frac{\pi_X}{\pi_Y}\frac{u'(x)}{u'(y)} > 0, \tag{8}$$

where $B = (1+\beta)$ if $x < y$ (so X is the less favorable state) and $B = (1+\beta)^{-1}$ if $x > y$ (so Y is less favorable). Thus the indifference curve has a kink on the diagonal $x = y$, with -slope $\frac{\pi_X}{\pi_Y}(1+\beta)$ on the right and -slope $\frac{\pi_X}{\pi_Y}(1+\beta)^{-1}$ on the left.

Suppose, as is common in the subsequent literature, that the underlying Bernoulli function u is CRRA with risk aversion coefficient $\gamma > 0$. The tangency condition $-\frac{dy}{dx} = \frac{p_x}{p_y}$ applies as usual when the optimal choice is interior (not on the diagonal nor at a corner of the budget set). Writing $b = \ln(1+\beta)$, and recalling that in this case $\frac{u'(x)}{u'(y)} = (\frac{x}{y})^{-\gamma}$, we see that disappointment aversion changes equation (7) to

$$\ln\frac{x}{y} = \frac{1}{\gamma}[L - b] \tag{9}$$

when $L > b$ and so $x > y$. By symmetry, when $L < -b$, we again have a tangency but with $x < y$ and with $-b$ replaced by $+b$ in (9). For values of $L \in [-b, b]$, a DA agent will choose at the kink in the indifference curve where the diagonal $x = y$ intersects the budget line. Of course, for extreme values of L and sufficiently small parameters b and γ, corner solutions (on the x or y axis) are also possible. Estimating equations (13) below will spell this out explicitly.

Prospect theory (Kahneman and Tversky, 1979) is another relevant generalization of EUH. It has two main elements. The first is the Value function V, whose shape may change on either side of a reference point. The natural reference point in our context is zero. In that case, V is equivalent to a Bernoulli function. The second element is a probability weighting function $w(\pi_X)$, whose "inverse-S shape" is intended to capture "diminishing sensitivity" to changes in probability over the middle range of probabilities (e.g., Kahneman and Tversky, 1979, Tversky and Kahneman, 1992, Camerer and Ho, 1994). Our experiment will focus on probabilities between $\pi_X = 0.3$ and 0.8, for which $\left|\frac{w(\pi_X)}{w(\pi_Y)} - 1\right| < \left|\frac{\pi_X}{\pi_Y} - 1\right|$; that is, Prospect Theory implies the diminished senstitivity of individuals to changes in the relative probabilities of the two states. Thus, if Prospect Theory's weighted probabilities replace the objective probabilities in the tangency (and corner) conditions developed earlier, then EUH implication 3 above must be modified to state that the coefficient on **objective** log odds will have smaller magnitude than the coefficient on log price ratio.

3.3 Non-parametric summary statistic

We have seen that we can recover an estimate of a subject's coefficient of relative risk aversion from her responses to a budget line elicitation task. If her Bernoulli function is approximately CRRA we can use equation (7), and if her choices are approximated by the DA model we can use (9).

For more general preferences over lotteries (and even for heuristics that are not consistent with a preference relation), the estimated γ can still be regarded as an indicator of a subject's risk preferences, but it no longer has a precise interpretation. An agnostic researcher may prefer some sort of nonparametric indicator, but we are not aware of any such statistic that is generally accepted and is defined and comparable across a broad class of elicitation tasks. We considered several possibilities and eventually settled on a normalized risk premium, defined as follows.

Let $M = \max_{(x,y)\in F} \pi_X x + \pi_Y y$ be the maximum feasible expected payoff in an elicitation task. When $L \neq 0$, there is a unique point (x_M, y_M) that achieves that maximum and would be selected by a risk neutral agent. As usual, define $\mu_M = \pi_X x_M + \pi_Y y_M$ and $\sigma_M^2 = \pi_X(x_M - \mu_M)^2 + \pi_Y(y_M - \mu_M)^2$; note that $\sigma_M > 0$ in all our elicitation tasks. Let $C = \pi_X x_C + \pi_Y y_C$ be the expected payoff of the subject's actual choice $(x_C, y_C) \in F$. Then the revealed Relative Risk Premium is

$$RRP = \frac{M - C}{\sigma_M} \tag{10}$$

if $L \neq 0$ and otherwise is 0. Thus RRP resembles a coefficient of variation or a Sharpe ratio, and captures the agent's willingness to forego expected payoff in order to reduce dispersion.

3.4 Stochastic Dominance

Suppose that $\pi_X = \pi_Y = 0.5$ and $p_x = 0.4$ while $p_y = 0.6$. No matter what her risk preferences, an agent facing these prices and probabilities should never choose a point on the budget line with $x < y$. For example, suppose she considered choosing $(x, y) = (7.5, 15)$, exhausting her budget $m = 12$. Since the states are equally likely, she'd be just as happy with $(15, 7.5)$, no matter what her Bernoulli function is. But the portfolio $(15, 7.5)$ costs only 10.5, so she could afford to spend 1.5 more on either Arrow security (or both) and be strictly better off than at $(x, y) = (7.5, 15)$.

The general result is expressed in terms of first order stochastic dominance (FOSD). Recall that lottery A (strictly) FOSDs lottery B iff $F_A(x) \leq F_B(x)$ for all x, with strict inequality for some x. The definition refers to the cumulative distribution function $F_Z(x)$, the probability that the realized payoff in lottery Z is no greater than x. Recall also (e.g., Mas-Colell, Whinston, and Green (1995), p. 195) that every expected utility maximizing agent prefers lottery A to B iff A FOSDs B.

Proposition 1 *A choice (x, y) on the budget line (2) is strictly first order stochastically dominated by another choice on the same budget line iff*

a. one Arrow state (e.g., X) is more likely and its security is less expensive (e.g., $\pi_X \geq \pi_Y$ and $p_x \leq p_y$), with at least one of these comparisons strict; and

b. the choice includes strictly less of the less-expensive-more-likely security (e.g., $x < y$).

See Appendix A for a proof. The Proposition tells us that every choice on the budget line can be rationalized by some Bernoulli function if the more likely state has a higher price, or if $L = 0$. But some choices will be dominated when prices are equal and probabilities differ, or the reverse, and when the more likely state has a lower price.[2] In those cases, we can test for the rationality of subjects without committing to a functional form. For example, in Figure 1 below, the budget line crosses the diagonal at (400, 400)/9; any choice on the budget line with $x > 400/9$ is strictly dominated by an interval of choices with $x < 400/9$.

4 Laboratory Procedures

We first present the six general elicitation tasks used in our experiment. Then we lay out our within-subjects design and implementation details.

4.1 Elicitation Tasks

Budget line (BL). One task is to choose from a simple budget line in the tradition of Choi et al. (2007), as in Figure 1. The subject sees the state probabilities and (x, y) coordinates in text, while the state prices and the cash endowment are implicit in the slope and intercepts of the displayed

[2]The proof readily extends to cover many extensions of expected utility, including prospect theory with symmetric probability weighting and also disappointment aversion.

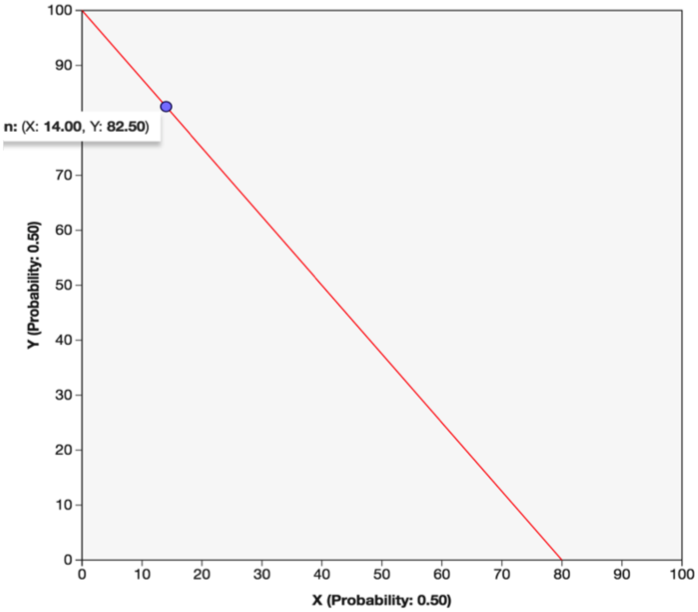

Figure 1: In treatment BL, the subject chooses an allocation of Arrow securities by clicking any point on a given budget line, then clicking Confirm bar (not shown). Text box shows values (x, y) at clicked point, here $(14, 82.5)$. Axis labels note π_X and π_Y; here, each is 0.5.

budget line. Our task varies the price ratio from 0.23 to 1.23, and varies the X state probability from 0.3 to 0.8.

Budget Jars (BJ and BJn). The interface for a new elicitation task is shown in Figure 2. The task has precisely the same feasible set as the BL task but it uses a much different graphical display. Subjects start with an explicit cash endowment (shown in green in the wide jar) and use sliders on the other two jars to buy the two Arrow securities. The level in the cash jar decreases (resp. increases) as the subject moves up (resp. down) the level in the red (security X) or blue (security Y) jar, at a rate proportional to the price of that security. The text below the jars spells out the state contingent payoffs (and state probabilities) at the current allocation. The subject clicks Submit bar to finalize the current allocation.[3]

The Submit bar is grayed out (not clickable) until the cash jar is empty in treatment BJn. That is, in this variant, cash must be exhausted. Note, however, that the set F of feasible allocations (x, y) shown in the next to last column ("Total") is the same as in treatment BJ (where cash balance can remain positive), and also the same as in treatment BL, for given state probabilities, prices and

[3]Proposition 1 implies that any choice involving a positive amount of a strictly more expensive security is first order stochastically dominated when state probabilities are equal. Suppressing the more expensive jar to prevent such choices would render the BJ task equivalent, from a decision theoretic perspective, to the Investment Game of Gneezy and Potters (1997). With their parameters — endowed cash is 4.0 and intercept (wolog, with the x-axis) is 10 — the price ratio is $4/(10 - 4) \approx 0.667$.

