# Experimental Payment Protocols and the Bipolar Behaviorist

by

Glenn W. Harrison and J. Todd Swarthout †

March 2014

ABSTRACT.

If someone claims that individuals behave as if they *violate* the independence axiom when making decisions over simple lotteries, it is invariably on the basis of experiments and theories that must *assume* the independence axiom through the use of the random lottery incentive mechanism. We refer to someone who holds this view as a Bipolar Behaviorist, exhibiting pessimism about the axiom when it comes to characterizing how individuals directly evaluate two lotteries in a binary choice task, but optimism about the axiom when it comes to characterizing how individuals evaluate multiple lotteries that make up the incentive structure for a multiple-task experiment. We reject the hypothesis about subject behavior underlying this stance: we find that preferences estimated with a model that assumes violations of the independence axiom are significantly affected when one elicits choices with *procedures* that require the independence assumption, as compared to choices elicited with *procedures* that do not require the assumption. The upshot is that one cannot consistently estimate popular models that relax the independence axiom using data from experiments that assume the validity of the random lottery incentive mechanism.

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); and Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout). E-mail contacts: gharrison@gsu.edu and swarthout@gsu.edu. We are grateful to Jim Cox, Jimmy Martínez, John Quiggin, Elisabet Rutström, Vjollca Sadiraj, Ulrich Schmidt, Uzi Segal and Nathaniel Wilcox for helpful discussions.

The independence axiom plays a central role in most formal statements of expected utility theory (EUT), as well as popular alternative models of decision-making under objective or subjective risk. One such alternative is rank-dependent utility (RDU) theory, which assumes that the independence axiom is invalid in a certain way. The axiom also plays a central role in virtually every experiment used to characterize the way in which risk preferences deviate from EUT, through the use of the random lottery incentive mechanism (RLIM). For example, if someone claims that individuals behave as if they "probability weight" outcomes, and hence *violate* the independence axiom (IA), it is almost always on the basis of experiments and theories that *assume* the IA if the incentives are to be taken seriously. But there is an obvious inconsistency with saying that individuals behave as if they violate the IA on the basis of evidence collected under the maintained assumption that the axiom is magically valid.

This inconsistency has long bothered theorists confronted with experimental data, and there have been responses from theorists and experimentalists. The primary theoretical response has been to argue that there is a way to write out a model of decision-making under risk that allows one to relax the IA but to still allow risk preferences to deviate from EUT predictions and for RLIM to be valid. In effect, to argue an existence proof: even though the inconsistency is real for the most popular alternatives to EUT, there exists a formal alternative to EUT where there is no inconsistency. We discuss this theoretical response later in section 5. The primary experimental response has been to develop some direct tests and some ingenious designs intended to trap the IA under some circumstances.[1] But these direct and indirect experimental tests of the IA have been inconclusive. This is frustrating: either the axiom applies or it does not.

The uneasy state of the literature has evolved to assuming the axiom for the purposes of making the payment protocol of an experiment valid, but rejecting it when characterizing the risk preferences

[1] Cubitt, Starmer and Sugden [1991; p. 119] explain the logic of the *indirect* tests: "Our strategy is to take as the maintained hypothesis that the random lottery design is unbiased. We test this hypothesis in situations in which we have *a priori* expectations that individuals' preferences violate the independence axiom in ways which, if the contamination hypothesis were true, would induce observable biases." All studies using indirect tests of this kind, which of course rest on premisses that might be false, also report direct tests.

exhibited in the same experiment using the standard alternatives to EUT.[2] Those characterizations seem to show evidence of rank-dependent probability weighting, when that very evidence calls into question a maintained assumption of the payment protocol used to generate the evidence. We refer to someone who holds this view as a Bipolar[3] Behaviorist, exhibiting pessimism about the IA when it comes to characterizing how individuals directly evaluate two lotteries in a binary choice task, but optimism about the IA when characterizing how individuals evaluate multiple lotteries that make up the incentive structure for a multiple-task experiment.

The standard payment protocol in individual risky choice experiments involves a subject making K>1 binary choices over objective lotteries, and then selecting one choice at random for payment. We call this protocol 1-in-K. Following Conlisk [1989], Starmer and Sugden [1991], Beattie and Loomes [1997], Cubitt, Starmer and Sugden [1998] and Cox, Sadiraj and Schmidt [2011], an alternative payment protocol, which we call 1-in-1, involves a subject making only one choice, and then being paid with certainty for the single choice.[4] The IA can have no role to play in the validity of the 1-in-1 protocol *per se* if we restrict choice to simple lotteries, but plays a defining role in the 1-in-K protocol. And the role that the IA plays in the theoretical and behavioral validity of the experimental payment protocol is quite distinct from the role that it might play in evaluating the actual binary choice or choices. Even with the 1-in-1 protocol being used, it is possible to ask if behavior is better characterized by violations of IA or not. Indeed, the whole point of our design is to highlight the dual role of the IA in 1-in-K protocols that seek to test violations of IA.

---

[2] An illustrative sample of studies estimating or testing models of rank-dependent utility, for instance, without questioning the inconsistent use of RLIM include Camerer [1989], Starmer [1992], Camerer and Ho [1994], Hey and Orme [1994], Wakker, Erev and Weber [1994] and Harrison and Rutström [2008][2009].

[3] Our use of the term "bipolar" is to convey diametrically-opposed views, and not to imply mental illness. Of course, one could instead view the term as a colorful metaphor, with a little bite to it. Indeed, we openly admit to such bipolar attitudes at times in our own research.

[4] Conlisk [1989; p.406] has a very clear statement of the problem, and the need for the 1-in-1 protocol. He uses the 1-in-1 protocol in his test of the Allais Paradox with real monetary consequences, incidentally finding no evidence whatsoever for the alleged anomaly, but does not test it behaviorally against the 1-in-K protocol. Starmer and Sugden [1991] were the first to undertake that behavioral comparison.

Testing the manner in which the IA interacts with payment protocols used to collect data on observed choice behavior is complicated by the possibility that the Reduction of Compound Lotteries (ROCL) axiom may be invalid behaviorally. This possibility lies at the heart of the theoretical response to the hypothesis about subject behavior underlying the stance of the Bipolar Behaviorist. If the objects of choice are themselves compound lotteries, as is the case in some famous experimental tasks such as the "preference reversal" experiments of Grether and Plott [1979], then one has to take a stand on the validity of ROCL anyway.[5] But if the objects of choice are simple lotteries, as here, then one can test the *implications of RLIM for the standard alternatives to EUT* without taking a position on the validity of ROCL.[6]

We offer direct tests of the effect of payment protocols on preferences for risk in general, and the evidence for probability weighting in particular. We do find statistically significant evidence for a difference in estimated risk preferences deriving from the use of different payment protocols and experimental tasks.

Using choices over simple lotteries, we find evidence of RDU probability weighting with the 1-in-1 protocol that does *not* rely on the validity of the IA. So this result establishes that there is theoretical and behavioral "cause for concern" when one assumes the validity of the IA for the 1-in-K protocol. We then find that this theoretical concern is empirically relevant. Estimated RDU risk preferences *are different* depending on whether one infers them from data collected with the 1-in-1 payment protocol or the 1-in-K payment protocol. It is not the existence of evidence for probability weighting that is the issue, it is the fact that the nature of probability weighting differs in the 1-in-1 *versus* 1-in-K protocol.

---

[5] In those experiments the elicitation procedure for the certainty-equivalents of simple lotteries was, itself, a compound lottery. Hence the validity of the incentives for this design required both Compound IA and ROCL, hence Mixture IA. Holt [1986] and Karni and Safra [1987] showed that if Compound IA was violated, but ROCL and transitivity was assumed, one might still observe choices that suggest "preference reversals." Segal [1988] showed that if ROCL was violated, but Compound IA and transitivity was assumed, that one might also still observe choices that suggest "preference reversals."

[6] To anticipate the language explained in section 1, we are then directly testing the Compound IA and not testing the Mixture IA (the Mixture IA implies the validity of the Compound IA *and* ROCL). Starmer and Sugden [1991], in fact, test the RLIM payment protocol by testing for the validity of ROCL.

In order to justify the use of the 1-in-K payment protocol, many studies appeal to the "isolation effect." This effect is often presented as a *behavioral* assertion that a subject views each choice in an experiment as independent of other choices in the experiment. When stated formally, the isolation effect is often expressed the same as the IA, and is indeed exactly the same as the IA in our choice context. We recognize that the isolation effect is often invoked informally as "an empirical matter," with either an appeal to prior evidence[7] or simply a conjecture that the isolation effect is a reasonable description human behavior. Given limited empirical support from prior studies and the tautology of support via conjecture, we present an experiment which provides a new test of the isolation effect.

In section 1 we describe the theoretical constructs needed for our design, in particular the various axioms that are at issue. In section 2 we present our experimental design, which allows comparison of risk preferences obtained from choice tasks over simple lotteries that do *not* require the IA with risk preferences obtained from tasks that *do* require the assumption. We also explain why we focus on differences in estimated preferences across treatments rather than just examine raw choice patterns. In section 3 we develop the econometric model used to estimate preferences. We pay particular attention to the manner in which between-subject heterogeneity is modeled. The reason for this attention is that the simplest way of avoiding reliance on the IA is to give some individuals only one

[7] Given the ubiquity of the RLIM in the laboratory, surely past studies have definitively verified the empirical validity of the isolation effect? Unfortunately, this is not the case. A few studies have focused on this issue, but conclusions differ and the verdict is still out on whether use of the RLIM biases behavior. All of these studies consider direct and indirect violations of the IA underlying the RLIM. Direct violations come from comparisons of choices 1-in-1 with 1-in-K payment procedures in the experiments, exactly as in our design, and indirect violations come from comparisons of choices that have a "trip-wire" prediction from EUT (and any decision-making model that assumes IA). These indirect violations are variants of the Allais phenomena known as "Common Ratio" effects and "Common Consequence" effects. Focusing just on the direct tests, comparable to our design, we find mixed results in the previous literature. Starmer and Sugden [1991] find the same pattern of choices in one 1-in-1 versus 1-in-2 comparison, and a different pattern in another comparison. Their design only had two pairs of choices, so K=2; indeed, all of the previous studies had a small K. Beattie and Loomes [1997] used K=4, and found no difference between the 1-in-1 and 1-in-4 choices over three pairs of binary lottery choices. They did find a difference between the 1-in-1 and 1-in-4 choices over the single "multiple lottery choice" task, patterned after Binswanger [1981]. Appendix C (available on request) contains a more complete summary of the previous literature.

choice to make, *necessitating* the pooling of choices across different individuals. In the absence of an assumption of homogeneity of risk preferences, or samples of sufficient power to allow randomization to mitigate the need for that assumption, we must address the econometric modeling of heterogeneity. In section 4 we examine the data from our experiments, and present the econometric analysis of hypotheses. In section 5 we draw some general implications of our results, and in section 6 offer general conclusions.

## 1. Theory

*A. Basic Axioms*

Following Segal [1987][1988][1990][1992], we distinguish between three axioms defined over objective lotteries.[8] In words, the **Reduction of Compound Lotteries** (ROCL) axiom states that a decision-maker is indifferent between a compound lottery and the actuarially-equivalent simple lottery in which the probabilities of the two stages of the compound lottery have been multiplied out. To use the language of Samuelson [1952; p.671], the former generates a *compound income-probability-situation*, and the latter defines an *associated income-probability-situation*, and that "...only algebra, not human behavior, is involved in this definition."

To state this more explicitly, with notation to be used to state all axioms, let X, Y and Z denote simple lotteries, A and B denote compound lotteries, ≻ express strict preference, and ~ express indifference. Then the ROCL axiom says that A ~ X if the probabilities and prizes in X are the actuarially-equivalent probabilities and prizes from A. Thus if A is the compound lottery that pays "double or nothing" from the outcome of the lottery that pays $10 if a coin flip is a head and $2 if the

[8] In general we focus throughout on lotteries defined over objective probabilities. Remarkably, Bade [2011] shows that the 1-in-K payment protocol does not immediately generate inferential problems for *some* models of choices over lotteries defined over ambiguous acts, such as the Maxmin Expected Utility model. However, for the popular "smooth" models of ambiguity aversion, the 1-in-K protocol does generate problems if the smooth model is compatible with the notion of stochastic independence.

coin flip is a tail, then X would be the lottery that pays \$20 with probability ½×½ = ¼, \$4 with probability ½×½ = ¼, and nothing with probability ½. From an observational perspective, one would have to see choices between compound lotteries and the actuarially-equivalent simple lottery to test ROCL.