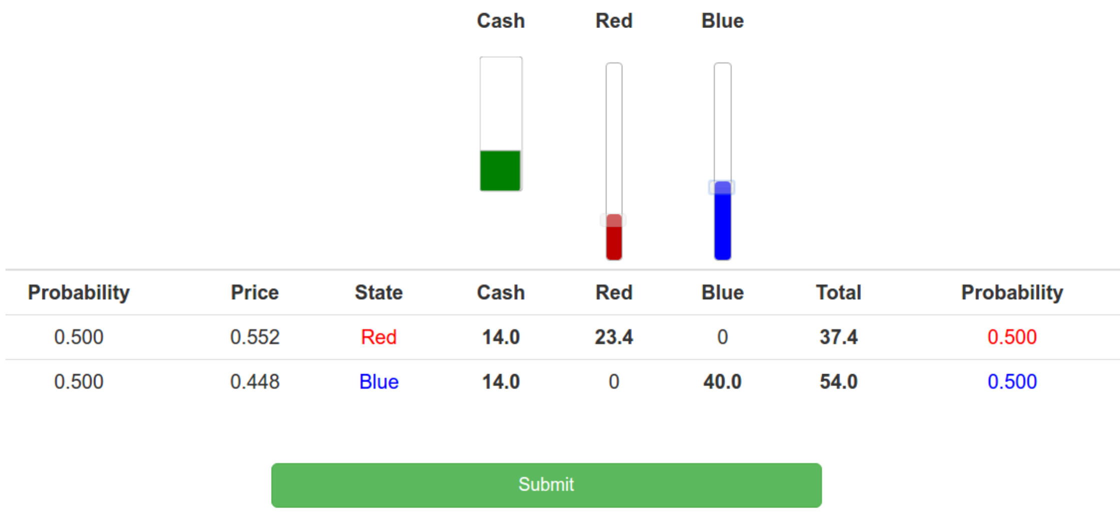

Figure 2: In treatment BJ, subjects choose an affordable allocation (x, y) by moving the sliders on the red and blue jars. The text below automatically updates so that x is shown in the "Total" column in the Red row, and y is shown below it in the Blue row. Clicking the Submit bar finalizes the allocation. In treatment BJn, the subject must empty the cash jar before the Submit bar becomes active.

cash income.

Budget Dots: Eckel and Grossman (BDEG). Eckel and Grossman (2002, 2008) ask subjects to chose an allocation (x, y) from a menu F of five or six possibilities, with equal state probabilities $\pi_X = \pi_Y = 0.5$. We modify their task by also considering unequal probabilities and by displaying F as discrete points in a graph otherwise similar to Figure 1. The Eckel-Grossman menus typically graph as in Figure 3a: equally spaced points on a budget line spanned by the more distant intercept (for the cheaper security) and the perfectly hedged portfolio ($x = y$); the stochastically dominated points are excluded.

Multiple price list (HL, BDHL). Perhaps the most widely used elicitation task in recent years is the multiple price list in text format (e.g., Holt and Laury, 2002). Each row in the text list has the same two allocations but different rows have different state probabilities. Our HL treatments use Holt & Laury's original pair of lotteries — $(x, y) = (2.00, 1.60)$, called "safe," and $(x, y) = (3.85, 0.10)$, called "risky" — and the six most relevant state probabilities, $\pi_X = 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$.[4] Treatment HL stacks six rows of text, each row representing choice be-

[4]The original list also included lines for $\pi_X = 0.1, 0.2, 0.9, 1.0$ but the vast majority of subjects choose the safe lottery for the first two omitted probabilities and choose the risky lottery for the last two. To streamline our design, we omit these four items; see Habib, Friedman, Crockett, and James (2017) for insight into the impact of such omissions.

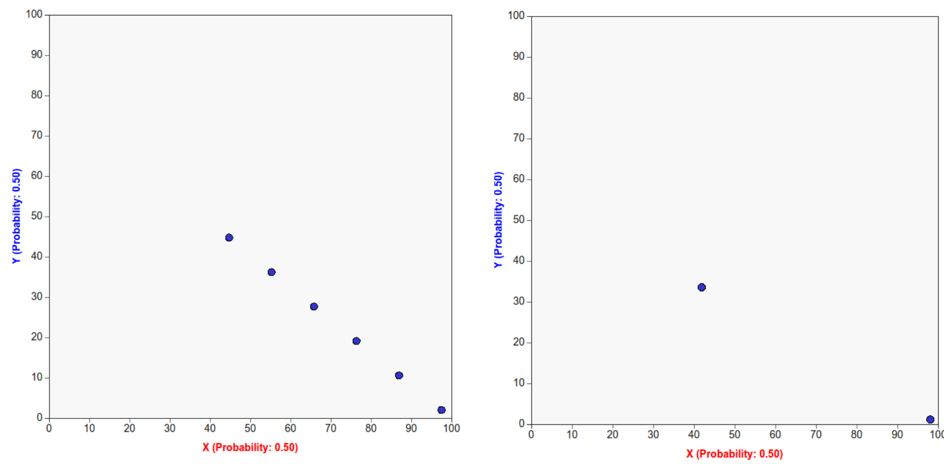

Figure 3: Discrete budget dots. Axis labels note π_X and π_Y; here, each is 0.5. (a) In treatment BDEG, the subject chooses an allocation of Arrow securities by clicking one of the six large dots on the given budget line, then clicking Confirm bar. (b) In treatment BDHL, subjects click one of two dots representing the two feasible HL allocations.

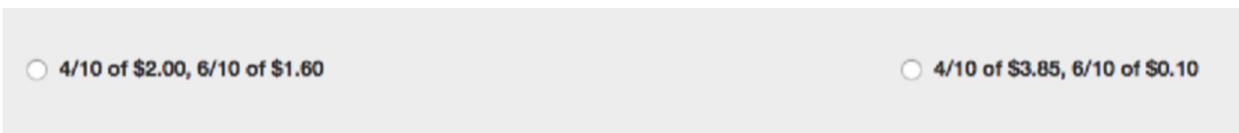

Figure 4: In treatment HL, subjects click a radio button to choose between two feasible allocations in each line; the X state probability (here 0.40) increases by 0.10 from one line to the next.

tween two lotteries as in Figure 4, with $\pi_X = 0.3$ in the top line increasing by 0.1 in each successive line and $\pi_Y = 1 - \pi_X$. Treatment BDHL takes the lotteries from one row (i.e., with a particular π_X value) from HL and displays the two feasible choices graphically, as in Figure 3(b), where the implicit price ratio is $p = -\frac{\Delta y}{\Delta x} = \frac{1.60-0.10}{3.85-2.00} \approx 0.81$. As further described below, successive trials vary the probabilities while keeping the price constant, and some sets of trials use an implicit price of .58 instead of the original 0.81.

4.2 Experimental Design

Each subject faces a total of 66 lottery choices, organized into 11 blocks. The first and last block for each subject is the six-line HL elicitation task with $p = \frac{p_x}{p_y} = 0.81$; unless otherwise noted, the analysis below treats each of these blocks as a single trial. The other blocks consist of six consecutive trials, each with a single lottery choice. The middle block (block 6 of 11) for each subject is always the BL task with $\pi = 0.5$ and price sequence $p = 0.23, 0.58, 0.81, 0.93, 1.00, 1.23$. The remaining blocks (2-5 and 7-10) use other elicitation tasks, but within each block the task is held constant.

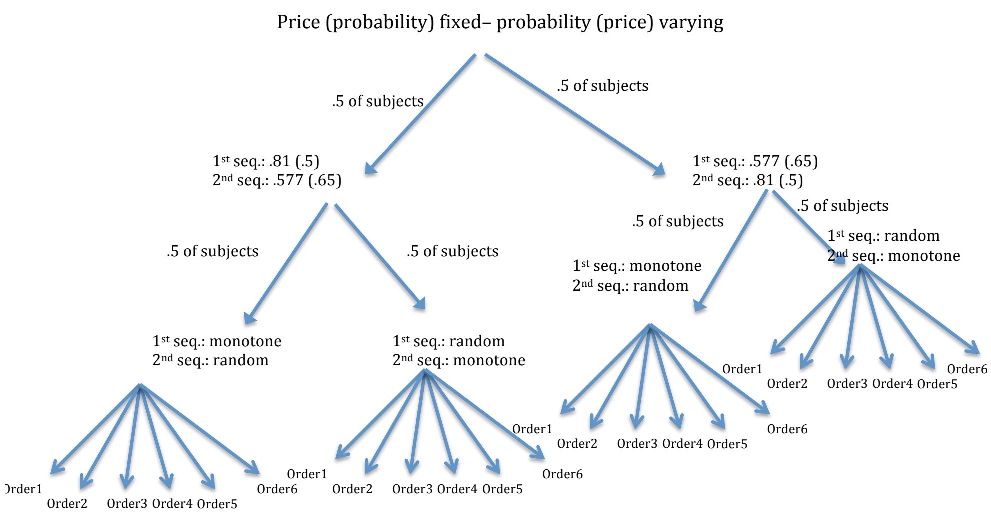

Figure 5: Experimental Design Summary. Sessions are balanced across several design dimensions: fixed price vs fixed probability, order of the two levels at which price or probability is fixed, monotone vs random sequencing of varying price (or probability), and treatment order across blocks.

We place equal numbers of subjects in two sorts of sessions: fixed price and fixed probability, according to the design of blocks 2-5 and 7-10. As summarized in Figure 5, half of the fixed price sessions keep price at .81 in blocks 2-5, and keep it fixed at .58 in blocks 7-10, while the two fixed prices are flipped in the other half of these sessions. Within each block of these fixed price sessions, each of the six probabilities (from $\pi_X = 0.3$ to 0.8) is used once. In half of these sessions a monotone increasing sequence of probabilities is used in blocks 2-5, while random sequences are used in blocks 7-10. In the other sessions random sequences are used in blocks 2-5 and the monotone sequence is used in blocks 7-10. Each fixed price session uses the elicitation tasks BL, BDHL, BJ, BJn once each in the first four blocks and once each in the last four blocks. Of the $4! = 24$ possible task sequences that could be employed before (or after) the middle block (6), we selected a balanced subset denoted Order1 thru Order6, and used them with equal frequency.

In the fixed probability sessions, we similarly keep π_X fixed at 0.5 in blocks 2-5 and at .65 in blocks 7-10, or the reverse. Within each of these blocks we present the six varying prices $p = \frac{p_x}{p_y} = 0.23, 0.58, 0.81, 0.93, 1.00, 1.23$ in monotone increasing sequence or in random sequence, and use Order1 through Order6 for the elicitation tasks across the monotone blocks. The set of four tasks in these sessions is the same as in the fixed price sessions except that BDEG (which requires fixed probabilities) replaces BDHL (which requires fixed prices).

4.3 Implementation

A total of 142 subjects from the LEEPS lab subject pool participated in 18 sessions between October 2016 and March 2017. After subjects read instructions (a copy is attached as Appendix C) privately the conductor explained the mechanics of each elicitation institution and allowed subjects to make practice decisions prior to the paid trials.

Each subject was paid for a single trial, determined by a ball drawn from a bingo cage with 56 numbered balls. If ball 1 or ball 56 came up, indicating a HL trial, then a roll of a six sided die determined the relevant line. The subject then rolled a ten-sided die to determine which state (X or Y) of the chosen lottery paid that period. Each session lasted about 60 minutes, and the final payments $[min, max]$ range, including \$7 show-up fee, was \$[7.00, 17.00], with average payout roughly \$10.

5 Results

To provide perspective on subsequent data analysis, we begin with the four scatterplots in Figure 6, each showing all choices in Blocks 2-10 for a single subject. The log allocation ratios for each choice are plotted against the log price - log odds composite variable L; open squares are used for the budget jar task (or open circles when there is a no-cash constraint), triangles for budget line task, and plusses for the relevant budget dot task. For reference, the solid line graphs what equation (7) predicts for a expected utility maximizer with the original Bernoulli function $u(c) = \ln(c)$. The dashed line plots risk-neutral optimal choices, while the dotted line plots the choices of an otherwise risk-neutral person with an interval of disappointment aversion.

The subject in Panel (a) could be characterized as a noisy CRRA expected utility maximizer with γ close to 1. Panel (b) shows another subject in a fixed probability session who, in 53 of 54 trials, maximized expected value (as in CRRA with $\gamma = 0$). Panels (c) and (d) show two different subjects in fixed price sessions who seem like noisy disappointment averters in the budget dot task (here BDHL), but who diverge in the other tasks (towards $\gamma = 1$ or $\gamma = 0$, respectively). These four examples exhibit only a small slice of our 142 subjects' diverse behavior.

Stochastically dominated choices in Figure 6 appear in the second or fourth quadrant, where L and $\ln x/y$ have opposite signs.[5] We see seven choices in those quadrants in panel (c) of the Figure,

[5]The converse is false: not all choices that plot in those quadrants are stochastically dominated. For example, a choice with $y > x$ is not a FOSD violation when $\ln \frac{p_y}{p_x} > \ln \frac{\pi_Y}{\pi_X} > 0$ even though that choice plots in the fourth quadrant. On the other hand, choices consistent with CRRA, or with most other common parametric Bernoulli functions, never plot in the second or fourth quadrant.

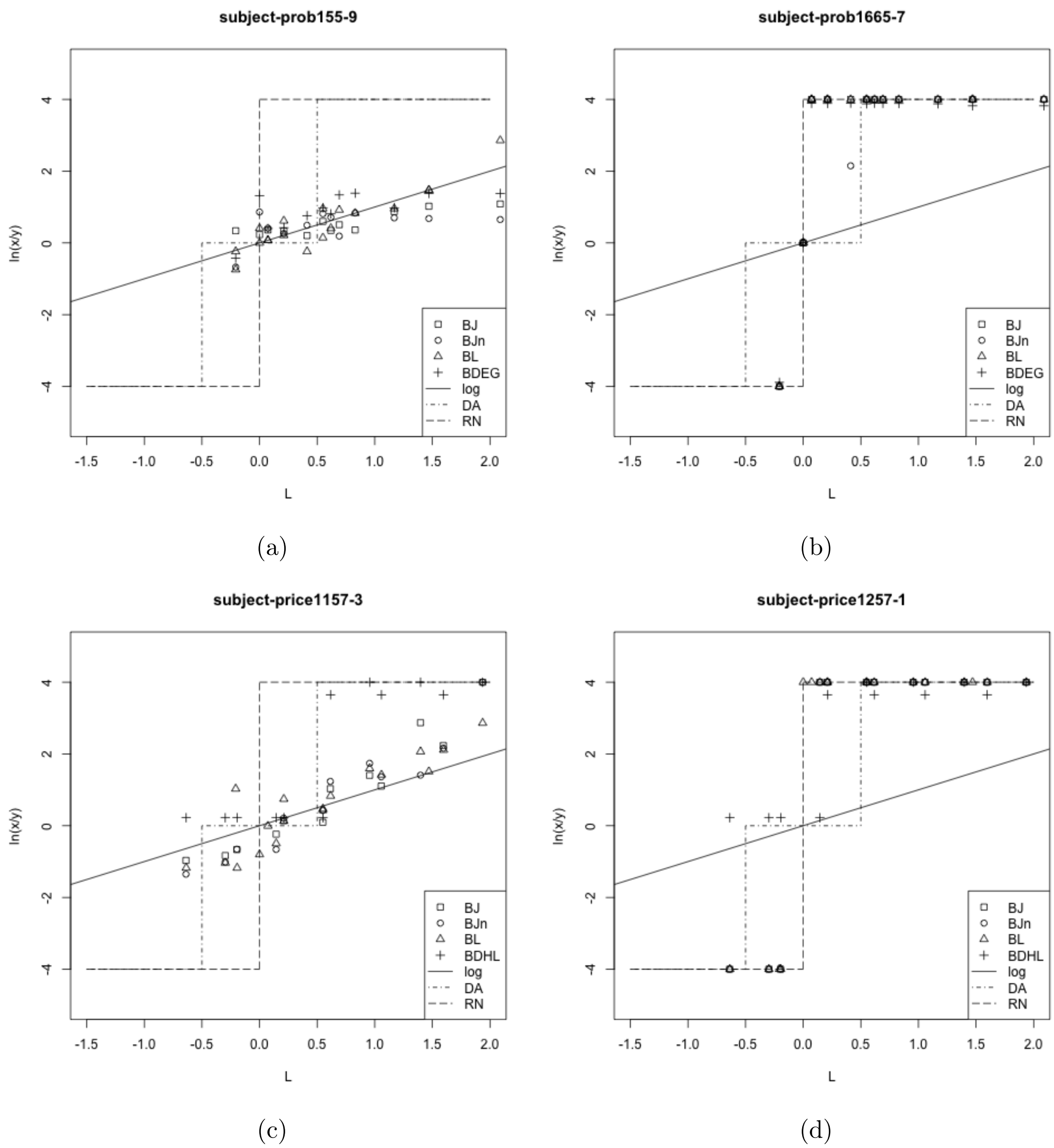

Figure 6: Choices (truncated at $\ln x/y = \pm 4.0$) for four subjects. Actual choices by elicitation task are plotted with open symbols, and theoretical predictions are plotted with lines, sorted according to $L = \ln \pi_X - \ln \pi_Y - \ln p_x + \ln p_y$. Log refers to CRRA model with $\gamma = 1$, RN refers to risk neutral choice ($\gamma = 0$), and DA refers to the disappointment averse model with $b = 0.5$ and $\gamma = 0$.

but rather few in the other panels; in most such cases, we are close to the quadrant boundaries.

Table 1 takes a more systematic look at violations of first order stochastic dominance (FOSD). The feasible set F contains no dominated choices in the BDEG and BDHL tasks, so they are omitted from the Table. The same is true for HL if we consider each row as a separate task, but multicrossings in 6-line HL trials (i.e., in the first and last trials) are FOSD violations, and these appear in the Table's last column. The other columns report FOSD violations in the remaining tasks, where Proposition 1 applies.

	BL	BJ	BJn	HL
Opportunities	2196	1438	1399	284
Violations	309	424	508	27
(Random)	872	571	558	257
Major Violations	16	9	19	-
(Random)	258	209	201	-

Table 1: Violations of FOSD. "Opportunites" is the number of trials involving a task that allowed FOSD violations. "Violations" are the number of trials in which subjects' actual choices violated FOSD. "Random" gives the expected number (to nearest integer) of violations given iid uniformly distributed random choices in each task. A violation (x, y) at L is deemed "major" if $L \times ln(\frac{x}{y}) \leq -1$.

We distinguish between minor violations (as illustrated in the previous Figure) and "major" violations, defining the latter as lying inside the rectangular hyperbola $ln(\frac{x}{y})L = -1$. Table 1 shows a fair number of minor violations of FOSD but (by this criterion, chosen a priori), relatively few major violations. Table B.5 in Appendix B looks at tighter criteria for major violations, and confirms that a large majority of actual violations are tiny, due to clicking just a few pixels away from a sensible choice. Subsequent analysis therefore will include all data including FOSD violations.

5.1 Prices and probabilities

Recall that implications 2 and 3 of standard decision theory predict that people will treat the composite variable $L = ln(\pi_X) - ln(\pi_Y) - \ln(P_x) + ln(P_y)$ as a sufficient statistic for prices and probabilities, and that they will react symmetrically to prices and to probabilities. Do actual choices support those strong predictions?