The **Compound Independence Axiom** (CIA) states that a compound lottery formed from two simple lotteries by adding a positive common lottery with the same probability to each of the simple lotteries will exhibit the same preference ordering as the simple lotteries. So this is a statement that the preference ordering of the two constructed compound lotteries will be the same as the preference ordering of the different simple lotteries that distinguish the compound lotteries, provided that the common prize in the compound lotteries is the same and has the same (compound lottery) probability. It says nothing about how the compound lotteries are to be evaluated, and in particular *it does not assume ROCL*. It only restricts the preference *ordering* of the two constructed compound lotteries to match the preference *ordering* of the original simple lotteries.

The CIA says that if A is the compound lottery giving the simple lottery X with probability $\alpha$ and the simple lottery Z with probability $(1-\alpha)$, and B is the compound lottery giving the simple lottery Y with probability $\alpha$ and the simple lottery Z with probability $(1-\alpha)$, then $A \succ B$ iff $X \succ Y$ $\forall\ \alpha \in (0,1)$. So the construction of the two compound lotteries A and B has the "independence axiom" cadence of the common prize Z with a common probability $(1-\alpha)$, but the implication is only that the *ordering* of the compound and constituent simple lotteries are the same. For example, Segal [1992; p.170] defines the CIA by assuming that the second-stage lotteries are replaced by their certainty-equivalent according to some (possibly non-EUT) model, "throwing away" information about the second-stage probabilities before one examines the first-stage probabilities at all. Hence one cannot then define the actuarially-equivalent simple lottery and hence state the ROCL axiom, by construction, since the informational bridge to that calculation has been burnt.

Finally, the **Mixture Independence Axiom** (MIA) says that the preference ordering of two simple lotteries must be the same as the two actuarially-equivalent simple lotteries derived from the two compound lotteries formed by combining a common outcome with one of the original simple lotteries, where the common outcome has the same(compound lottery) probability. That is, $X \succ Y$ iff the actuarially-equivalent simple lottery of $\alpha X + (1-\alpha)Z$ is strictly preferred to the actuarially-equivalent simple lottery of $\alpha Y + (1-\alpha)Z$, $\forall\ \alpha \in (0,1]$. So stated, it is clear that the MIA strengthens the CIA by making a definite statement that the constructed compound lotteries are to be evaluated in a way that is ROCL-consistent. Construction of the compound lottery in the MIA is actually implicit: the axiom only makes observable statements about two pairs of simple lotteries. To restate Samuelson's point about the definition of ROCL, the experimenter testing the MIA could have constructed the associated income-probability-situation without knowing the risk preferences of the individual (although the experimenter would need to know how to multiply).

The reason these three axioms are important for the evaluation of alternatives to EUT is that the failure of MIA does not imply the failure of CIA *and* ROCL. It does imply the failure of one *or* the other, but it is far from obvious which one. Indeed, one could imagine some individuals or task domains where only CIA might fail, only ROCL might fail, or both might fail. Moreover, specific types of failures of ROCL lie at the heart of many important models of decision-making under uncertainty and ambiguity. We use the acronym IA when we mean "CIA or MIA" and the acronyms CIA or MIA directly when the difference matters.

*B. Experimental Payment Protocols*

Turning now to experimental procedures, the most popular payment protocol used in individual choice experiments assumes the validity of the CIA. This payment protocol is called the **Random Lottery Incentive Mechanism** (RLIM). It entails the subject making K choices and then one of the K choices being selected at random to be played out. Typically, and without loss of generality, assume that the selection of the $k^{th}$ task to be played out uses a discrete uniform distribution over the K tasks. Since the other K-1 tasks will generate a payoff of zero, the payment protocol can be seen as a compound lottery that assigns probability $\alpha = 1/k$ to the selected task and $(1-\alpha) = (1-(1/k))$ to the other K-1 tasks as a whole. If the experiment consists of binary choices between simple lotteries X and Y, then the RLIM can be immediately seen to entail an application of the CIA, where Z = U($0) and $(1-\alpha) = (1-(1/k))$, for the utility function U(·). Hence, under the CIA, the preference ordering of X and Y is independent of all of the choices in the other tasks (Holt [1986]).

The CIA can be avoided by setting K=1, and asking each subject to answer one binary choice task for payment. Unfortunately, this comes at the cost of another assumption if one wants to compare choice patterns over two simple lottery pairs, as in most of the popular tests of EUT such as the Allais Paradox and Common Ratio test: the assumption that risk preferences across subjects are the same. This is a strong assumption, obviously, and one that leads to inferential tradeoffs in terms of the "power" of tests of EUT relying on randomization that will vary with sample size. Further, experimenters in this area are wont to ignore the implications this assumption has on power calculations, and so the results from small-sample studies which assume homogeneity of preferences can be questionable due to low power.

The assumption of homogeneous preferences can be diluted, however, by changing it to a conditional form: that risk preferences are homogeneous conditional on a finite set of observable characteristics. Although this sounds like an econometric assumption, and it certainly has statistical implications, it is as much a matter of (operationally meaningful) theory as formal statements of the CIA, ROCL and MIA.

## 2. Experiment

*A. Basic Design Issues*

Our experimental design focuses directly on the risk preferences that one can infer from binary choices over pairs of simple lotteries. This task is canonical, in terms of testing EUT against alternatives such as RDU, as well as for estimating risk preferences. Our design builds on a comparison of the risk preferences implied by 1-in-1 and 1-in-K choice tasks. We let K equal 30, to match the typical risky choice experiment in which there are many choices (e.g., Hey and Orme [1994]). Figure 1 shows the interface given to our subjects. A standard, fixed show-up fee, in our case $7.50, was paid to every subject independently of their lottery choices.

*B. Specific Design*

Table 1 summarizes our experimental design. In **treatment A**, each subject undertakes one 1-in-1 binary choice, where each subject's single lottery pair is drawn at random from a set of 69 lottery pairs shown in Appendix A (available on request). These lottery pairs span five monetary prize amounts, $5, $10, $20, $35 and $70, and five probabilities, 0, ¼, ½, ¾ and 1. The prizes are combined in ten "contexts," defined as a particular triple of prizes.[9] They are based on a battery of lottery pairs developed by Wilcox [2010] for the purpose of robust estimation of EUT and RDU models.[10] These lotteries also contain a number of pairs in which the "EUT-safe" lottery has a *higher* EV than the "EUT-risky" lottery: this is designed deliberately to evaluate the extent of risk premia deriving from probability pessimism rather than diminishing marginal utility.

In treatment A we do *not* have to assume the CIA in order for observed choices to reflect risk preferences under EUT or RDU. In effect, it represents the behavioral Gold Standard benchmark in terms of internal validity, against which the other payment protocols are to be evaluated.

In **treatment B** we move to the 1-in-30 case, which is typical of the usual risk elicitation setting.

---

[9] For example, the first context consists of lotteries defined over the prizes $5, $10 and $20, and the tenth context consists of lotteries defined over the prizes $20, $35 and $70. The significance of the prize context is explained by Wilcox [2010][2011].

[10] The original battery includes repetition of some choices, to help identify the "error rate" and hence the behavioral error parameter, defined later. In addition, the original battery was designed to be administered in its entirety to every subject. We decided *a priori* that 30 choice tasks was the maximum that our subject pool could focus on in any one session, given the need in some sessions for there to be later tasks.

In some cases there was an additional task after the 30 lottery choices, and in some cases there was no other task. In the former case the instructions raised the possibility of a future task for payment.[11] In the latter case we explicitly told subjects that there were no further salient tasks affecting their earnings after the risky lottery task, to avoid them even tacitly thinking of forming a portfolio over the risky lottery tasks and any future tasks. We later test for an effect of there being an extra task and find none, so pool these minor variants into one treatment B.

Every random event determining payouts was generated by the rolling of one or more dice. These dice were illustrated visually during the reading of the instructions,[12] and each subject rolled their own dice.

*C. Why Not Just Look At Raw Choice Patterns?*

Prior tests looked directly at choice patterns. In contrast, we focus on the risk preferences implied by the observed choice data, and do not examine the choice patterns themselves. The reason is that there are limits on what can be inferred by just looking at choice patterns. Since much of the literature on the evaluation of the axioms of EUT has done precisely that, we explain why we believe this to be less informative than trying to make inferences about the underlying latent preferences. This may be particularly important because one might wonder how they *could* differ: after all, if preferences are just rationalizing observed choices, and if observed choices appear to violate the predictions of EUT or IA, how can it be that the implied preferences might not?

[11] To be precise, when there *was* an extra task subjects were told that "All payoffs are in cash, and are in addition to the $7.50 show-up fee that you receive just for being here, as well as any other earnings in other tasks." In all other cases subjects were told that "All payoffs are in cash, and are in addition to the $7.50 show-up fee that you receive just for being here. The only other task today is for you to answer some demographic questions. Your answers to those questions will not affect your payoffs."

[12] A video camera captured images at the front table and broadcasted the images to displays throughout the lab. In addition to a large projection screen in the front, there are 3 wide-screen TV displays spread throughout the lab so that every cubicle has a clear view of the images.

Behavioral Errors

In an important sense, our task would be easier if humans never made mistakes. This would allow us to test deterministic theories of choice, and *any* deviation from the predictions of the theory would provide *prima facie* evidence of a failure of the theory. However, humans do make errors in behavior, and so our task is more complex. The canonical evidence for behavioral errors is the fraction of "switching behavior" observed when subjects are given literally the same lottery pair at different points in a session (e.g., Wilcox [1993]). Any analysis of individual choice ought to account for such behavioral errors. Indeed, some previous analyses of choice patterns have attempted to account for "mistakes" by implementing "trembles" (e.g., Conlisk [1989; Appendix I] and Harless and Camerer [1994]). Such trembles are agnostic about the way any behavioral error might affect the latent components of the choice. A more satisfactory approach would incorporate behavioral errors into the choice process in a more structural manner (Wilcox [2008]).

It is worth emphasizing that behavioral errors are quite distinct conceptually from sampling errors. The former refer to some latent component of the theoretical structure generating a predicted choice. The latter refer to the properties of an estimate of the parameters of that theoretical structure. To see the difference, and assuming a consistent estimator, if the sample size gets larger and larger the sampling errors must get smaller and smaller, but the (point estimate of the) behavioral error need not.[13] In the first instance behavioral errors are the business of theorists, not econometricians, as stressed by Wilcox [2008][2011].

[13] An additional subtlety arises if one posits random coefficients. In this case, the estimates for any structural parameter, such as the behavioral error parameter, will have a distribution that characterizes the population. If that population distribution is assumed to be Gaussian, as is often the case, there will be a point estimate and standard error estimate of the population mean, and a point estimate and standard *error* estimate of the population standard *deviation*. With a consistent estimator, increased sample sizes imply that both standard *error* estimates will decrease, but the point estimate of the population standard *deviation* need not.

Do Choice Patterns Use All Available Information?

Once we recognize that there can be some imprecision in the manner in which preferences translate into observed choices, we obtain another informational advantage from making inferences about preferences estimated from a structural model: a theory about how the *intensity* of a preference for one lottery over another matters. For any given utility function and set of parameter values, and assuming EUT for exposition, a larger difference in the EU of two lotteries matters more for the likelihood of the presumed preferences than a difference in the EU that is close to zero. To see this, assume some parameter values characterizing preferences, and two lottery pairs. One lottery pair, evaluated at those parameter values, implies an EU for the left lottery that is $\varepsilon$ greater than the EU for the right lottery. Another lottery pair, similarly evaluated at those same parameter values, implies an EU for the left lottery that is much greater than $\varepsilon$ more than the EU for the right lottery. An observed choice that is inconsistent with the predicted choice for the second lottery pair matters more for the validity of the assumed parameter values than an inconsistent observed choice for the first lottery pair. This is not the case when one simply looks at the number of consistent and inconsistent choice pairs, as all inconsistent choices are treated as informationally equivalent.

Of course, one has to define the term "intensity" for a given utility representation, and there are theoretical and econometric subtleties involved in normalizing EU differences over different choice contexts, discussed later and in Wilcox [2008][2011]. Structural estimation also typically entails some parametric assumptions, also discussed later, that are not involved with the usual analysis of choice patterns. But there is simply more information used when one evaluates estimated preferences with a structural model. The difference is akin to limited-information inference versus full-information inference in statistics: *ceteris paribus*, it is always better to use more information than less, and the only (statistically) efficient estimators use all information. We admit immediately that things are not all equal, and that *some* parametric assumptions will be needed to undertake what we call the full-information approach here. But we argue that the preference estimation approach is complementary to studying choice patterns, and not an inferior and less direct method of conducting the same analysis.

Are the Stimuli Representative?