To find out, we run regressions on pooled data of all 142 subjects. The dependent variable is $\ln x/y$, the log portfolio ratio of the actual choice, trial by trial. Of course, that variable is not defined at corner choices, where $\ln x/y = \pm\infty$. Therefore, the regressions reported below truncate at $\ln x/y = \pm 4.0$, i.e., we lump together all observations within $\exp(-4) \approx 2\%$ of a corner. The key regression is

$$\ln(\frac{x}{y}) = \alpha_{lo}\ln(\frac{\pi_X}{\pi_Y}) + \alpha_{lp}\ln(\frac{P_x}{P_y}) + \varepsilon, \tag{11}$$

in which symmetry is imposed via the parameter restriction

$$\alpha_{lo} + \alpha_{lp} = 0. \tag{12}$$

	Dependent variable:ln(x/y)			
	(1)	(2)	(3)	(4)
logOdds	1.288***	1.320***	1.390***	1.473***
	(0.080)	(0.077)	(0.070)	(0.056)
logPrice	−1.101***	−1.130***	−1.233***	−1.334***
	(0.082)	(0.079)	(0.078)	(0.079)
Observations	5,964	6,804	7,668	9,372
R^2	0.430	0.441	0.460	0.477
F-stat	16.16***	18.78***	13.48***	12.02***
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2: Estimates of Equation (11), with errors clustered at subject level. Column 1 data includes only BL, BJ and BJn trials. Column 2 also includes BDEG trials, and column 3 includes BDHL trials as well. Column 4 also includes the HL list data, with each line treated as a separate trial. F-stats are for the parameter restriction (12).

Table 2 presents the results for different subsets of the data. The F-tests strongly reject the null hypothesis of symmetry in favor of the one-sided alternative that subjects respond more strongly to probabilities than to prices. Note that this is the opposite direction from the "diminishing sensitivity" prediction of Prospect Theory.

Still, the estimated probability (or log odds) responsiveness of about 1.3 or 1.4 is not dramatically different than the price responsiveness of 1.1 to 1.3. Thus the theoretical prediction that L is a sufficient statistic for the state prices and state probabilities misses a significant and unexpected regularity, but is nevertheless a decent first approximation. Subsequent analysis will therefore often use L to summarize the prices and probabilities.

5.2 Disappointment Aversion

How important is the tendency (first noted in Choi et al., 2007) for subjects to pick almost perfectly hedged portfolios, even when L is not very close to zero? To gain some insight, we estimate the two parameter model introduced in Section 3.2. Recall that we seek to estimate the coefficient of relative risk aversion γ as well as the disappointment aversion parameter $b = \ln(1+\beta)$, using L as an explanatory variable for the log choice ratio $\ln x/y$.

The model says that a subject will choose a point on the diagonal $x = y$ (where $\ln x/y = 0$) when L falls within a certain range defined by b. Applying the logic of Section 3.2, we obtain the estimating equations

$$\begin{aligned} \ln x/y &= 0 \text{ if } L \in (-b, b) \\ &= \frac{1}{\gamma}[L - b] \text{ if } L \in [b, 4\gamma + b] \\ &= \frac{1}{\gamma}[L + b] \text{ if } L \in [-4\gamma - b, -b] \\ &= 4 \text{ if } L > [4\gamma + b] \\ &= -4 \text{ if } L < -[4\gamma - b] \end{aligned} \tag{13}$$

and use NLLS. That is, for a given subject (label suppressed), let $R(L(\pi, p), b, \gamma)$ be the right-hand side of (13). Then

$$(\hat{b}, \hat{\gamma}) = \text{argmin} \sum_{t=1}^{54} [\tilde{\ln}\ x_t/y_t - R(L_t, b, \gamma)]^2. \tag{14}$$

The tilde in $\tilde{\ln}\ x_t/y_t$ is to remind us that that dependent variable is truncated at ± 4.

Figure 7 compares the predictive power of that two parameter model to that of a simple CRRA model (as in equation (7) or, equivalently, imposing the restriction $b = 0$ in equations (13 - 14)). For each subject, we designate one of the observations as the prediction target, estimate both models on the remaining 53 of the 54 observations, predict the target observation and compute the prediction error for each model. The table reports squared prediction errors summed over all 54 possible prediction targets.

The Figure shows that the majority of our 142 subjects have relatively small prediction errors (SSE $<$ 150) that hardly differ between the two models and so fall almost on top of the diagonal line. The largest outlier has a larger DA error (about 500) than simple CRRA error (about 350). The DA model does slightly better than the CRRA model with the subjects with largest SSE's, but (by definition) neither model predicts their behavior very well. See Appendix B for more analysis of

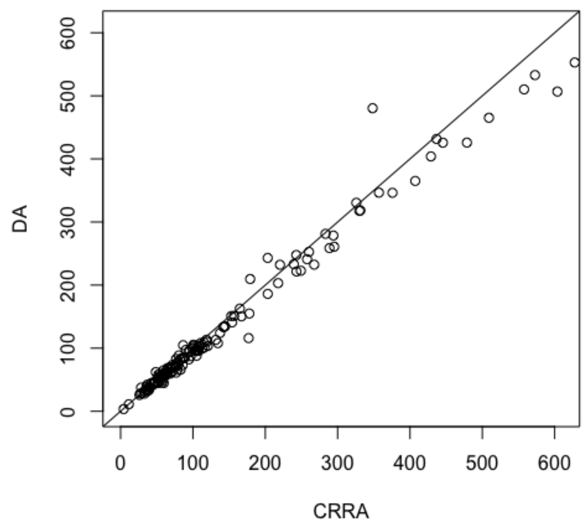

Figure 7: Predictive power comparison. Vertical axis is sum of squared prediction errors (SSE) for DA model (13 - 14), and horizontal axis is SSE for simple CRRA model ($b = 0$).

the two parameter model and more model comparisons. We conclude that we don't sacrifice much predictive accuracy by using the simpler (one-parameter CRRA) model in the analysis to follow.

5.3 Distributions of estimated risk aversion

We now disaggregate by both subject and task. To compute the parametric risk aversion measure γ for tasks in the BL family, we return to our maintained hypothesis that prices and probabilities enter only via the composite variable L, but now include task-specific interactive dummy variables to detect differences across elicitation tasks. For each subject we estimate

$$\ln(x/y) = (\beta_1 + \beta_2 BJ + \beta_3 BJn)\, L + \varepsilon, \tag{15}$$

for the relevant seven blocks, which contain 42 observations: 18 from BL, and 12 each from BJ and BJn. Revealed risk aversion then is computed from the resulting coefficient estimates via

$$\hat{\gamma}_{i\tau} = 1/(\hat{\beta}_1 + \hat{\beta}_k), \tag{16}$$

where $k = 2$ for treatment $\tau =$ BJ and $k = 3$ for $\tau =$ BJn. Of course, for the omitted treatment $\tau =$ BL, the revealed value is just $1/\hat{\beta}_1$.

We also obtain γ estimates from the six-line HL trials using the traditional crossover method, as

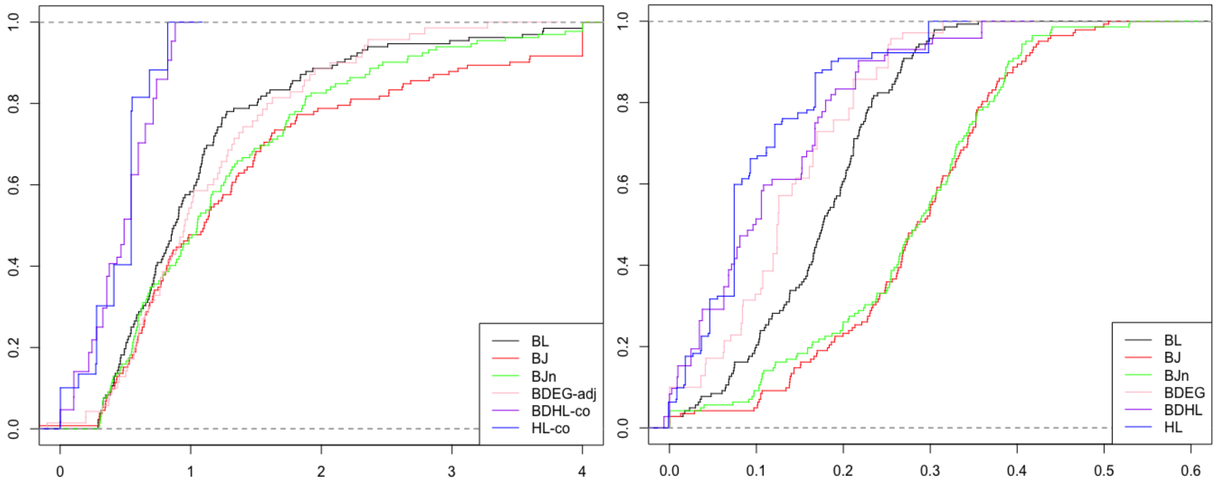

Figure 8: (a) Cumulative distribution functions of $\hat{\gamma}_{i\tau}$ for all subjects. HL-co, BDHL-co, and BDEG-adj estimates use standard crossover method, and other estimates use equations (15, 16). (b) Cumulative distribution functions of RRP for all subjects.

follows. Let the subject cross from the safe allocation $(2.00, 1.60)$ to the risky allocation $(3.85, 0.10)$ between rows k and $k+1$, and let γ_j be the parameter value that would make a CRRA EU-maximizing agent indifferent between the two allocations in row j. Then $\hat{\gamma} = 0.5(\gamma_k + \gamma_{k+1})$ for that trial. For each subject we record $\hat{\gamma}_{HL-co}$, the mean $\hat{\gamma}$ of the first and last trials. We also compute $\hat{\gamma} = 0.5(\gamma_k + \gamma_{k+1})$ for each block of six BDHL trials the subject encountered in fixed price sessions, and record the average $\hat{\gamma}_{BDHL-co}$ over the two BDHL blocks for that subject. HL-co (or BDHL-co) is not defined for subjects who never cross or who multicross in either HL trial (or in any BDHL block).