Comparison of choice patterns from a paradox test with two pairs of lotteries may support or refute the theory under consideration, but how confident are we that the result is representative of choices over all possible lottery pairs? What if multiple tests using distinct choice *patterns* are conducted and only a single test *pattern* suggests a failure of the theory? Perhaps some theorists are content with a single case of falsification, but others may be concerned that the single failure is a rare exception. For example, it is well-known that violations of EUT tend to occur less frequently when lotteries are in the "interior" of the Marschak-Machina triangle (e.g., Starmer [2000; p.358]). Hence one might draw one negative set of qualitative conclusions about EUT from one battery of stimuli and a different, positive set of qualitative conclusions about EUT from a different battery of stimuli (e.g., Camerer [1989][1992] on the importance of "boundary effects"). As a general model for all sets of stimuli, EUT is still in trouble in this case, to be sure, but inferences about the validity of EUT then need to be nuanced and conditional.

Model estimation can address this "representativeness" issue by presenting subjects with a wide range of lottery pairs, a point first stressed in the experimental economics literature by Hey and Orme [1994]. Of course, there is a tradeoff in doing this: with the 1-in-1 protocol we cannot conduct choice pattern comparisons due to low sample sizes for any given lottery pair.

The Homogeneity Assumption

Another theoretical reason one might want to estimate a structural model of preferences, rather than examine choice data alone, is to better account for heterogeneity of preferences in the 1-in-1 treatment. The analysis of choice patterns must assume preference homogeneity, or perhaps minimally condition on an observable characteristic or two, such as assuming homogeneity within samples of men and women. Some might appeal to large-sample randomization in an attempt to avoid the assumption of homogeneity, but rarely does anyone conduct appropriate power analyses to justify that appeal. By using structural model estimation, observed preference heterogeneity can be ameliorated through the use of demographics controls (e.g., Harrison and Rutström [2008]), and unobserved preference heterogeneity can be ameliorated through the use of random coefficient models (e.g., Andersen, Harrison, Hole, Lau and Rutström [2012]).

*D. Data*

A total of 283 subjects were recruited to participate in experiments at Georgia State University between February 2011 and April 2011. The general recruitment message did not mention the show-up fee or any specific range of possible earnings, and subjects were undergraduate students recruited from across the campus. Table 1 shows the allocations of subjects over our treatments. Instructions for all treatments are presented in Appendix B (available on request). Every subject received a copy of the instructions, printed in color, and had time to read them after being seated in the lab. The instructions were then projected on-screen and read out word-for-word by the same experimenter in every session. Every subject also completed a demographic survey covering standard characteristics. All subjects were paid in cash at the end of each session.

## 3. Econometrics

Our interest is in making inferences about the latent risk preferences underlying observed choice behavior. The estimation approach is to write out a structural model of decision-making, assuming some functional forms if necessary. We focus initially on EUT as the appropriate null, but also consider RDU models of decision-making under risk. The lottery parameters in our design also allow us to estimate the structural model assuming non-parametric specifications of the utility and probability weighting functions, and these non-parametric estimations will be the main focus of inferences whenever possible.

Appendix D (available on request) provides the specifications of our econometric model, and generally follows Harrison and Rutström [2008].

## 4. Results

We consider behavior observed under treatments A and B, and evaluate the hypothesis that risk preferences are the same across the treatments. The initial estimates assume preference homogeneity across subjects, to be able to focus on the interpretation of non-parametric estimates of the utility and probability weighting functions. We then allow for preference heterogeneity using observable demographic characteristics. Finally, we consider parametric estimates, which may be more intuitive and familiar.

Appendix D (available on request) contains the detailed estimates of all structural models. If we use a non-parametric specification, we cannot find any statistically significant effect of the payment protocols when we assume EUT. This is true where we assume preference homogeneity (*p*-value = 0.78) or allow for preference heterogeneity (*p*-value = 0.81). But we do find significant differences when we assume RDU. When we assume preference homogeneity we observe a statistically significant effect at a *p*-value of 0.078. With preference heterogeneity the *p*-value across all parameters, those characterizing utility and probability weighting, is only 0.15; but if we evaluate the effect for just the probability weighting parameters the p-value is 0.07.

Turning to parametric estimates, since they are familiar and easier to visualize, we again find no effect of payment protocols under EUT (*p*-value = 0.98). In this case we use the Expo-Power utility function proposed by Saha [1993], and implemented by Holt and Laury [2002], since it allows for non-constant relative risk aversion. We find a *p*-value of 0.06 with the RDU model, across all parameters, and note again that the culprit appears to be the probability weighting behavior (the *p*-value on just those parameters is 0.07). In this case we again use the Expo-Power utility function, as well as the flexible two-parameter Prelec [1998] probability weighting function These results assume preference heterogeneity across individuals Figure 1 shows the difference between RDU probability weighting functions when moving from the 1-in-1 payment protocol to the 1-in-K payment protocol. The differences are striking, quantitatively *and* qualitatively.

## 5. Implications

A first implication of our results is to encourage theorists to come up with payment protocols that allow one to elicit multiple choices but do not require that one violate an assumption required for the coherent specification of the particular decision model. This challenge has been directly addressed, and partially met, by Cox, Sadiraj and Schmidt [2011]. There are no known, or obvious, payment protocols that can be used for RDU and Cumulative Prospect Theory (CPT).

A second implication of our results is to question inferences made about *specific alternative hypotheses* to EUT when the 1-in-K protocol has been employed. That is, in literally every test of specific alternatives to EUT of which we are aware. This is not to say that EUT is valid, just that tests of the validity of specific alternatives rest on a maintained assumption that appears to be false.[14]

A third, costly implication of our results, then, is to place a premium on collecting choice data in smaller doses, using 1-in-1 payment protocols. Anyone proposing new or robust anomalies should be encouraged to do this, and demonstrate that the alleged misbehavior persists when one removes the obvious theoretical confound.

A fourth, modeling implication of the need for 1-in-1 choice data is to place greater urgency on the use of rigorous econometric methods to flexibly characterize heterogeneous preferences.

[14] Our results suggest a research strategy to properly evaluate the validity of EUT in an efficient manner. Examine the catalog of anomalies that arise in choice tasks over simple lotteries using a 1-in-K payment protocol, for some large K, and then for those anomalies that survive, drill down with the more expensive 1-in-1 protocol. This strategy does run the risk that there could be "offsetting violations" of EUT in the 1-in-K payment protocol, but that is a tradeoff that many scholars would, we believe, be willing to take in the interests of efficient use of an experimental budget. And the alternative to the tradeoff is simple enough: replicate every anomaly using the 1-in-1 payment protocol, as in the non-hypothetical experiments of Conlisk [1989].

A fifth implication is to consider more rigorously the learning behavior that might change behavior towards lottery choices such as these.[15] The argument is that one would expect 1-in-1 behavior to differ from 1-in-30 behavior since the latter reflects some learning behavior. The problem with this line of argument is that it is silent as to what should be compared to what, and does not provide a metric for defining when learning is finished.[16]

A sixth implication is to formally model the effects of treating behavior as if generated by portfolio formation for the experiment as a whole. In effect, this is the implication of the theoretical "Recursive RDU" response to the problems posed by the RLIM payment procedure.[17] Even though nobody has ever estimated or empirically implemented such a model, the fact that one can formally write one out is regarded as enough to validate the use of the RLIM. The Recursive RDU model is not the standard RDU model, and it is the latter model that is the one that appears in all empirical work. It is apparent that the estimation of a Recursive RDU model for $K \ggg 0$, as in our experimental design, is infeasible.

A final implication is to just be honest when presenting experimental findings on RDU and CPT models about the assumed neutrality of the experimental payment protocol. In effect this is just saying

---

[15] Binmore [2007; p. 6ff.] has long made the point that we ought to recognize that the artefactual nature of the usual laboratory tasks, and indeed some tasks in the field, means that we should allow subjects to learn how to behave in that environment before drawing unconditional conclusions. Although his immediate arguments are about the study of strategic behavior in games, they are general. These arguments also suggest that a 1-in-1 payment protocol might not be the Gold Standard if one is interested in external validity, whatever its role in terms of internal validity.

[16] One could mitigate the issue by providing subjects with lots of experience in one session, and then invite them back for further experiments, either 1-in-1 or 1-in-30, arguing on *a priori* grounds that any differences in behavior then should reflect longer-run, steady-state behavior for this task. We are sympathetic to this view, and indeed it was implicit in the early days of experimental economics where "experience" meant that subjects has participated in some task and then had time to "sleep on it" before the next session. The hypothesis implied here is that the differences we find would diminish if subjects were given "enough" experience, which is of course testable if someone can ever define what "enough" means.

[17] In fact, one of the earliest statements of this Recursive RDU model by Segal [1988] was in the context of offering an explanation of preference reversals behavior being logically consistent with the validity of the Compound IA. It is therefore theoretically possible that the RLIM procedure is suspect but the Compound IA is valid, so that one does not have to take a bipolar stance about the Compound IA after all. On the other hand, evidence that payment protocols do affect behavior is then evidence against the Mixture IA, so the hypothesis to be tested to support the stance of the Bipolar Behaviorist is then with respect to ROCL.

that there might be two independence axioms at work: one for the evaluation of a given lottery in a binary choice setting, and another one for an evaluation of sets of choices in 1-in-K settings. If one estimates RDU and CPT models with a 1-in-K protocol one might claim to be allowing the first axiom to be relaxed while maintaining the second. It is logically possible for the latter axiom to be empirically false while the former axiom is empirically true. In the absence of better alternatives, we do this in our own ongoing research using 1-in-K protocols.

## 6. Conclusions

We have demonstrated that estimated non-EUT preferences are sensitive to whether behavior is elicited with the RLIM or instead with a "one-shot" design. We do so by considering a simple and popular alternative to EUT: a standard RDU model that relaxes the CIA. Of course, this RDU specification is identical to the "gain frame" part of CPT, so CPT inherits all of the issues we have raised with respect to inferring risk preferences with RDU. Experimental economists should take this result seriously, and recognize the apparent problem of inferring preferences through use of the RLIM and treating these results "as if" they are the same as those from a 1-in-1 scenario.

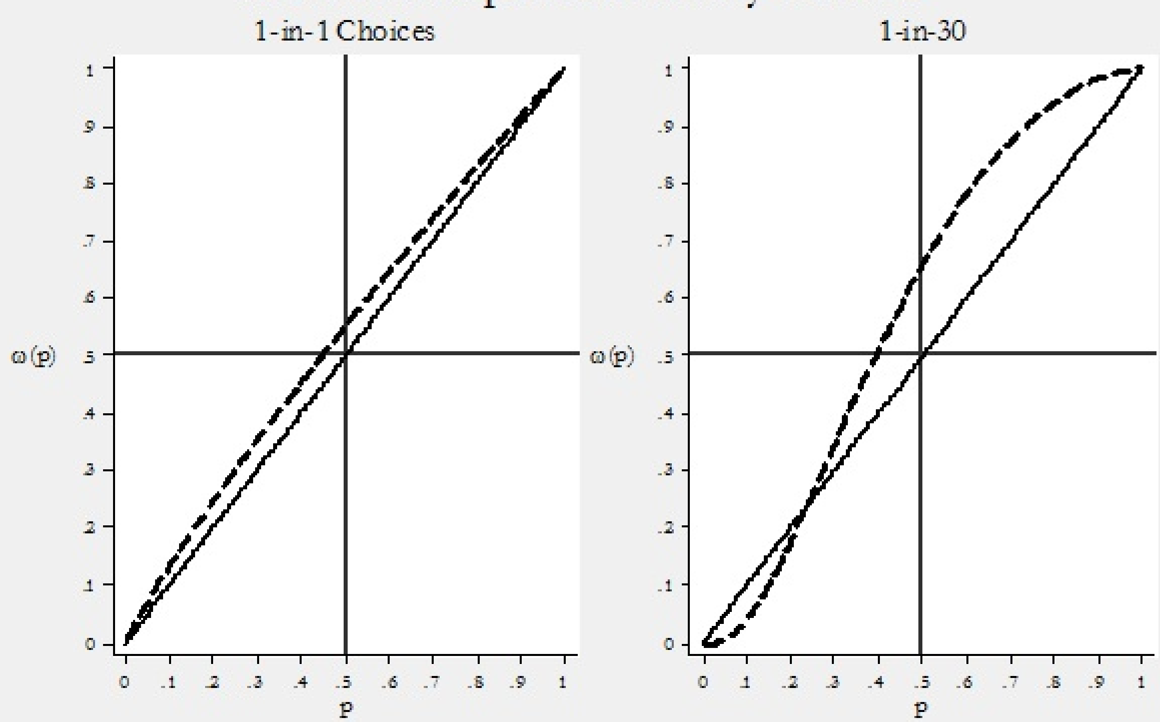
**Table 1: Experimental Design**

All choices drawn from the same battery of 69 lottery pairs at random.
All subjects receive a $7.50 show-up fee.
Unless otherwise noted for treatment B, subjects were told that there would be no other salient task in the experiment.

| Treatment | Subjects | Choices |
|---|---|---|
| **A**. 1-in-1 | 75 | 75 |
| **B**. 1-in-30 † | 208 | 6240 |

Notes: † For 171 subjects, and 5130 choices, there was an additional task after the 30 lottery choices. This was a time-discounting choice, and these subjects were told at the outset that there could be additional salient tasks.

Figure 1: Bipolar Probability Weighting Functions for Rank-Dependent Utility Models



## References


Andersen, Steffen; Harrison, Glenn W.; Hole, Arne Rise, Lau, Morten I., and Rutström, E. Elisabet, "Non-Linear Mixed Logit," *Theory and Decision*, 73, 2012, 77-96.

Bade, Sophie, "Independent Randomization Devices and the Elicitation of Ambiguity Averse Preferences," *Working Paper*, Max Planck Institute for Research on Collective Goods, Bonn, 2011.

Beattie, J., and Loomes, Graham, "The Impact of Incentives Upon Risky Choice Experiments," *Journal of Risk and Uncertainty*, 14, 1997, 149-162.

Binmore, Ken, *Does Game Theory Work? The Bargaining Challenge* (Cambridge, MA: MIT Press, 2007).

Binswanger, Hans P., "Attitudes Toward Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, August 1980, 395-407.

Camerer, Colin F., "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 1989, 61-104.

Camerer, Colin F., "Recent Tests of Generalizations of Expected Utility Theory," in W. Edwards (ed.), *Utility Theories: Measurements and Applications* (Boston: Kluwer, 1992).

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

Conlisk, John, "Three Variants on the Allais Example," *American Economic Review*, 79(3), June 1989, 392-407.

Cubitt, Robin P.; Starmer, Chris, and Sugden, Robert, "On the Validity of the Random Lottery Incentive System," *Experimental Economics*, 1(2), 1998, 115-131.

Cox, James C.; Sadiraj, Vjollca, and Schmidt, Ulrich, "Paradoxes and Mechanisms for Choice under Risk," *Working Paper 2011-12*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2011 (revised 2014); forthcoming, *Experimental Economics*.

Grether, David M., and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69(4), September 1979, 623-638.

Harless, David W., and Camerer, Colin F., "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62(6), November 1994, 1251-1289.

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility *And* Prospect Theory: One Wedding and A Decent Funeral," *Experimental Economics*, 12(2), June 2009, 133-158.

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Holt, Charles A., "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, June 1986, 508-514.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Karni, Edi, and Safra, Zvi, "Preference Reversals and the Observability of Preferences by Experimental Methods," *Econometrica*, 55, 1987, 675-685.

Loomes, Graham, and Sugden, Robert, "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 1998, 581-598.

Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66, 1998, 497-527.

Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.

Saha, Atanu, "Expo-Power Utility: A Flexible Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75(4), November 1993, 905-913.

Samuelson, Paul A., "Probability, Utility, and the Independence Axiom," *Econometrica*, 20, 1952, 670-678.

Segal, Uzi, "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28, 1987, 145-154.

Segal, Uzi, "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review*, 78(1), March 1988, 233-236.

Segal, Uzi, "Two-Stage Lotteries Without the Reduction Axiom," *Econometrica*, 58(2), March 1990, 349-377.

Segal, Uzi, "The Independence Axiom Versus the Reduction Axiom: Must We Have Both?" in W. Edwards (ed.), *Utility Theories: Measurements and Applications* (Boston: Kluwer Academic Publishers, 1992).

Starmer, Chris, "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect," *Review of Economic Studies*, 59, 1992, 813-830.

Starmer, Chris, and Sugden, Robert, "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review*, 81, 1991, 971-978.

Wakker, Peter P.; Erev, Ido, and Weber, Elke U., "Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories," *Journal of Risk and Uncertainty*, 9, 1994, 195-230.

Wilcox, Nathaniel T., "Lottery Choice: Incentives, Complexity, and Decision Time," *Economic Journal*, 103, 1993, 1397-1417.

Wilcox, Nathaniel T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison," in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Wilcox, Nathaniel T., "A Comparison of Three Probabilistic Models of Binary Discrete Choice Under Risk," *Working Paper*, Economic Science Institute, Chapman University, March 2010.

Wilcox, Nathaniel T., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 162(1), May 2011, 89-104.

**APPENDICES TO**

# Experimental Payment Protocols and the Bipolar Behaviorist

by

Glenn W. Harrison and J. Todd Swarthout

March 2014

**Table of Contents**

## Appendix A: Parameters of Experiments

### Table A1: Lotteries in Experiments

| Pair | Context | Prizes Low | Prizes Middle | Prizes High | "Safe" Lottery Probabilities Low | "Safe" Middle | "Safe" High | "Risky" Lottery Probabilities Low | "Risky" Middle | "Risky" High | EV Safe | EV Risky |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | $5 | $10 | $20 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $10.00 | $16.25 |
| 2 | 1 | $5 | $10 | $20 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $8.75 | $12.50 |
| 3 | 1 | $5 | $10 | $20 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $10.00 | $12.50 |
| 4 | 1 | $5 | $10 | $20 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $7.50 | $8.75 |
| 5 | 1 | $5 | $10 | $20 | 0 | 1 | 0 | 0.25 | 0.5 | 0.25 | $10.00 | $11.25 |
| 6 | 1 | $5 | $10 | $20 | 0.25 | 0.5 | 0.25 | 0.5 | 0 | 0.5 | $11.25 | $12.50 |
| 7 | 1 | $5 | $10 | $20 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $15.00 | $16.25 |
| 8 | 1 | $5 | $10 | $20 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $12.50 | $12.50 |
| 9 | 1 | $5 | $10 | $20 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $8.75 | $8.75 |
| 10 | 1 | $5 | $10 | $20 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $10.00 | $8.75 |
| 11 | 2 | $5 | $10 | $35 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $10.00 | $20.00 |
| 12 | 2 | $5 | $10 | $35 | 0 | 0.75 | 0.25 | 0.25 | 0 | 0.75 | $16.25 | $27.50 |
| 13 | 2 | $5 | $10 | $35 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $8.75 | $12.50 |
| 14 | 2 | $5 | $10 | $35 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $22.50 | $27.50 |
| 15 | 2 | $5 | $10 | $35 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $16.25 | $20.00 |
| 16 | 2 | $5 | $10 | $35 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $10.00 | $12.50 |
| 17 | 3 | $5 | $10 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $8.75 | $37.50 |
| 18 | 3 | $5 | $10 | $70 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $10.00 | $37.50 |
| 19 | 3 | $5 | $10 | $70 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $7.50 | $21.25 |
| 20 | 3 | $5 | $10 | $70 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $10.00 | $21.25 |
| 21 | 4 | $5 | $20 | $35 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $20.00 | $27.50 |
| 22 | 4 | $5 | $20 | $35 | 0 | 0.75 | 0.25 | 0.25 | 0 | 0.75 | $23.75 | $27.50 |
| 23 | 4 | $5 | $20 | $35 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $27.50 | $27.50 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 4 | $5 | $20 | $35 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $20.00 | $20.00 |
| 25 | 4 | $5 | $20 | $35 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $12.50 | $12.50 |
| 26 | 4 | $5 | $20 | $35 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $23.75 | $20.00 |
| 27 | 4 | $5 | $20 | $35 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $16.25 | $12.50 |
| 28 | 5 | $5 | $20 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $16.25 | $37.50 |
| 29 | 5 | $5 | $20 | $70 | 0 | 0.75 | 0.25 | 0.25 | 0 | 0.75 | $32.50 | $53.75 |
| 30 | 5 | $5 | $20 | $70 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $12.50 | $21.25 |
| 31 | 5 | $5 | $20 | $70 | 0.25 | 0.5 | 0.25 | 0.5 | 0 | 0.5 | $28.75 | $37.50 |
| 32 | 5 | $5 | $20 | $70 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $16.25 | $21.25 |
| 33 | 5 | $5 | $20 | $70 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $45.00 | $53.75 |
| 34 | 6 | $5 | $35 | $70 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $35.00 | $53.75 |
| 35 | 6 | $5 | $35 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $27.50 | $37.50 |
| 36 | 6 | $5 | $35 | $70 | 0 | 0.75 | 0.25 | 0.25 | 0 | 0.75 | $43.75 | $53.75 |
| 37 | 6 | $5 | $35 | $70 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $20.00 | $21.25 |
| 38 | 6 | $5 | $35 | $70 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $52.50 | $53.75 |
| 39 | 6 | $5 | $35 | $70 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $43.75 | $37.50 |
| 40 | 6 | $5 | $35 | $70 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $27.50 | $21.25 |
| 41 | 6 | $5 | $35 | $70 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $35.00 | $21.25 |
| 42 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $20.00 | $28.75 |
| 43 | 7 | $10 | $20 | $35 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $17.50 | $22.50 |
| 44 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.25 | 0.25 | 0.5 | $20.00 | $25.00 |
| 45 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $20.00 | $22.50 |
| 46 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.25 | 0.5 | 0.25 | $20.00 | $21.25 |
| 47 | 7 | $10 | $20 | $35 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $23.75 | $22.50 |
| 48 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.5 | 0.25 | 0.25 | $20.00 | $18.75 |
| 49 | 7 | $10 | $20 | $35 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $17.50 | $16.25 |
| 50 | 7 | $10 | $20 | $35 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $20.00 | $16.25 |
| 51 | 8 | $10 | $20 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $17.50 | $40.00 |
| 52 | 8 | $10 | $20 | $70 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $15.00 | $25.00 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53 | 8 | $10 | $20 | $70 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $17.50 | $25.00 |
| 54 | 9 | $10 | $35 | $70 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $35.00 | $55.00 |
| 55 | 9 | $10 | $35 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $28.75 | $40.00 |
| 56 | 9 | $10 | $35 | $70 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $52.50 | $55.00 |
| 57 | 9 | $10 | $35 | $70 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $43.75 | $40.00 |
| 58 | 10 | $20 | $35 | $70 | 0 | 1 | 0 | 0.25 | 0 | 0.75 | $35.00 | $57.50 |
| 59 | 10 | $20 | $35 | $70 | 0.25 | 0.75 | 0 | 0.5 | 0 | 0.5 | $31.25 | $45.00 |
| 60 | 10 | $20 | $35 | $70 | 0 | 0.75 | 0.25 | 0.25 | 0 | 0.75 | $43.75 | $57.50 |
| 61 | 10 | $20 | $35 | $70 | 0 | 1 | 0 | 0.5 | 0 | 0.5 | $35.00 | $45.00 |
| 62 | 10 | $20 | $35 | $70 | 0.5 | 0.5 | 0 | 0.75 | 0 | 0.25 | $27.50 | $32.50 |
| 63 | 10 | $20 | $35 | $70 | 0 | 1 | 0 | 0.25 | 0.5 | 0.25 | $35.00 | $40.00 |
| 64 | 10 | $20 | $35 | $70 | 0.25 | 0.5 | 0.25 | 0.5 | 0 | 0.5 | $40.00 | $45.00 |
| 65 | 10 | $20 | $35 | $70 | 0 | 0.5 | 0.5 | 0.25 | 0 | 0.75 | $52.50 | $57.50 |
| 66 | 10 | $20 | $35 | $70 | 0 | 1 | 0 | 0.5 | 0.25 | 0.25 | $35.00 | $36.25 |
| 67 | 10 | $20 | $35 | $70 | 0.25 | 0.75 | 0 | 0.75 | 0 | 0.25 | $31.25 | $32.50 |
| 68 | 10 | $20 | $35 | $70 | 0 | 0.75 | 0.25 | 0.5 | 0 | 0.5 | $43.75 | $45.00 |
| 69 | 10 | $20 | $35 | $70 | 0 | 1 | 0 | 0.75 | 0 | 0.25 | $35.00 | $32.50 |