The γ estimate for a single trial of BDEG of a fixed probability session is, as customary, the average of two γ values: one for which a CRRA agent would be indifferent between the chosen dot and the one above it, and one for which the indifference γ is between the chosen dot and the one below. (For trials where the chosen dot is extreme, so there is either no dot above or else no dot below, we simply use the γ algebraically implied by equation (7). For each subject, BDEG-adj is the average over the 12 BDEG trials (two blocks of 6).

Panel (a) of Figure 8 collects the results for subjects in all sessions. The lowest γ estimates come from the HL and BDHL tasks; both are approximately uniformly distributed between 0 and 0.8. At the other extreme, the BJ and BJn regression estimates have medians around 1.0; their distributions are relatively dispersed and skewed towards higher values. The no-cash constraint seems to reduce the fraction of outlying estimates ($\hat{\gamma} > 4.0$) from about 10% to under 4%. The regression estimates from BL and BDEG produce middling distributions concentrated in the range [0.5, 3.0], with greater density towards the lower end.

Panel (b) of Figure 8 also shows distributions across all subjects of estimates for each relevant task, but it uses the non-parametric risk preference measure RRP. Recall from equation (10) that, for BL-like tasks (including BJ and BJn as well as BDEG) the RRP measure is defined as the shortfall in expected value (of the risk-neutral choice minus that of the actual choice) normalized by the standard deviation of the risk-neutral choice. We treat each six-line HL trial as a compound lottery in computing RRP, i.e., C in equation (10) is the expected value of $1/6$ chance of playing the chosen (safe or risky) lottery for each of the 6 lines, and M and σ_M are similarly calculated for the six risk-neutral choices. We treat the six trials in a BDHL block in the same manner.

In Panel b, as in Panel a, the BJ and BJn distributions of RRP are close to each other, and indicate that subjects respond to these tasks as if considerably more risk averse than in the HL and BDHL tasks. Again, the distributions for the BL and BDEG tasks lie in between, but now are closer to the HL distributions.

	BDHL-co	BDEG-adj	BL	BJ	BJn
HL-co	0.01	0.00	0.00	0.00	0.00
BDHL-co	-	0.00	0.00	0.00	0.00
BDEG-adj	-	-	0.49	0.28	0.66
BL	-	-	-	0.01	0.05
BJ	-	-	-	-	0.75
HL-co	0.05	0.00	0.00	0.00	0.00
BDHL-co	-	0.03	0.00	0.00	0.00
BDEG-adj	-	-	0.00	0.00	0.00
BL	-	-	-	0.00	0.00
BJ	-	-	-	-	0.87

Table 3: Kolmogorov-Smirnov p-values for equality of γ distributions (top panel) and RRP (lower panel). Tests consider all choices by all subjects, who each faced five of the six tasks; subjects in fixed price (resp. fixed probability) sessions did not face the BDEG (resp. BDHL) task.

Table 3 reports Kolmogorov-Smirnov test results, and firmly rejects the null hypothesis that most pairs of tasks have the same distribution of risk aversion parameters γ. Confirming visual impressions from Figure 8, the exceptions are that the distribution of BDEG is similar to that for other tasks in the BL family, and that BJ and BJn have rather similar distributions. KS tests on the RRP distributions rejects the null hypothesis for all pairs except for BJ and BJn. Overall, we see that different tasks generally lead to substantially different risk estimates.

Is the same true for tasks that decision theory would deem identical? Table 4 compares individual subjects' γ estimates in the first round HL trial (which implements exactly the traditional HL method) with their γ estimates in later HL trials: the (identically implemented) last round HL trial, and BDHL blocks under different price treatments. The table shows that re-test stability

		$\frac{P_x}{P_y} = .81$		$\frac{P_x}{P_y} = .57$	
	Last HL-co	Monotone	Random	Monotone	Random
t test	-0.02	0.02	0.14	0.03	-0.17
	(0.45)	(0.74)	(0.04)	(0.64)	(0.01)
df	120	25	20	19	22
KS test	0.05	0.47	0.42	0.34	0.49
	(0.99)	(0.00)	(0.02)	(0.11)	(0.00)
corr	0.63	0.38	.42	0.33	0.11

Table 4: HL consistency tests (and p-values). Matched pair t-tests for the null hypothesis that, for each subject, γ calculated from the first HL trial (via the standard crossover method) minus γ calculated from specified later trials is zero. Kolmogorov-Smirnov (KS) test compares the cumulative distributions, and corr reports the Spearman rank correlation coefficients, across the same subsets of trials. All subjects (except for those who multi-cross) are included in the first column; remaining columns are restricted to the 72 fixed price subjects, of whom 43 saw BDHL blocked at price .81 and BDHL random at price .58 while the remaining 29 subjects saw the reverse. Subjects who never cross were assigned $\gamma = 0$ (resp. $\gamma = 4$) if they always chose the risky (resp. safe) gamble.

of subject behavior over earlier and later administrations of HL cannot be rejected in mean or in distribution. The same table also shows, however, that divergences emerge in comparisons to between the first administration of HL and administrations of BDHL. BDHL implementing the same implicit price ratio, $P = 0.81$, as HL, but in random order, supports significantly less risk averse behavior. BDHL implementing a $P = 0.57$ price ratio, in random order, supports significantly more risk averse behavior; this finding is directionally consistent with the results on the effect of payoff transforms shown by Habib et al. (2017), where a payoff transform equivalent to a $P = 0.23$ price ratio in an MPL supported substantially more risk averse behavior than did $P = 0.81$. We will present further results on the effects of visual format and of environment sequencing in Section 5.5.

5.4 Within subject (in)consistency

The tests so far look at population distributions and so do not address subject-level consistency. As stated near the beginning of Section 3, the expected utility hypothesis is that a human subject not only chooses as if maximizing some Bernoulli function, but also that her Bernoulli function is fixed. Such preference stability is crucial from a scientific perspective — the whole point of using some artificial task to elicit risk preferences is that those revealed preferences should enable the researcher to predict behavior in other risky settings of more direct economic interest. As a step towards checking preference stability or consistency, we now examine the power of individual subject's choices in one elicitation task to predict behavior in other elicitation tasks.

Given the substantial differences already observed in the population distributions, the main concern now is the relative ranking among subjects. Does a subject who, say, reveals herself to be among the most (or least) risk averse subjects in one task tend in other elicitation tasks to reveal a relatively high (or low) degree of risk aversion? Specifically, are subjects' Spearman rank correlations across tasks near $\rho = 1.0$?

An affirmative answer seems especially plausible for tasks that basic decision theory considers identical. Since they employ precisely the same feasible set F = a budget line, the BL, BJ, and BJn tasks are the same according to basic decision theory, and they are closely related to the tasks, BDHL and BDEG, that use finite subsets of the budget line. By the same token, the initial and final period trials of HL are identical to each other and to $p = 0.81$ blocks of BDHL. Therefore decision theory (and, a fortiori, EUH) predicts identical rankings within these task families.

Fixed Price	BDHL	BL	BJ	BJn
HL	0.35	0.23	0.23	0.22
BDHL	-	0.51	0.46	0.54
BL	-	-	0.72	0.81
BJ	-	-	-	0.80
Fixed Prob	**BDEG**	**BL**	**BJ**	**BJn**
HL	0.15	-0.07	0.03	0.16
BDEG	-	0.46	0.56	0.53
BL	-	-	0.68	0.62
BJ	-	-	-	0.86

Table 5: Within-subject Spearman rank correlation of estimated γ.

Table 5 collects the rank correlations for the parametric measure γ of revealed risk preference. In the fixed price sessions we indeed get fairly high correlations within the BL family, in the 0.7 - 0.8 range. Correlation between HL and BDHL is .35, lower than one might expect given the within family consistency seen in Figure 8, and given the correlations around 0.5 of BDHL with members of the BL family. In the fixed probability sessions, the HL correlation with members of the BL family is close to zero; other correlations are roughly similar to those in the fixed price sample.

Table 6 tells a generally similar story for the non-parametric measure RRP. In the Fixed Price data, the within family correlation is $\rho = 0.50$ for the two HL members, and is roughly .5 to .8 within the three BL members, while cross family correlations are around .6 for BDHL and .3 - .4 for HL. Again, in the Fixed Probability data, the cross family correlations are near zero (with the possible exception of the HL-BDEG correlation of 0.21) while the four members of the BL family have within-family correlations ranging from a bit under 0.5 to almost 0.7.

Fixed Price	BDHL	BL	BJ	BJn
HL	0.50	0.41	0.33	0.37
BDHL	-	0.61	0.51	0.59
BL	-	-	0.71	0.82
BJ	-	-	-	0.80
Fixed Prob	BDEG	BL	BJ	BJn
HL	0.21	0.08	0.06	0.06
BDEG	-	0.61	0.57	0.44
BL	-	-	0.58	0.47
BJ	-	-	-	0.69

Table 6: Within-subject Spearman rank correlation of estimated RRP.

5.5 Sensitivity to task attributes

The previous two subsections show clearly that a decision theoretic focus is too narrow to explain many of the most interesting regularities in our data. Can some of the observed differences across elicitation tasks be explained instead by the way human subjects react to the presentation of the choices? More specifically, could some task attributes that are irrelevant according to decision theory nonetheless affect the central tendencies of revealed risk preferences?

	HL	BDHL	BDEG	BL	BJ	BJn
Spatial	-	✓	✓	✓	-	-
2Dots	✓	✓	-	-	-	-
6Dots	-	-	✓	-	-	-
Cash	-	-	-	-	✓	-
FixProb	-	-	✓	*	*	*
Random	-	*	*	*	*	*

Table 7: Task attributes. A ✓in the column for a given task indicates that the attributes in that row is always present, a − indicates an attribute never present, and a * indicates an attribute present in some but not all trials using the task.