## Appendix B: Instructions

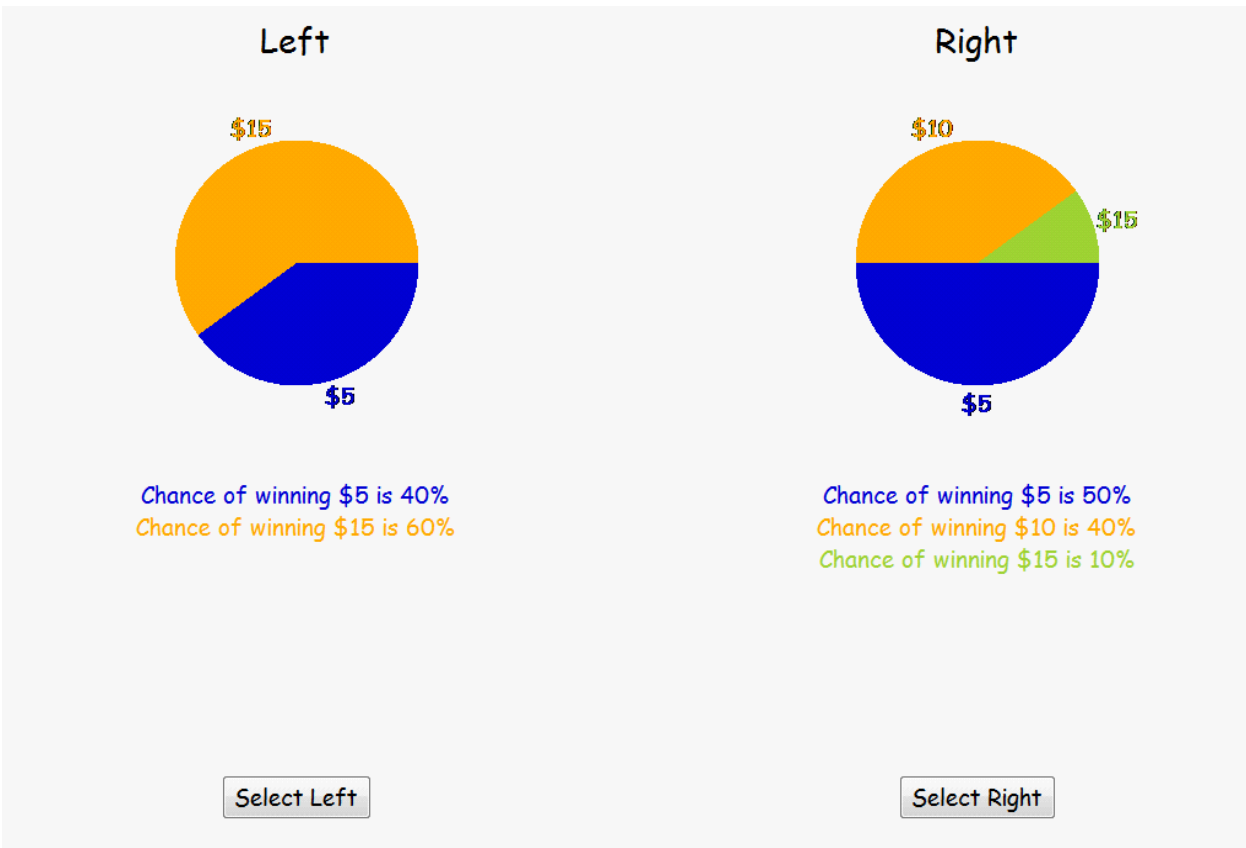
The original instructions used in all experiments are available on request.

*Treatment A: 1-in-1*

### Choices Over Risky Prospects

This is a task where you will choose between prospects with varying prizes and chances of winning. You will be presented with one pair of prospects where you will choose one of them. You should choose the prospect you prefer to play. You will actually get the chance to play the prospect you choose, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer.

Here is an example of what the computer display of a pair of prospects will look like.




The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

In the above example the left prospect pays five dollars ($5) if the number drawn is between 1 and 40, and pays fifteen dollars ($15) if the number is between 41 and 100. The blue color in the pie chart corresponds to 40% of the area and illustrates the chances that the number drawn will be between 1 and 40 and your prize will be $5. The orange area in the pie chart corresponds to 60% of the area and illustrates the chances that the number drawn will be between 41 and 100 and your prize will be $15.

Now look at the pie in the chart on the right. It pays five dollars ($5) if the number drawn is between 1 and 50, ten dollars ($10) if the number is between 51 and 90, and fifteen dollars ($15) if the number is between 91 and 100. As with the prospect on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $15 pie slice is 10% of the total pie.

The pair of prospects you choose from is shown on a screen on the computer. On that screen, you should indicate which prospect you prefer to play by clicking on one of the buttons beneath the prospects.

After you have made your choice, raise your hand and an experimenter will come over. It is certain that your one choice will be played out for real. You will roll the two ten-sided dice to determine the outcome of the prospect you chose.

For instance, suppose you picked the prospect on the left in the above example. If the random number was 37, you would win $5; if it was 93, you would get $15. If you picked the prospect on the right and drew the number 37, you would get $5; if it was 93, you would get $15.

Therefore, your payoff is determined by two things:

- by which prospect you selected, the left or the right; and
- by the outcome of that prospect when you roll the two 10-sided dice.

Which prospects you prefer is a matter of personal taste. The people next to you may be presented with a different prospect, and may have different preferences, so their responses should not matter to you. Please work silently, and make your choices by thinking carefully about the prospect you are presented with.

All payoffs are in cash, and are in addition to the $7.50 show-up fee that you receive just for being here. The only other task today is for you to answer some demographic questions. Your answers to those questions will not affect your payoffs.

*Treatment B1: 1-in-30 Sequential*

**Choices Over Risky Prospects**

This is a task where you will choose between prospects with varying prizes and chances of winning. You will be presented with a series of pairs of prospects where you will choose one of them. There are 30 pairs in the series. For each pair of prospects, you should choose the prospect you prefer to play. You will actually get the chance to play one of the prospects you choose, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer.

Here is an example of what the computer display of such a pair of prospects will look like.

SAME DISPLAY AS FOR TREATMENT A

The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

In the above example the left prospect pays five dollars ($5) if the number drawn is between 1 and 40, and pays fifteen dollars ($15) if the number is between 41 and 100. The blue color in the pie chart corresponds to 40% of the area and illustrates the chances that the number drawn will be between 1 and 40 and your prize will be $5. The orange area in the pie chart corresponds to 60% of the area and illustrates the chances that the number drawn will be between 41 and 100 and your prize will be $15.

Now look at the pie in the chart on the right. It pays five dollars ($5) if the number drawn is between 1 and 50, ten dollars ($10) if the number is between 51 and 90, and fifteen dollars ($15) if the number is between 91 and 100. As with the prospect on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $15 pie slice is 10% of the total pie.

Each pair of prospects is shown on a separate screen on the computer. On each screen, you should indicate which prospect you prefer to play by clicking on one of the buttons beneath the prospects.

After you have worked through all of the pairs of prospects, raise your hand and an experimenter will come over. You will then roll a 30-sided die to determine which pair of prospects will be played out. Since there is a chance that any of your 30 choices could be played out for real, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will roll the two ten-sided dice to determine the outcome of the prospect you chose.

For instance, suppose you picked the prospect on the left in the above example. If the random number was 37, you would win $5; if it was 93, you would get $15. If you picked the prospect on the right and drew the number 37, you would get $5; if it was 93, you would get $15.

Therefore, your payoff is determined by three things:

- by which prospect you selected, the left or the right, for each of these 30 pairs;
- by which prospect pair is chosen to be played out in the series of 30 such pairs using the 30-sided die; and
- by the outcome of that prospect when you roll the two 10-sided dice.

Which prospects you prefer is a matter of personal taste. The people next to you may be presented with different prospects, and may have different preferences, so their responses should not matter to you. Please work silently, and make your choices by thinking carefully about each prospect.

All payoffs are in cash, and are in addition to the $7.50 show-up fee that you receive just for being here. The only other task today is for you to answer some demographic questions. Your answers to those questions will not affect your payoffs.

*Treatment B2: 1-in-30 With an Additional Paid Task*

These instructions were identical to those for Treatment B, apart from the language changes in the final paragraph described in the text.

## Appendix C: Literature Review

Starmer and Sugden [1991], Beattie and Loomes [1997] and Cubitt, Starmer and Sugden [1998] have directly studied the Random Lottery Incentive Method, which of course relies on the validity of the IA. All of these studies consider direct and indirect violations of the IA.[1] Direct violations come from comparisons of choices 1-in-1 with 1-in-K payment procedures in the experiments, and indirect violations come from comparisons of choices that have a "trip-wire" prediction from EUT (and any decision-making model that assumes IA). These indirect violations are variants of the Allais [1953] phenomena known as "Common Ratio" effects and "Common Consequence" effects.[2]

Following **Cubitt, Starmer and Sugden** [1998; p.119], let a and b be monetary prizes, such that a>b>0. Consider the risky prospects

R1: {a, λ; 0, 1-λ}     R2: {a, λp; b, 1-p; 0, (1-λ)p}     R3: {a, λp; 0, 1-λp}

and the safe prospects

S1: {b, 1}     S2: {b, 1}     S3: (b, p; 0, 1-p},

---

[1] This is clearly recognized by Starmer and Sugden [1991; p.973]: "The success of the experiment depended on our finding systematic violations of expected-utility theory for real choices. Provided we found these, we would be able to test whether subjects behaved according to the reduction principle by investigating whether the same violations were found with the random-lottery design. The experiment also allows a second kind of test of the random-lottery design: if random-lottery experiments elicit true preferences, we should expect to find no significant difference between subjects' responses to the random lottery and real-choice designs." Their reduction principle applies the IA, and their "random lottery" design is what we refer to as a 1-in-K payment protocol.

[2] The first test of the common consequence form of the Allais Paradox using incentivized 1-in-1 choices, that do not assume IA, is due to Conlisk [1989]. He found striking evidence of no violations of IA. On the other hand, this result is not welcome in some circles. Cubitt, Starmer and Sugden [1989; p. 130] comment that these results are "... sometimes quoted as evidence that violations of EUT are less frequent in single choice than in random lottery designs. Conlisk investigated the Common Consequence effect using a single choice design. In each of the two relevant tasks, almost all subjects (26 out of 27 in one case, 24 out of 26 in the other) chose the riskier option. Clearly, this distribution of responses between riskier and safer choices is far too asymmetric for the experiment to be a satisfactory test for systematic deviations from EUT." The logic of the final sentence is hard to ascertain. Moreover, the evidence for the Common Consequence effect in *incentivized* 1-in-K choices is decidedly mixed: Burke, Carter, Gominiak and Ohl [1996] and Fan [2002] find no evidence of an EUT violation, whereas Starmer and Sugden [1991] do, as discussed below.

where $0<\lambda<1$ and $0<p\leq 1$. With these lotteries, the IA under EUT implies that preferences over R1 and S1 will be the same as preferences over R2 and S2 and also over R3 and S3. The Common Consequence effect is said to occur when one observes a greater fraction of risky choices over R3 and S3 than over R2 and S2, and a Common Ratio effect is said to occur when one observes a greater fraction of risky choices over R3 and S3 than over R1 and S1. These differences in the fractions of risky choices indirectly imply a statistically significant difference in risk preferences.

**Starmer and Sugden** [1991] present subjects with two pairs of lotteries, making up a Common Consequence test of EUT. In one 1-in-1 treatment 40 subjects were given each lottery pair to make a choice over, and over two 1-in-2 treatments 80 subjects were given the two lotteries to make a choice over. Their Common Consequence test between 1-in-1 choices shows evidence of a clear violation of the EUT prediction, and the IA. Using a *one-sided* Fisher Exact test, since there is an *a priori* prediction of direction, albeit from previously observed behavior from hypothetical tasks, we calculate a *p*-value of 0.021 on the prediction of the EUT hypothesis.[3] Their direct tests of the IA axiom, from comparisons of choices in the same lotteries across the 1-in-1 and 1-in-2 payment protocols, show mixed results. Since there is no prior hypothesis as to the direction of the effect of relying on the IA in the 1-in-2 treatment, it is appropriate in this case to use *two-sided* Fisher Exact tests of the hypothesis that the patterns of choice in each treatment are the same. For one lottery pair the *p*-value on this hypothesis is 0.23, and for the other lottery pair the *p*-value is 0.055. These data provide clear evidence for outright pessimism with respect to the IA: nothing bipolar here.