Table 7 considers six such attributes for our six elicitation tasks, four of which pertain to the elicitation format and two of which pertain to the wider environment. The attribute "Spatial" refers a budget line display in Arrow-Debreu 2D space, either allowing choice anywhere on the line (in BL) or on a subset of points (BDHL and BDEG). "2Dots" refers to tasks with only binary choices, either via radio buttons selecting lotteries presented in text (HL) or via two points in Arrow-Debreu 2D space (BDHL). The attribute "6Dots" refers to the other discrete choice possibility on the budget line. "Cash" refers to the attribute (used only in treatment BJ) allowing retention of cash in the cash jar (as opposed to its explicit exhaustion in BJn, and implicit budget exhaustion in all other treatments). The environmental attributes are Fix[ed]Prob[ability] sessions (versus Fixed Price) and Random (versus monotone) ordering of price or probability sequences.

	γ		RRP	
	(1)	(2)	(3)	(4)
Spatial	-0.124^{*}	-0.124^{***}	-0.081^{***}	-0.081^{***}
	(0.073)	(0.025)	(0.006)	(0.017)
2Dots	-0.347^{***}	-0.347^{***}	-0.167^{***}	-0.167^{***}
	(0.053)	(0.064)	(0.009)	(0.035)
6Dots	0.468^{***}	0.468^{***}	-0.011	-0.011^{**}
	(0.076)	(0.027)	(0.007)	(0.005)
Cash	0.050	0.050^{***}	0.020^{***}	0.020
	(0.083)	(0.018)	(0.007)	(0.012)
FixProb	-0.019	-0.019	-0.087^{***}	-0.087^{***}
	(0.156)	(0.040)	(0.019)	(0.026)
Random	0.095^{***}	0.095^{*}	0.024^{***}	0.024^{***}
	(0.053)	(0.052)	(0.009)	(0.008)
period	0.003	0.003	-0.0003	-0.0003^{***}
	(0.001)	(0.002)	(0.0002)	(0.0001)
Spatial:FixProb	-0.140	-0.140^{***}	0.013	0.013
	(0.113)	(0.041)	(0.011)	(0.021)
Cash:FixProb	0.136	0.136^{***}	-0.006	-0.006
	(0.134)	(0.018)	(0.010)	(0.012)
FixProb:Random	-0.083^{***}	-0.083	-0.023	-0.023
	(0.105)	(0.111)	(0.020)	(0.015)
FixProb:period	-0.001	-0.001	0.001^{*}	0.001^{***}
	(0.003)	(0.003)	(0.001)	(0.0002)
Observations	6,752	6,752	6,815	6,815
R^2	0.246	0.246	0.653	0.653
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: OLS regression coefficients (and standard errors) for risk preference measures γ and RRP. Regressions include subject level fixed effects with errors clustered at the subject level (columns 1 and 3) and task level (columns 2 and 4).

Table 8 reports regressions of revealed risk aversion on attribute dummies. The first two columns report results for the γ measure and the other two columns for the RRP measure. One can see that there are economically significant and statistically significant effects beyond those recognized by decision theory. For example, the Spatial coefficient suggests that, consistent with the results in Habib et al. (2017), subjects responding to budget lines (or dots) drawn in a two-dimensional space reveal less risk aversion (lower gamma estimates) than subjects responding to text. The effect is substantial (≈ -0.12) and is compounded in the fixed probability environment (≈-.12 - .02 - .14 = -.28). The BDEG attribute of restricting choice to 6 dots on the same side of the perfect hedge increases the elicited γ by almost 0.5, while the discrete restriction in BDHL to 2 dots (one of which is near the corner of the budget line) reduces the elicited γ by about 0.35. Such differences are huge in terms of economic behavior: the difference between risk neutrality and highly risk averse "square root" utility (as typical of FPSB data estimated under CRRAM (Cox, Smith, and Walker, 1988) is, of course, 0.5.

Another intriguing result is that the Random sequence attribute seems to impact fixed price sessions but not fixed probability sessions. This suggests that the mechanism behind the sequence effect (and also the Lévy-Garboua et al. (2012) result on variations in implementation of Holt-Laury) may be that subjects rely on the monotone ordering of probabilities (as across rows of standard Holt-Laury) to guide their decision-making. There is, however, no such reliance in budget allocation it seems not to matter whether changes in the slope of a budget line are random or monotone. Whether cash retention is allowed or budget exhaustion is enforced (explicitly or implicitly) seems to matter more in the sessions where probability is fixed and price ratios (budget line slopes) vary. The impact of attributes such as discreteness (e.g. “6 Dot") may be less surprising (the choice set is restricted in BDEG relative to BL, for instance), but it is still worth documenting them, and of course controlling for them.

6 Discussion

We report a laboratory experiment in which each of our 142 subjects responds to a variety of direct choice lottery tasks, each of which is intended to elicit personal risk preferences. For each task and for each subject, we summarize the elicited preferences parametrically, via the coefficient γ of relative risk aversion, and also non-parametrically, via a relative risk premium (RRP). Our main findings can be summarized briefly.

First, the expected utility hypothesis (EUH) — that each subject chooses among lotteries as if maximizing the expectation of some personal Bernoulli function — is a decent approximation for each elicitation task considered separately. Our subjects often violate first order stochastic dominance restrictions, but the vast majority of violations are tiny. (Biais, Mariotti, Moinas, and Pouget (2017) reach a similar conclusion in independent recent work.) Also contrary to EUH, some subjects choose almost perfectly safe lotteries at moderately unfavorable prices and jump to much riskier lotteries at slightly more extreme prices, consistent with the Disappointment Aversion DA model in Gul (1991). However, the usual two-parameter DA model has essentially the same predictive power in our data as a one-parameter restriction that is consistent with EUH.

Perhaps the most intriguing deviation from EUH, not previously documented to our knowledge, is that our subjects overall respond more strongly to probabilities than to prices. EUH predicts symmetric responses, and the "diminishing sensitivity" of Prospect Theory's probability weighting function predicts asymmetry in the opposite direction from what we find. (One of the authors had also conjectured, incorrectly, that more salient visual information regarding prices — the slopes of budget lines — would provoke a stronger response than probabilities conveyed by mere axis labels.) Although the asymmetry is statistically quite significant, the symmetric EUH prediction remains a decent approximation for most economic purposes.

Our second main finding is that the EUH is a rather poor approximation when comparing behavior across elicitation tasks. The EUH allows subjects to differ in their elicited risk preferences, and indeed our subjects do differ considerably. However, EUH does not allow for differences across elicitation tasks; indeed, for elicited personal Bernoulli functions to be scientifically useful, they must have some sort of stability across tasks. To the contrary, we find large differences across elicitation tasks in the population distributions of γ and RRP. We also find considerable instability in the ordering of individual subjects within a distribution, especially when we compare tasks that are not very closely related. For example, for both the parametric and nonparametric measures, the binary list tasks (HL and BDHL) produced distributions that are more compressed than those for the budget line tasks (BL, BJ and BJn), and the by-subject rank correlations are across the two families are low to nil.

Our third main finding is that, to some degree, these differences and inconsistencies across elicitation tasks can be related to attributes that, according to basic decision theory, should be irrelevant. For example, spatial presentation of information reveals less risk aversion than text presentation. Presenting monotone sequences of prices or probabilities encourages subjects to reveal less risk averse than random sequences. Such differences are inconsequential from the perspective of decision theory, yet the estimated differences in elicited preferences are economically quite large, comparable to the difference between a square root and a linear Bernoulli function.

This last result sharpens the findings of earlier studies such as Isaac and James (2000)), Berg et al. (2005), Loomes and Pogrebna (2014), and Pedroni et al. (2017), among others. These studies, and ours, establish that that any given subject is likely to reveal different risk preferences across different elicitation tasks. It has been less clear up to this point why that happens, and what decision theorists and experimenters should do about it. By restricting our tasks to direct choice between lotteries, we avoid ongoing debates over tangential issues such as possible (bad) strategic bidding or a gap between willingness to pay and willingness to accept, and instead are able to isolate effects associated with specific aspects of task interfaces. It is now clear that conventional decision theory by itself can not explain the nontrivial differences we find across tasks that have exactly the same feasible sets defined by probabilities and prices. The differences in location and dispersion of population distributions, and the scrambled ordering across subjects, requires some other sort of explanation.

Our results on these attributes may help point the way forward. Expanding on a result of Habib et al. (2017), that representing binary lottery choices as rotating cylinders instead of text pushes revealed preferences towards risk neutrality, we find that quite generally it matters whether probability and price information is represented spatially (e.g., as budget lines) or in text. Expanding on Lévy-Garboua et al. (2012), we find that that a monotone (as opposed to random) ordering of probabilities and prices also pushes subjects' behavior towards risk neutrality. Likewise, the theoretically redundant addition of cash to the set of available lotteries seems to influence elicited preferences.

On the theoretical side, such findings suggest to us that decision theory might benefit from incorporating models that are sensitive to how information is acquired and processed, in the spirit of, for example, Pleskac and Busemeyer (2010), and Massaro and Friedman (1990). On the empirical side, these findings suggest seeking new laboratory (and perhaps field) experiments testing for the persistence and robustness of behavioral differences across tasks with different formats but identical opportunities and consequences.

References

James Andreoni and William T. Harbaugh. Unexpected utility: Experimental tests and five key questions about preferences over risk. 2009.

James Andreoni, Michael A. Kuhn, and Charles Sprenger. Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior and Organization*, 116:451–464, 2015.

Gordon M. Becker, Morris H. DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232, 1964.

Joyce Berg, John Dickhaut, and Kevin McCabe. Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences*, 102(11):4209–4214, 2005.

Bruno Biais, Thomas Mariotti, Sophie Moinas, and Sébastien Pouget. Asset pricing and risk—sharing in a complete market: An experimental investigation. 2017.

Hans P. Binswanger. Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, 62(3):395–407, 1980.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Pavlo R. Blavatskyy. Stronger utility. *Theory and Decision*, 76(2):265–286, 2014.

Colin F. Camerer and Teck-Hua Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196, 1994.

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *The American Economic Review*, 97(5):1921–1938, 2007.

Sean M. Collins and Duncan James. Response mode and stochastic choice together explain preference reversals. *Quantitative Economics*, 6:825–856, 2015.

James C. Cox, Bruce Roberson, and Vernon L. Smith. Theory and behavior of single object auctions. In Vernon L. Smith, editor, *Research in Experimental Economics, Vol. 2*, pages 1–44. JAI Press, 1982.

James C. Cox, Vernon L. Smith, and James M. Walker. Theory and individual behavior of first-price auctions. *Journal of Risk and Uncertainty*, 1(1):61–99, 1988.

James C. Cox, Vjollca Sadiraj, and Ulrich Schmidt. Paradoxes and mechanisms for choice under risk. *Experimental Economics*, 18(2):215–250, 2015.

Barbara Erin Crockett and Sean Crockett. Endowments and risky choice. 2018.

Paolo Crossetto and Antonio Filippin. The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65, 2013.

Catherine Eckel and Philip J. Grossman. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4):281–295, 2002.

Catherine Eckel and Philip J. Grossman. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization*, 68(1):1–17, 2008.

David Freeman, Yoram Halevy, and Terri Kneeland. Eliciting risk preferences using choice lists. 2016.

Daniel Friedman, R. Mark Isaac, Duncan James, and Shyam Sunder. *Risky Curves: On the Empirical Failure of Expected Utility*. Routledge, 2014.

Uri Gneezy and Jan Potters. An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645, 1997.

David Grether. Financial incentive effects and individual decisionmaking. Technical Report 401, California Institute of Technology Social Science Working Paper, 1981.

Faruk Gul. A theory of disappointment aversion. *Econometrica*, 59(3):667–686, 1991.

Sameh Habib, Daniel Friedman, Sean Crockett, and Duncan James. Payoff and presentation modulation of elicited risk preferences in MPLs. *Journal of the Economic Science Association*, 3(2): 183–194, 2017.

John Hey and Chris Orme. Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326, 1994.

Charles A. Holt and Susan K. Laury. Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655, 2002.

Leonid Hurwicz. On informationally decentralized systems. In C. B. McGuire and Roy Radner, editors, *Decision and Organization*, pages 297–336. North Holland, Amsterdam, 1972.

R. Mark Isaac and Duncan James. Just who are you calling risk averse? *Journal of Risk and Uncertainty*, 20(2):177–187, 2000.

Duncan James. Incentive compatible elicitation procedures. Perth, Australia, December 2011. 19th International Congress on Modelling and Simulation.

Steven J. Kachelmeier and Mohamed Shehata. Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *The American Economic Review*, 82(5):1120–1141, 1992.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

Edi Karni and Zvi Safra. 'preference reversal' and the observability of preferences by experimental methods. *Econometrica*, 55(3):675–85, 1987.

C. W. Lejuez, Jennifer P. Read, Christopher W. Kahler, Jerry B. Richards, Susan E. Ramsey, Gregory L. Stuart, David R. Strong, and Richard A. Brown. Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (bart). *Journal of Experimental Psychology*, 8(2): 75–84, 2002.

Louis Lévy-Garboua, Hela Maafi, David Masclet, and Antoine Terracol. Risk aversion and framing effects. *Experimental Economics*, 15(1):128–144, 2012.

Sarah Lichtenstein and Paul Slovic. Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1):46–55, 1971.

Sarah Lichtenstein and Paul Slovic. Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology*, 101(1):16–20, 1973.

Graham Loomes and Ganna Pogrebna. Measuring individual risk attitudes when preferences are imprecise. *The Economic Journal*, 124(576):569–593, 2014.

Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

Dominic W Massaro and Daniel Friedman. Models of integration given multiple sources of information. *Psychological review*, 97(2):225, 1990.

Andreas Pedroni, Renato Frey, Adrian Bruhin, Gilles Dutilh, Ralph Hertwig, and Jörg Rieskamp. The risk elicitation puzzle. *Nature Human Behavior*, 1:803–809, 2017.

Timothy J Pleskac and Jerome R Busemeyer. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological review*, 117(3):864, 2010.

Vernon L. Smith. Experiment economics: Induced value theory. *American Economic Review, Papers and Proceedings*, 66(2):274–279, 1976.

Vernon L. Smith. Theory, experiment and economics. *Journal of Economic Perspectives*, 3(1): 151–169, 1989.

Charles Sprenger. An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy*, 123(6):1456–1499, 2015.

Amos Tversky and Daniel Kahneman. Prospect theory: An analysis of decision under risk. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

Appendix A Proof of Proposition

A budget line is the set of lotteries $(x, y) \in \mathbb{R}^2$ satisfying $xp_x + yp_y = m$, where m is an (implicit or explicit) endowment of cash, and $p_x > 0$ and $p_y > 0$ are the prices of the two Arrow securities, with state probabilities $\pi_X, \pi_Y > 0$ and $\pi_X + \pi_Y = 1$.

Recall that a lottery L FOSD's another lottery M if their cumulative distribution functions (cdf's) satisfy $F_M(z) - F_L(z) \geq 0$ for all $z \in \mathbb{R}$, and that the lottery ordering is strict if the inequality is strict for some $z \in \mathbb{R}$.

Proposition 2 . *A lottery $(x, y) \in \mathbb{R}^2$ is strictly first order stochastically dominated by another lottery on the same budget line iff*

a. one Arrow state is more likely and its security is less expensive (e.g., $\pi_X \geq \pi_Y$ and $p_x \leq p_y$), with at least one of these comparisons strict; and

b. the lottery includes strictly less of the less expensive security (e.g., $x < y$).

Proof.[6] First consider the case $\pi_X \geq \pi_Y$ and $p_x < p_y$, and suppose that $x < y$. The cdf for lottery (x, y) is

$$\begin{aligned} F(z) &= 0 \quad \text{if} \quad z < x \\ &= \pi_X \quad \text{if} \quad x \leq z < y \\ &= 1 \quad \text{if} \quad z \geq y \end{aligned}$$

We will construct another lottery (a, b) on the same budget line as (x, y) in two steps, and show that it strictly FOSD's (x, y). First set $a = y$ and $b' = x$, and let G be is corresponding cdf. Then $F(z) - G(z) = 0$ for $z < x$ and $z > y$, but $F(z) - G(z) = \pi_X - \pi_Y \geq 0$ for $x \leq z < y$, so the lottery (a, b') FOSD's (x, y). Now set $b = b' + c/p_y$, where $c = (y - x)(p_y - p_x) > 0$ by hypothesis, and let H be the cdf for the lottery (a, b). Clearly $G(z) = H(z)$ except for $y < z \leq y + c/p_y$, where $G(z) - H(z) = 1 - \pi_X > 0$. Thus (a, b) strictly FOSD's (a, b') and thus, by transitivity, strictly FOSDs (x, y). To complete the proof for the present case we need only verify that the expenditure on (a, b) is the same as on (x, y):

$$ap_x + bp_y = yp_x + (x + c/p_y)p_y = yp_x + xp_y + c = yp_x + xp_y + (y - x)(p_y - p_x) = xp_x + yp_y = m.$$

The other cases have very similar proofs. For example, if $\pi_X > \pi_Y$ and $p_x \leq p_y$, then the conclusion follows from the fact that (a, b') strictly FOSD's (x, y). Of course, we can only guarantee

[6] Adapted slightly from an original proof by Brett Williams.

weak FOSD of (x, y) with $y > x$ when both $\pi_X \geq \pi_Y$ and $p_x \leq p_y$. To show that (x, y) with $y < x$ is FOSD'd when $\pi_X \leq \pi_Y$ and $p_x \geq p_y$, we use precisely the same approach interchanging the roles of X and Y.

To complete the proof, we need only show that no lottery on the budget line is strictly FOSD'd when (i) $\pi_X > \pi_Y$ and $p_x > p_y$ or (ii) $\pi_X < \pi_Y$ and $p_x < p_y$, and to check subcases where the inequalities are weak. Of course, the arguments are the same for (ii) as for (i) due to the symmetric roles of X and Y, so it suffices to consider only case (i). For this case, let F, G be the cdfs for lotteries $(x, y) \neq (a, b)$ on the same budget line. Since the line is negatively sloped, one of the points, say (x, y), is northwest of the other, so $x < a$ and $b < y$. There are now three subcases.

1. Both points are above the diagonal $x' = y'$. Since $p_x > p_y$, we have $x < a < b < y$. It follows that $F(z) - G(z) = \pi_X > 0$ for $x \leq z < a$ but $F(z) - G(z) = \pi_X - 1 < 0$ for $b \leq z < y$. Hence neither point FOSD's the other.

2. Both points are below the diagonal $x' = y'$. Since $p_x > p_y$, we have $b < y < x < a$. It follows that $F(z) - G(z) = 0 - \pi_Y < 0$ for $b \leq z < y$ but $F(z) - G(z) = 1 - \pi_Y > 0$ for $x \leq z < a$; again, no FOSD ranking.

3. $x < y$ but $a > b$. We can not have $x < b < y < a$, as this would imply that the budget line has -slope $\frac{y-b}{a-x} < 1$ but the hypothesis $p_x > p_y$ implies -slope > 1. The other three orderings $b < x < a < y, b < x < y < a, x < b < y < a$, are possible, but each implies a change in the sign of $F(z) - G(z)$. For example, with $b < x < y < a$, we have $F(z) - G(z) = 0 - \pi_Y < 0$ for $b \leq z < x$ but $F(z) - G(z) = 1 - \pi_Y > 0$ for $y \leq z < a$.

The subcases where the inequalities are weak follow from taking limits as $\frac{p_x}{p_y} \to 1$ and $\frac{\pi_X}{\pi_Y} \to 1$.