**Beattie and Loomes** [1997] examined 4 lottery choice tasks. The first 3 tasks involved a binary choice between two lotteries, and the fourth task involved the subject selecting one of four possible lotteries. For each of the 4 choice tasks they had a 1-in-1 treatment, and a 1-in-4 treatment,

[3] A comparable test, using the data from the 1-in-2 choices, also rejects the EUT hypothesis, in this case with a *p*-value of 0.018.

conducted on a between subjects basis. Sample sizes were 48, 47, 48 and 50 subjects in each of the 1-in-1 treatments for the 4 tasks, and 48 subjects for the 1-in-4 treatment. The *p*-values for the two-sided test of the 1-in-1 and 1-in-4 choices for the same lottery pairs are 0.42, 0.84, 0.77, and 0.058, respectively, for each choice task. Over all 4 tasks, the *p*-value is only 0.51. But there is a significant effect for one of the 4 tasks, and this is a task that is essentially the same as the popular method developed by Binswanger [1980]: subjects are offered an ordered set of choices that increase the average payoff while increasing variance. There is no direct evidence of an effect from the RLIM in the binary choice tasks.

On the other hand, there is clear, but indirect, evidence of a violation of the IA in binary choices by a comparison of two Common Ratio pairs in their set of choice tasks. For the 1-in-1 choices for these two pairs, a Fisher Exact test can reject the hypothesis of the same choices, as predicted under EUT, with a *p*-value of less than 0.001.[4] Taken with the direct evidence for these two pairs, when the IA is tested and *not* rejected via the RLIM payment procedure, these results provide striking support for the Bipolar Hypothesis advanced earlier.

**Cubitt, Starmer and Sugden** [1998] focus exclusively on Common Consequence and Common Ratio pairs of pairs, across three sets of experiments.

In the first set of experiments they compare 1-in-1 choices with 1-in-3 choices. Their comparison rests on subjects *not* having extreme risk-loving preferences over the other lotteries in the 1-in-3 treatment, but this is an *a priori* plausible assumption, and generally supported by their data. The two-sided *p*-values for these tests are 0.14 and 0.045, providing evidence against the application of the IA.

In the second set of experiments they compare 1-in-1 choices with 1-in-4 choices, with samples of 51 and 46 for the 1-in-1 treatments and 53 for the 1-in-4 treatment. Tests of the

---

[4] The same result occurs for the 1-in-4 pairs in this comparison.

hypothesis of the same choices in each leg of the Common Ratio pair of pairs have *p*-values of 0.84 and 0.16, implying that the direct test of the IA via the payment procedure had no significant effect. But in this case, in contrast to Beattie and Loomes [1997], there is no evidence for the Common Ratio effect in the 1-in-1 comparisons: the *p*-value on these choice patterns, spanning *both* legs of the Common Ratio pair of pairs, is 0.31. Of course, if EUT appears to be alive and well in terms of this familiar trip-wire test, then there is no *theoretical* expectation that the IA would be violated via the RLIM payment procedure.

In the third set of experiments with virtually the same Common Ratio pairs of pairs, and in fact the same lottery probabilities and prizes as the comparable Common Ratio pair of Beattie and Loomes [1997], they compare 1-in-1 choice patterns with 1-in-20 choice patterns. Sample sizes are 49 and 56 for the 1-in-1 treatments, and 97 for the 1-in-20 treatments. The direct tests of the IA via the RLIM procedure have *p*-values of 0.41 and 0.32, and the indirect Common Ratio test of the IA using the 1-in-1 treatment has a *p*-value of 0.10. At the risk of mixing psychiatric disorder metaphors, this is evidence for a Borderline Bipolar Hypothesis.

One common feature of virtually all of these studies is the use of a small value of K in the 1-in-K treatments. The rationale for this is explained by Cubitt, Starmer and Sugden [1998; p. 125]:

> First, as in the case of Experiments 1 and 2 [which used K=3 and K=4, respectively], we wanted to test the contamination hypothesis in a context in which we could expect the independence axiom to be violated. For Experiment 3 [which used K=20], however, we chose a somewhat different approach. In Experiments 1 and 2, the random lottery treatments involved only two tasks. In practical applications of the random lottery design, there are usually many tasks, and we wished to test the contamination hypothesis in such a setting. There are some reasons for expecting the extent of any bias in the random lottery design to depend on the number of tasks. On the one hand, it might be argued that, the more tasks there are in a random lottery experiment, the more likely subjects are to use the simplifying heuristic of treating each task in isolation. On the other hand, the more tasks there are, the more incentives are diluted; thus if bias is a product of dilution, its extent will increase with the number of tasks.

With the notable exception of the multiple price list design of Holt and Laury [2002], which uses

K=10, most applications of the RLIM do use large values of K. Hey and Orme [1994], for example, had K=100, Harrison and Rutström [2009] had K=60, Hey and Lee [2005a][2005b] have employed K=30 like us, and Wilcox [2010] and Hey [2001] bravely use K=300 and K=500, respectively, to obtain a rich data set for each individual subject.

Another common feature of all of these studies is that subjects were able to see all lotteries before having to make any choices. Starmer and Sugden [1998b] provided all lotteries in a booklet, and allowed subjects to make choices at any order they wanted. The specific lotteries of interest here were presented together on the same page of the booklet (their Figure 1, p. 975). Beattie and Loomes [1997; p.157] note that their 1-in-4 lotteries were "... presented together on a single sheet of paper." Cubitt, Starmer and Sugden [1998; p. 127] note that the software interface they used "... allowed subjects in all groups to backtrack at any point in the experiment, going back to previous tasks and changing their responses if they wished. After they had made all [...] responses they were reminded of this option. In this way [...] we gave subjects the opportunity to treat the whole experiment as a single decision problem if they so wished." They find that only one-third of subjects used this backtrack option, and they did not record if there were any changes in choices. Of course, the remaining two-thirds of subjects could still have viewed all tasks as one decision problem, making choices in later stages as a function of choices in earlier stages (e.g., "tend to pick safe options early, to ensure a certain payoff, then go for more risky options"). Cox, Sadiraj and Schmidt [2011] simply gave subjects all choices at the outset, each on one of K unbound sheets of paper, and then allowed them to enter choices on a computer interface that presented them sequentially.

Camerer [1989] studied the IA, literally in his design as an "afterthought." After subjects had made a number of choices using the RLIM, one was selected for payment, and the subject asked if he wanted to change the choice. Very few did, as Camerer [1989] notes:

Only two of 80 subjects did change. Therefore, either the independence axiom holds

> or subjects exhibit an isolation effect. Since the data below suggest that independence is often violated, we must conclude that there is an isolation effect. This is puzzling for theorists, but comforting for experimenters because it implies that allowing subjects to play some randomly chosen gambles generates meaningful responses for all gambles.

Our design provides a more direct test along these lines, without the 1-in-1 choice being an afterthought where the subject might feel compelled to stick with the initial choice rather than appear confused or capricious to the experimenter.

### Additional References


Burke, Michael S.; Carter, John R.; Gominiak, Robert D., and Ohl, Daniel F., "An Experimental Note on the Allais Paradox and Monetary Incentives," *Empirical Economics*, 21, 1996, 617-632.

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

## Appendix D: Econometric Model and Results

*A. The Basic Model*

Assume that the utility of income is defined by a completely non-parametric utility function. We exploit the fact that, by design, the lottery pairs in our experiment span only 5 monetary prize amounts, \$5, \$10, \$20, \$35 and \$70. Set the utility for the smallest prize to 0 and the utility of the largest prize to 1, and directly estimate the utility of the intermediate prizes:

$$U(\$0) = 0,\ U(\$10) = \varkappa_{10},\ U(\$20) = \varkappa_{20},\ U(\$35) = \varkappa_{35},\ U(\$70) = 1 \tag{1}$$

with the constraint that $\varkappa_{10}$ , $\varkappa_{20}$ and $\varkappa_{35}$ lie in the unit interval. This is precisely the approach employed by Hey and Orme [1994] and Wilcox [2010].

Let there be J possible outcomes in a lottery. The probability $p_i(M_j)$ of each outcome $M_j$ in lottery i is induced by the experimenter, so the expected utility (EU) of lottery i is simply the probability weighted utility of each outcome j:

$$EU_i = \sum_{j=1,J} [\ p_i(M_j) \times U(M_j)\ ]. \tag{2}$$

The EU for each lottery pair is calculated for candidate estimates of $\varkappa_{10}$ , $\varkappa_{20}$ and $\varkappa_{35}$, and the index

$$\nabla EU = EU_R - EU_L \tag{3}$$

calculated, where $EU_L$ is the "left" lottery and $EU_R$ is the "right" lottery of a given lottery pair as presented to subjects. The latent index $\nabla EU$, based on latent preferences, is then linked to observed choices using a standard cumulative normal distribution function $\Phi(\nabla EU)$. This "probit" function takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the probit link function,

$$\text{prob(choose lottery R)} = \Phi(\nabla EU) \tag{4}$$

The logistic function is similar and leads instead to the "logit" specification.

Thus the likelihood of the observed responses, conditional on the EUT specification being true, depends on the estimates of $\varkappa_{10}$ , $\varkappa_{20}$ and $\varkappa_{35}$ given the above statistical specification and the observed choices. The "statistical specification" here includes assuming some functional form for the cumulative density function (CDF). The conditional log-likelihood is then

$$\ln L(\varkappa_{10}, \varkappa_{20}, \varkappa_{35}; y, \mathbf{X}) = \sum_i [ (\ln \Phi(\nabla EU) \times \mathbf{I}(y_i = 1)) + (\ln (1-\Phi(\nabla EU)) \times \mathbf{I}(y_i = -1)) ] \tag{5}$$

where $\mathbf{I}(\cdot)$ is the indicator function, $y_i = 1(-1)$ denotes the choice of the Option R (L) lottery in risk aversion task i, and **X** is a vector of individual characteristics reflecting age, sex, race, and so on.

It is a simple matter to generalize this analysis to allow the core parameters $\varkappa_{10}$, $\varkappa_{20}$ and $\varkappa_{35}$ to each be a linear function of observable characteristics of the individual or task. We would then extend the model to allow $\varkappa_{10}$, for example, to be $\varkappa_{10} + \theta \times \mathbf{X}$, where $\varkappa_{10}$ is a fixed parameter and $\theta$ is a vector of effects associated with each characteristic in the variable vector **X**. In effect the unconditional model just estimates $\varkappa_{10}$ and assumes implicitly that $\theta$ is a vector of zeroes. This extension significantly enhances the attraction of structural maximum likelihood (ML) estimation, particularly for responses pooled over different subjects. Characterizing heterogeneous preferences is a central issue here because of treatment A, since one can then condition estimates of preferences on observable characteristics of the task or subject.

Harrison and Rutström [2008; Appendix F] review procedures that can be used to estimate structural models of this kind, as well as more complex non-EUT models. The goal is to illustrate how experimental economists can write explicit ML routines that are specific to different structural choice models. It is a simple matter to correct for multiple responses from the same subject ("clustering"), or heteroskedasticity, as needed.

*B. Behavioral Errors*

An important extension of the core structural model is to allow for subjects to make some behavioral errors. We employ a Fechner error specification, popularized by Hey and Orme [1994], that posits the latent index

$$\nabla EU = (EU_R - EU_L)/\mu \tag{3'}$$

instead of (3). In this specification μ is a structural "noise parameter" used to allow some errors from the perspective of the deterministic EUT model.[5] The index ∇EU is in the form of a cumulative probability distribution function defined over differences in the EU of the two lotteries and the noise parameter μ. Thus, as $\mu \to 0$ this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries; but as μ gets larger and larger the choice essentially becomes random. When $\mu=1$ this specification collapses to (3). Thus μ can be viewed as a parameter that flattens out the link function in (4) as μ gets larger.

An important contribution to the characterization of behavioral errors is the "contextual error" specification proposed by Wilcox [2011]. It is designed to allow robust inferences about the primitive "more stochastically risk averse than," and consistent inferences when one estimates over prize contexts in order to get better estimates. It posits the latent index

$$\nabla EU = ((EU_R - EU_L)/\nu)/\mu \tag{3''}$$

instead of (3′), where ν is a normalizing term for each lottery pair L and R. The normalizing term ν is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair. The value of ν varies, in principle, from lottery choice to lottery choice: hence it is said to be "contextual." For the Fechner error specification, dividing by ν ensures that the *normalized* EU difference $[(EU_R - EU_L)/\nu]$ remains in the unit interval. Our utility normalization (1) automatically ensures that the EU difference remains in the unit interval, but later specifications relax that, and normalization is needed then.

---

[5] This is just one of several different types of error story that could be used, and Wilcox [2008] provides a review of the implications of the alternatives. Some specifications place the error at the final choice between one lottery or after the subject has decided which one has the higher expected utility; some place the error earlier, on the comparison of preferences leading to the choice; and some place the error even earlier, on the determination of the expected utility of each lottery.