$\square$

Appendix B Supplementary Analysis

	Dependent variable:ln(x/y)			
	(1)	(2)	(3)	(4)
$\log(\pi_X)$	1.927***	2.048***	1.114***	2.229***
	(0.231)	(0.223)	(0.200)	(0.158)
$\log((1 - \pi_X))$	−0.805***	−0.722***	−1.501***	−0.669***
	(0.197)	(0.194)	(0.173)	(0.136)
$\log(P_x)$	−1.529***	−1.604***	−1.012***	−1.588***
	(0.187)	(0.181)	(0.172)	(0.143)
$\log(P_y)$	0.357	0.138	1.213***	−0.634***
	(0.270)	(0.259)	(0.228)	(0.180)
Observations	5,964	6,804	7,668	9,372
R^2	0.432	0.444	0.463	0.489
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table B.1: Estimates of Equation (17), with errors clustered at subject level. Column 1 data includes only BL, BJ and BJn trials. Column 2 also includes BDEG trials, and column 3 includes BDHL trials as well. Column 4 also includes the HL list data, with each line treated as a separate trial.

One interesting nuance that presents itself when working with binary choice data is whether to treat those data as a whole, as in Holt and Laury, or as a series of individual instances of choice, as per Hey and Orme (1994). Or to put it another way, what would MPL data generate in terms of γ, if they were instead handled by means of regression (akin to Hey and Orme (1994))? Table B.3 shows that within a particular format of MPL (within HL, or within BDHL) the distributions of γ estimates shift (K-S<.001), but ordering within each distribution is largely preserved (0.88 Spearman rank correlation in each case). Otherwise there is substantial scrambling of ordering and/or shift in distribution (when comparing different combinations of data handling and format of MPL). We do not suggest that one or the other method of handling binary choice data is necessarily better or worse, rather that whichever approach one employs will impart its own influence to the estimates derived thereby. This mutability of γ is in addition to the other sources of mutability due to other sources, documented elsewhere in this paper. A point to consider here is that estimates obtained from regressions including HL and BDHL individual decisions also include data from all our other institutions (BL, BJ, BJn and BDEG), unlike in Hey and Orme (1994) where all decisions

Fixed Price	HL-co	BDHL-co	BL	BJ	BJn
HL	0.89	0.47	0.37	0.34	0.32
BDHL	0.30	0.89	0.56	0.50	0.54
BL	0.23	0.54	0.99	0.73	0.82
BJ	0.17	0.41	0.68	0.99	0.78
BJn	0.22	0.56	0.77	0.80	0.99
Fixed Prob.	**HL-co**	**BDEG-adj**	**BL**	**BJ**	**BJn**
HL	0.82	0.26	-0.03	0.03	0.13
BDEG	0.21	0.70	0.56	0.60	0.55
BL	0.09	0.48	0.90	0.68	0.66
BJ	-0.02	0.50	0.64	0.82	0.77
BJn	0.04	0.40	0.50	0.78	0.88

Table B.2: Within-subject Spearman rank correlation by task between RRP (rows) and γ (columns). Fixed price data reported in top panel and fixed probability data in lower panel.

are obtained from binary choices.

	HL-co	BDHL-co	HL	BDHL
HL-co	-	0.03	0.00	0.00
BDHL-co	0.35	-	0.00	0.00
HL	0.88	0.48	-	0.46
BDHL	0.34	0.88	0.49	-

Table B.3: Kolmogorov Smirnov tests p-values (above diagonal), and Spearman rank correlations (below diagonal) for γ estimates obtained from the standard cross-over method (HL-co and BDHL-co) vs treating each line as an observation to estimate γ from equations (15, 16),

$$\ln(\frac{x}{y}) = \alpha_1 ln(\pi_X) + \alpha_2 ln(\pi_Y) + \alpha_3 \ln(P_x) + \alpha_4 ln(P_y) + \varepsilon \tag{17}$$

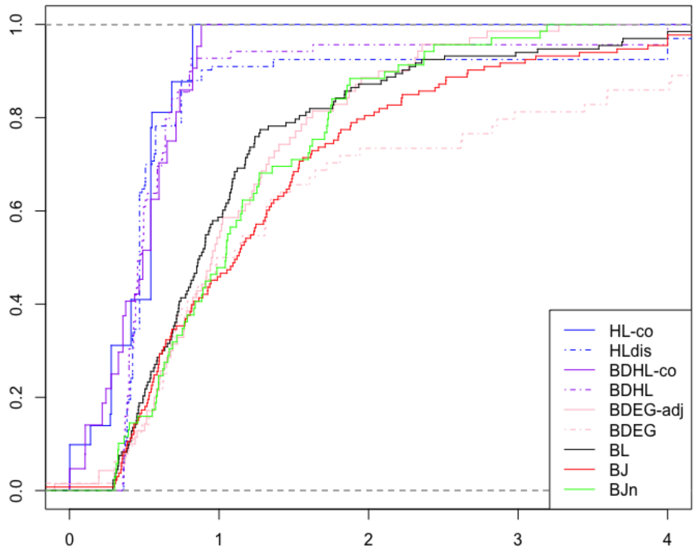

Figure B.1: γ estimates from equations (15, 16) using all observations for all decisions, including each line of the HL list and BDHL as separate observations.

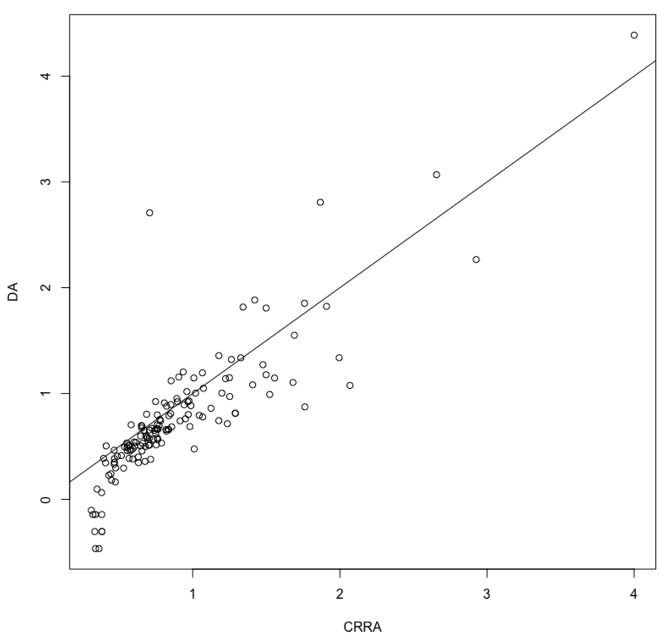

Figure B.2: γ estimates from CRRA and DA models.

	Dependent variable: γ	
	(1)	(2)
spatial	0.074	0.074
	(0.051)	(0.095)
Cash	0.148*	0.148
	(0.079)	(0.107)
dots2	−0.389***	−0.389***
	(0.060)	(0.111)
dots6	0.467***	0.467***
	(0.075)	(0.030)
binary	0.173***	0.173*
	(0.025)	(0.099)
FixProb	0.174	0.174*
	(0.155)	(0.100)
random	0.190***	0.190***
	(0.048)	(0.062)
period	0.001	0.001
	(0.001)	(0.002)
spatial:FixProb	−0.370***	−0.370***
	(0.103)	(0.107)
Cash:FixProb	0.038	0.038
	(0.133)	(0.107)
FixProb:random	−0.180*	−0.180*
	(0.092)	(0.102)
FixProb:period	0.002	0.002
	(0.003)	(0.002)
Observations	10,090	10,090
R^2	0.207	0.207
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table B.4: Dependent variable is implied γ for all decisions, except for HL list, BDHL and BDEG, where data is used twice; once as implied γ for each decision/line and another time where the HL list and BDHL list/screens are combined to generate one γ based on methods intended by authors. For BDEG a single decision is considered twice; once as implied γ and another as intended by the authors. The dummy binary is 1 for the case where the dependent variable is γ calculated from cross point. dots 2 is a dummy equal to 1 for all HL and BDHL decisions regardless of method of calculating γ and similarly for dots6 for BDEG.

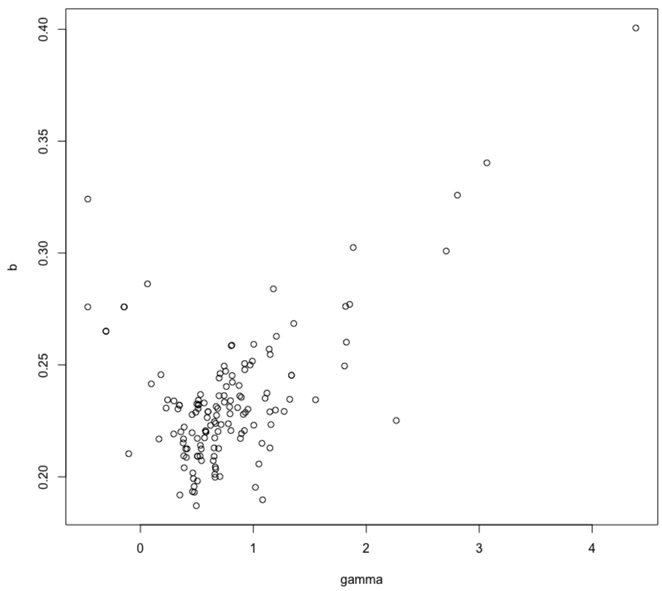

Figure B.3: Scatter plot of γ and b estimtes from DA model for each subject.

Major Violations of FOSD										
Major Cutoff $c =$	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1.0
BL Major Violations	85	60	44	32	30	26	24	22	16	16
(Random)	722	603	519	456	407	367	333	305	280	258
BJC Major Violations	47	26	22	20	16	11	10	10	9	9
(Random)	498	437	391	352	320	292	268	246	226	209
BJ Major Violations	62	43	34	30	28	27	24	21	20	19
(Random)	483	423	377	339	308	281	257	236	218	201

Table B.5: As in Table 1 of the text, except that that the cutoff value c in $L \times ln(\frac{x}{y}) \leq c$ varies as indicated.