*C. Rank-Dependent Model*

The RDU model extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same non-parametric specification (1) considered for EUT. To calculate decision weights under RDU one replaces expected utility defined by (2) with RDU

$$RDU_i = \sum_{j=1,J} [\ w(p_i(M_j)) \times U(M_j)\ ] = \sum_{j=1,J} [\ w_{ij} \times U(M_j)\ ] \tag{2′}$$

where

$$w_{ij} = \omega(p_{ij} + ... + p_{iJ}) - \omega(p_{i,\,j+1} + ... + p_{iJ}) \tag{6a}$$

for j=1,... , J-1, and

$$w_j = \omega(p_{ij}) \tag{6b}$$

for j=J, with the subscript j ranking outcomes from worst to best, and ω(·) is some probability weighting function.

We could adopt the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ:

$$\omega(p) = p^{\gamma} \tag{7}$$

So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function is said to reflect "pessimism." If one assumes for simplicity a linear utility function, this implies a risk premium: since $\omega(p) < p \ \ \forall p$, the "RDU EV" in which monetary prizes are weighted by ω(p) instead of p has to be less than the EV weighted by p. Hence the certainty-equivalent under RDU has to be less than the true EV.

We use instead a non-parametric specification of the probability weighting function which exploits the fact that our main lottery parameters only use the 5 probabilities, 0, ¼, ½, ¾ and 1. If we constrain the extremes to have weight 0 and 1, we then have

$$\omega(0) = 0,\ \omega(1/4) = \varphi_{1/4}\ ,\ \omega(1/2) = \varphi_{1/2}\ ,\ \omega(3/4) = \varphi_{3/4}\ \text{ and } \omega(1) = 1 \tag{8}$$

and directly estimate $\varphi_{1/4}$ , $\varphi_{1/2}$ and $\varphi_{3/4}$ with the constraint that each lie in the unit interval. Note that the values ¼, ½ and ¾ refer to cumulative probabilities, consistent with (6a) and (6b). This is the approach employed by Gonzalez and Wu [1996] and Wilcox [2010]. The rest of the ML specification for the RDU model is identical to the specification for the EUT model, but with different and additional parameters to estimate.

*D. Non-Parametric Estimates Assuming Preference Homogeneity*

Baseline Estimates

Start with non-parametric estimates of the EUT and RDU models in the payoff environment that does not assume IA: the 1-in-1 treatment A. Of course, EUT assumes IA, so EUT estimates under payoff environments that require IA, such as the 1-in-30 treatment B, will also be theoretically consistent with EUT estimates from treatment A. But the estimates for RDU will not generally be theoretically consistent unless we use the 1-in-1 payoff environment. So the estimates in Table D1 provide the first estimates, to the best of our knowledge, of RDU when those estimates are not contaminated by having to assume the IA in the form of the Bipolar Behavioral Hypothesis. The estimates also provide the basis for testing our main hypothesis: that risk preferences estimated under EUT or RDU change when one moves away from payoff environments that assume the IA to be valid. Of course, as stressed earlier, the "bad news" theoretically is that one must make an assumption of homogeneous preferences across individuals to interpret these estimates as reflecting risk preferences. Popular as that assumption is, we can and will relax it.

Panel A in Table D1 shows the EUT estimates for each interior prize. The point estimates are increasing in the prize value, consistent with non-satiation, so $\partial U(x)/\partial x > 0$. The 95% confidence intervals are generally tight, in the sense of allowing one to rule out the hypothesis that these estimates are statistically indistinguishable from 0 or 1.[6] They also suggest that the estimates satisfy non-satiation even when one allows for sampling error. For example, the 95% confidence interval for the U($10) estimate is between 0.05 and 0.26, and the 95% confidence interval for the U($20) estimate is between 0.33 and 0.55, so there is no overlap. There is some slight overlap between the 95% confidence interval for U($20) and the interval for U($35), which is between 0.49 and 0.78. The statistical significance of this overlap is tested directly in the next two lines with $\Delta U_{20 \leftrightarrow 35}$, which is the difference in the utilities: if this is positive, and statistically significantly different from zero, as it is, then we can be confident that these estimates satisfy non-satiation. The same is true of the increment from U($10) to U($20), shown by $\Delta U_{10 \leftrightarrow 20}$.

We also directly test for diminishing marginal utility, $\partial^2 U(x)/\partial x^2 < 0$, by evaluating the marginal utility of each increment in utility, and then seeing if the difference between the first and second marginal utility is positive. The estimates show that each of the marginal utilities is positive, as one would expect from the non-satiation results, and that there is evidence of statistically significant diminishing marginal utility.

Not surprisingly, the RDU estimation in Panel B of Table D1 has an aggregate log-likelihood that is better than the EUT alternative. The most interesting feature of these estimates is the striking role of diminishing marginal utility and the minor role of probability weighting. The estimated probability weights for the ¼, ½ and ¾ probabilities are 0.29, 0.37 and 0.68, respectively, and in each case the 95% confidence interval includes the true probability. The 95% confidence interval for $\varphi_{1/2}$ is between 0.17 and 0.58, and overlaps with the 95% confidence interval for $\varphi_{1/4}$. In fact, the increase of

[6] In a numerical sense this might not be surprising, since we estimate these parameters by using a non-linear transform that ensures that they lie in the unit interval, as theory suggests. But it is still possible for the sampling errors to be large enough that the 95% confidence intervals get very close to 0 or 1, and as a practical matter for finite samples this can occur. The "delta method" is used to infer standard errors from non-linear transformations of this kind (Oehlert [1992]), and it includes some approximation error which can be particularly noticeable when point estimates are close to the boundary.

8.5 percentage points from $\varphi_{1/4}$ to $\varphi_{1/2}$ has a *p*-value of 0.151; although a one-sided hypothesis test would be appropriate here, given our prior of an increasing probability weighting function, this two-sided test still implies a *p*-value of 0.075. A $\chi^2$ test of the hypothesis that all three of these estimated probability weights are equal to the corresponding probability has a *p*-value of 0.037, implying that *there is evidence of statistically significant probability weighting.* The estimated utility function under RDU exhibits the familiar properties of non-satiation and diminishing marginal utility. Again, these conclusions are all under the maintained assumption of preference homogeneity across subjects.

The Effect of Being Bipolar

These estimates provide the baseline for evaluating the effect of the 1-in-30 payoff treatment on risk preferences. Table D2 shows more estimates, again assuming that risk preferences are homogeneous across individuals. In this case we employ all of the data from treatments A and B, and include binary dummy variables to specify the payoff procedure. The first three lines in Panel A of Table D2 show estimates of $\varkappa_{10}$ , $\varkappa_{10}^{pay1}$ and $\varkappa_{10}^{ra\_idr}$ from $U(\$10) = \varkappa_{10} + \varkappa_{10}^{pay1} \times \mathbf{pay1} + \varkappa_{10}^{ra\_idr} \times \mathbf{ra\_idr}$, where **pay1** is a binary dummy variable for the 1-in-1 treatment A and **ra_idr** is a binary dummy variable for the portion of observations in the 1-in-30 treatment B in which there was an additional, salient, individual discount rate elicitation task after the lottery choices. In each case we show the marginal effect of the binary variable, so we see that $U(\$10) = 0.201 - 0.071 \times \mathbf{pay1} + 0.036 \times \mathbf{ra\_idr}$.

Panel A of Table D2 presents the EUT estimates. We find no statistically significant effect of the individual treatments on the estimated utility values under EUT. A test that **pay1** has no joint effect across all three structural parameters has a *p*-value of only 0.78. In one respect this is just comforting, and not “news,” since EUT assumes the IA and the IA is what makes treatment B formally the same as treatment A.

The RDU results are more interesting, since Table D1 suggested that there was evidence for probability weighting overall, and that the IA axiom was therefore significantly violated. If the IA is significantly violated, then we might expect to see different risk preferences under RDU when we merge in the 1-in-30 choices. This is indeed what we see with the RDU estimates in Panel B of Table D2, although it is not obvious from examination of the *individual* significance levels. However, a $\chi^2$ test indicates that the central treatment dummy **pay1** is a significant factor across all estimated coefficients, with a *p*-value of 0.078. So we *do see a statistically significant effect of the payoff treatment on elicited preferences under RDU*. Again, however, we stress that this is still under the maintained assumption of preference homogeneity across subjects. It is time to relax that assumption and re-evaluate the inferences about the payoff treatments.

*E. Non-Parametric Estimates Allowing Preference Heterogeneity*

We extend the estimation to include a set of observable characteristics of the individual, and employ a series of binary variables: **female** is 1 for women, and 0 otherwise; **freshman**, **sophomore**, and **senior** are 1 for whether that was the current stage of undergraduate education at GSU, and 0 otherwise; **asian** and **white** are 1 based on self-reported ethnic status, and 0 otherwise; and **gpaVHI** is 1 for those reporting a cumulative GPA between 3.5 and 4.0 (mostly A's), and 0 otherwise. The demographic characteristics as a whole are statistically significant for both models.[7]

We find that *allowing for subject heterogeneity does not change the inferences about risk preferences under EUT*. Again, this is expected, given that the 1-in-30 treatments should theoretically have no effect on elicited risk preferences if the IA holds, and EUT assumes the IA. A $\chi^2$ test of the significance of the treatment variables **pay1** across all structural parameters has a *p*-value of 0.81, confirming that conclusion. Figure D1 illustrates the predicted values of utility across all subjects and treatments,

[7] For the EUT model a $\chi^2$ test on this hypothesis has a *p*-value of 0.046. For the RDU model the *p*-value is 0.003 for the utility parameters and 0.006 for the probability weighting parameters (and less than 0.001 for all parameters).

using the estimated model with demographic heterogeneity to generate these predictions.[8]

With the RDU model we find that the **pay1** variable has significant individual effects on some probability weights and utility parameters. The *p*-values for U($10), U($20), U($35), ω(¼), ω(½) and ω(¾), respectively, are 0.11, 0.08, 0.12, 0.54, 0.05 and 0.16. Moreover, we *find a significant overall joint effect from the 1-in-1 treatment on all probability weights parameters.* A $\chi^2$ test on the hypothesis that the **pay1** treatment dummy has no effect on all three probability weights *and* all three utility parameters can only be rejected with a *p*-value of 0.15, but when looking at the joint effect on probability weights the *p*-value is 0.07. Figure D2 illustrates the predicted probability weights generated from the full model, with heterogeneity and all treatments, underlying these estimates.

In summary, and allowing for observable heterogeneity in preferences, we conclude that

- there is no evidence that estimated EUT preferences are affected by the two experimental payment protocols employed; and
- there is evidence that estimated RDU preferences are also affected by the use of an experimental payment protocol that requires the validity of the IA.

These results imply that the Bipolar Behaviorist is in urgent need of medication. It is not possible to simultaneously maintain that the IA is invalid in the latent specification of choices over pairs of lotteries, and that the IA is magically valid when paying subjects for more than one choice. We often hear the "isolation effect" invoked to allow this discord to stand, as noted earlier, but we have not seen that effect stated in a formal manner that explains how it differs from the IA. It is used in scientific rhetoric more in the manner of a behavioral "get out of jail free card" in the parlor game *Monopoly*.

[8] These predictions reflect the point estimates of the model allowing for treatment effects and demographic heterogeneity, and not the sampling errors. Formal hypothesis tests take those sampling errors into account.

*F. Parametric Estimates*

We employ familiar specifications for the parametric utility and probability weighting functions. Instead of (1) for the utility function, we use the Expo-Power (EP) utility function proposed by Saha [1993]. Following Holt and Laury [2002], the EP function can be defined as

$$U(x) = [1-\exp(-\alpha x^{1-r})]/\alpha, \tag{9}$$

where $\alpha$ and r are parameters to be estimated. RRA is then $r + \alpha(1-r)x^{1-r}$, so RRA varies with income x if $\alpha \neq 0$. This function nests CRRA (as $\alpha \rightarrow 0$) and CARA (as $r \rightarrow 0$), so can be unbounded or bounded depending on particular parameter values. Instead of (8) for the probability weighting function, we employ the flexible two-parameter Prelec [1998] function,

$$w(p) = \exp\{-\eta(-\ln p)^{\varphi}\}, \tag{10}$$

which is defined for $0<p<1$, $\eta>0$ and $0<\varphi<1$. When $\varphi=1$ this function collapses to the venerable power function $\omega(p) = p^{\eta}$ defined earlier by (7).

For the EUT model, the joint hypothesis that the **pay1** treatment dummies on the structural coefficients r and $\alpha$ are equal to zero cannot be rejected, with a *p*-value of 0.98. This confirms our earlier finding that under EUT there is no statistically significant difference in elicited risk preferences across the payment protocols.

For RDU the joint hypothesis that the **pay1** treatment dummies on the structural coefficients r, $\alpha$, $\eta$ and $\varphi$ are all equal to zero can be rejected with a *p*-value of 0.06 even with controls for demographic heterogeneity. It is noteworthy that, consistent with the non-parametric findings, the culprit is the probability weighting parameters: the *p*-values for the r, $\alpha$, $\eta$ and $\varphi$ coefficients of **pay1** *alone* are 0.43, 0.31, 0.74 and less than 0.07, respectively, when we control for demographic heterogeneity and all treatment effects. Figure 2 in the main text shows the effects of moving from the 1-in-1 payment protocol to the 1-in-K payment protocols for the RDU model, assuming heterogeneous preferences across all subjects. The differences are striking, quantitatively *and* qualitatively.

**Table D1: Non-Parametric Estimates Assuming Homogeneity and 1-in-1 Choices**

| Parameter | Point Estimate | Standard Error | *p*-value | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | *A. Expected Utility Theory* (LL = -38.2) | | | |
| $\varkappa_{10}$ | 0.155 | 0.052 | 0.003 | 0.053 | 0.256 |
| $\varkappa_{20}$ | 0.438 | 0.057 | <0.001 | 0.326 | 0.549 |
| $\varkappa_{35}$ | 0.630 | 0.074 | <0.001 | 0.486 | 0.775 |
| $\Delta U_{10 \leftrightarrow 20}$ | 0.283 | 0.051 | <0.001 | 0.183 | 0.383 |
| $\Delta U_{20 \leftrightarrow 35}$ | 0.193 | 0.051 | <0.001 | 0.092 | 0.293 |
| $\Delta U_{10 \leftrightarrow 20} \div 10$ | 0.028 | 0.005 | <0.001 | 0.018 | 0.038 |
| $\Delta U_{20 \leftrightarrow 35} \div 15$ | 0.013 | 0.003 | <0.001 | 0.006 | 0.020 |
| $\partial^2 U(x)/\partial x^2$ | 0.015 | 0.006 | 0.017 | 0.003 | 0.028 |
| | | *B. Rank-Dependent Utility Theory* (LL = -34.9) | | | |
| $\varkappa_{10}$ | 0.161 | 0.082 | 0.048 | 0.001 | 0.321 |
| $\varkappa_{20}$ | 0.379 | 0.124 | 0.002 | 0.136 | 0.622 |
| $\varkappa_{35}$ | 0.584 | 0.121 | <0.001 | 0.346 | 0.821 |
| $\Delta U_{10 \leftrightarrow 20}$ | 0.218 | 0.054 | <0.001 | 0.112 | 0.324 |
| $\Delta U_{20 \leftrightarrow 35}$ | 0.205 | 0.042 | <0.001 | 0.122 | 0.287 |
| $\Delta U_{10 \leftrightarrow 20} \div 10$ | 0.022 | 0.005 | <0.001 | 0.011 | 0.032 |
| $\Delta U_{20 \leftrightarrow 35} \div 15$ | 0.014 | 0.003 | <0.001 | 0.008 | 0.019 |
| $\partial^2 U(x)/\partial x^2$ | 0.008 | 0.007 | 0.220 | -0.005 | 0.021 |
| $\varphi_{1/4}$ | 0.288 | 0.084 | 0.001 | 0.123 | 0.453 |
| $\varphi_{1/2}$ | 0.373 | 0.106 | <0.001 | 0.166 | 0.580 |
| $\varphi_{3/4}$ | 0.680 | 0.097 | <0.001 | 0.490 | 0.869 |
| $\Delta p_{1/4 \leftrightarrow 1/2}$ | 0.085 | 0.059 | 0.151 | -0.031 | 0.201 |
| $\Delta p_{1/2 \leftrightarrow 3/4}$ | 0.307 | 0.049 | <0.001 | 0.211 | 0.402 |
| $\varphi_{1/4} - 1/4$ | 0.038 | 0.084 | 0.652 | -0.127 | 0.203 |
| $\varphi_{1/2} - 1/2$ | -0.127 | 0.106 | 0.228 | -0.334 | 0.080 |
| $\varphi_{3/4} - 3/4$ | -0.070 | 0.097 | 0.467 | -0.260 | 0.119 |

**Table D2: Non-Parametric Estimates Assuming Homogeneity**

Data from treatments A and B

| Parameter | Point Estimate | Standard Error | *p*-value | 95% Confidence Interval | |
|---|---|---|---|---|---|
| *A. Expected Utility Theory* (LL = -3756.4) | | | | | |
| $\varkappa_{10}$ constant | 0.201 | 0.029 | <0.001 | 0.144 | 0.259 |
| $\varkappa_{10}$ pay1 | -0.071 | 0.065 | 0.273 | -0.198 | 0.056 |
| $\varkappa_{10}$ ra_idr | 0.036 | 0.033 | 0.269 | -0.028 | 0.101 |
| $\varkappa_{20}$ constant | 0.441 | 0.041 | <0.001 | 0.360 | 0.520 |
| $\varkappa_{20}$ pay1 | -0.019 | 0.075 | 0.801 | -0.167 | 0.129 |
| $\varkappa_{20}$ ra_idr | 0.051 | 0.045 | 0.256 | -0.037 | 0.140 |
| $\varkappa_{35}$ constant | 0.639 | 0.036 | <0.001 | 0.569 | 0.709 |
| $\varkappa_{35}$ pay1 | -0.031 | 0.089 | 0.728 | -0.207 | 0.144 |
| $\varkappa_{35}$ ra_idr | 0.017 | 0.039 | 0.669 | -0.061 | 0.094 |
| *B. Rank-Dependent Utility Theory* (LL = -3718.69 | | | | | |
| $\varkappa_{10}$ constant | 0.201 | 0.035 | <0.001 | 0.132 | 0.268 |
| $\varkappa_{10}$ pay1 | -0.082 | 0.062 | 0.186 | -0.205 | 0.040 |
| $\varkappa_{10}$ ra_idr | 0.034 | 0.042 | 0.426 | -0.049 | 0.116 |
| $\varkappa_{20}$ constant | 0.445 | 0.051 | <0.001 | 0.346 | 0.545 |
| $\varkappa_{20}$ pay1 | -0.088 | 0.096 | 0.357 | -0.275 | 0.099 |
| $\varkappa_{20}$ ra_idr | 0.045 | 0.059 | 0.446 | -0.071 | 0.161 |
| $\varkappa_{35}$ constant | 0.643 | 0.040 | <0.001 | 0.564 | 0.722 |
| $\varkappa_{35}$ pay1 | -0.108 | 0.109 | 0.321 | -0.322 | 0.105 |
| $\varkappa_{35}$ ra_idr | 0.008 | 0.047 | 0.870 | -0.085 | 0.101 |
| $\varphi_{1/4}$ constant | 0.185 | 0.048 | <0.001 | 0.091 | 0.280 |
| $\varphi_{1/4}$ pay1 | 0.125 | 0.089 | 0.161 | -0.050 | 0.300 |
| $\varphi_{1/4}$ ra_idr | 0.021 | 0.055 | 0.703 | -0.087 | 0.129 |
| $\varphi_{1/2}$ constant | 0.555 | 0.047 | <0.001 | 0.463 | 0.646 |
| $\varphi_{1/2}$ pay1 | -0.157 | 0.099 | 0.113 | -0.351 | 0.037 |
| $\varphi_{1/2}$ ra_idr | -0.034 | 0.055 | 0.542 | -0.142 | 0.075 |
| $\varphi_{3/4}$ constant | 0.777 | 0.046 | <0.001 | 0.687 | 0.868 |
| $\varphi_{3/4}$ pay1 | -0.060 | 0.100 | 0.557 | -0.256 | 0.138 |
| $\varphi_{3/4}$ ra_idr | -0.015 | 0.053 | 0.767 | -0.119 | 0.088 |

## References

Andersen, Steffen; Harrison, Glenn W.; Hole, Arne Rise, Lau, Morten I., and Rutström, E. Elisabet, "Non-Linear Mixed Logit," *Theory and Decision*, 73, 2012, 77-96.

Bade, Sophie, "Independent Randomization Devices and the Elicitation of Ambiguity Averse Preferences," *Working Paper*, Max Planck Institute for Research on Collective Goods, Bonn, 2011.

Beattie, J., and Loomes, Graham, "The Impact of Incentives Upon Risky Choice Experiments," *Journal of Risk and Uncertainty*, 14, 1997, 149-162.

Binmore, Ken, *Does Game Theory Work? The Bargaining Challenge* (Cambridge, MA: MIT Press, 2007).

Binswanger, Hans P., "Attitudes Toward Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, August 1980, 395-407.

Camerer, Colin F., "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 1989, 61-104.

Camerer, Colin F., "Recent Tests of Generalizations of Expected Utility Theory," in W. Edwards (ed.), *Utility Theories: Measurements and Applications* (Boston: Kluwer, 1992).

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

Conlisk, John, "Three Variants on the Allais Example," *American Economic Review*, 79(3), June 1989, 392-407.

Cubitt, Robin P.; Starmer, Chris, and Sugden, Robert, "On the Validity of the Random Lottery Incentive System," *Experimental Economics*, 1(2), 1998, 115-131.

Cox, James C.; Sadiraj, Vjollca, and Schmidt, Ulrich, "Paradoxes and Mechanisms for Choice under Risk," *Working Paper 2011-12*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2011.

Epstein, Larry, "A Paradox for the 'Smooth Ambiguity Model of Preference'," *Econometrica*, 78, 2010, 2085-2099.

Gilboa, Itzhak, and Schmeidler, David, "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 1989, 141-153.

Gonzalez, Richard, and Wu, George, "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 1999, 129-166.

Grether, David M., and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69(4), September 1979, 623-638.

Harless, David W., and Camerer, Colin F., "The Predictive Utility of Generalized Expected Utility

Theories," *Econometrica*, 62(6), November 1994, 1251-1289.

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility *And* Prospect Theory: One Wedding and A Decent Funeral," *Experimental Economics*, 12(2), June 2009, 133-158.

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Holt, Charles A., "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, June 1986, 508-514.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Karni, Edi, and Safra, Zvi, "Preference Reversals and the Observability of Preferences by Experimental Methods," *Econometrica*, 55, 1987, 675-685.

Klibanoff, Peter; Marinacci, Massimo, and Mukerji, Sujoy, "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73(6), November 2005, 1849-1892.

Klibanoff, Peter; Marinacci, Massimo, and Mukerji, Sujoy, "On the Smooth Ambiguity Model: a Reply," *Econometrica*, 80, 1303-1323.

Loomes, Graham, and Sugden, Robert, "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 1998, 581-598.

Oehlert, Gary W., "A Note on the Delta Method," *The American Statistician*, 46(1), February 1992, 27-29.

Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66, 1998, 497-527.

Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.

Saha, Atanu, "Expo-Power Utility: A Flexible Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75(4), November 1993, 905-913.

Samuelson, Paul A., "Probability, Utility, and the Independence Axiom," *Econometrica*, 20, 1952, 670-678.

Schmidt, Ulrich, and Zank, Horst, "A Simple Model of Cumulative Prospect Theory," *Journal of Mathematical Economics*, 45(3-4), March 2009, 308–319.

Segal, Uzi, "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28, 1987, 145-154.

Segal, Uzi, "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review*, 78(1), March 1988, 233-236.

Segal, Uzi, "Two-Stage Lotteries Without the Reduction Axiom," *Econometrica*, 58(2), March 1990, 349-377.

Segal, Uzi, "The Independence Axiom Versus the Reduction Axiom: Must We Have Both?" in W. Edwards (ed.), *Utility Theories: Measurements and Applications* (Boston: Kluwer Academic Publishers, 1992).

Starmer, Chris, "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect," *Review of Economic Studies*, 59, 1992, 813-830.

Starmer, Chris, and Sugden, Robert, "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review*, 81, 1991, 971-978.

Wakker, Peter P.; Erev, Ido, and Weber, Elke U., "Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories," *Journal of Risk and Uncertainty*, 9, 1994, 195-230.

Wilcox, Nathaniel T., "Lottery Choice: Incentives, Complexity, and Decision Time," *Economic Journal*, 103, 1993, 1397-1417.

Wilcox, Nathaniel T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison," in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Wilcox, Nathaniel T., "A Comparison of Three Probabilistic Models of Binary Discrete Choice Under Risk," *Working Paper*, Economic Science Institute, Chapman University, March 2010.

Wilcox, Nathaniel T., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 162(1), May 2011, 89-104.

Yaari, Menahem E., "The Dual Theory of Choice under Risk," *Econometrica*, 55(1), 1987, 95-115.