Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence

by

Glenn W. Harrison, Jimmy Martínez-Correa and J. Todd Swarthout †

July 2015

ABSTRACT.

The reduction of compound lotteries axiom (ROCL) has assumed a central role in the evaluation of behavior towards risk and uncertainty. We present experimental evidence on its validity in the domain of objective probabilities. Our battery of lottery pairs includes simple one-stage lotteries, two-stages compound lotteries, and their actuarially-equivalent one-stage lotteries. We find *violations of ROCL* and that behavior is better characterized by a source-dependent version of the Rank-Dependent Utility model rather than Expected Utility Theory. Since we use the popular "1-in-K" random lottery incentive mechanism payment procedure in our main test, our experiment explicitly recognizes the impact that this payment procedure may have on preferences. Thus we also collect data using the "1-in-1" payment procedure. We *do not infer any violations of ROCL* when subjects are only given one decision to make. These results are supported by both structural estimation of latent preferences as well as non-parametric analysis of choice patterns. The random lottery incentive mechanism, used as payment protocol, itself induces an additional layer of "compounding" by design that might create confounds in tests of ROCL. Therefore, we provide a word of caution for experimenters interested in studying ROCL for other purposes, such as the relationship between ambiguity attitudes and attitudes towards compound lotteries, to carefully think about the design to study ROCL, payment protocols and their interaction with the preferences being elicited.

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, Copenhagen Business School, Denmark (Martínez-Correa); and Department of Economics and Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout). Harrison is also affiliated with the School of Economics, University of Cape Town and IZA – Institute for the Study of Labor. E-mail contacts: gharrison@gsu.edu, jima.eco@cbs.dk and swarthout@gsu.edu. A working paper includes all appendices and can be obtained from http://cear.gsu.edu/working-papers/ as Working Paper 2012-05. We are grateful to the reviewers for helpful comments.

Table of Contents

The reduction of compound lotteries axiom (ROCL) has assumed a central role in the evaluation of behavior towards risk, uncertainty and ambiguity. We present experimental evidence on its validity in domains defined over objective probabilities, where the tests are as clean as possible.[1] Even in this setting, one has to pay close attention to the experimental payment protocols used and their interaction with the experimental task, so that one does not inadvertently introduce confounds that may contaminate hypothesis testing. Using the popular random lottery incentive mechanism (RLIM) we find violations of ROCL, but when RLIM is not used we find that behavior is consistent with ROCL.

We therefore show that a fundamental methodological problem with tests of the ROCL assumption is that one cannot use an incentive structure that may induce subjects to behave in a way that could be confounded with violations of ROCL. This means, in effect, that experimental tests of ROCL must be conducted with each subject making only one choice.[2] Apart from the expense and time of collecting data at such a pace, this also means that evaluations must be on a between-subjects basis, in turn implying the necessity of modeling assumptions about heterogeneity in behavior.

In sections 1 and 2 we define the theory and experimental tasks used to examine ROCL in the context of objective probabilities. In section 3 we present evidence from our experiment. We find *violations of ROCL*, and observed behavior is better characterized by the Rank-Dependent Utility model (RDU) rather than Expected Utility Theory (EUT). However, *violations of ROCL* only occur when *many* choices are given to each subject and RLIM is used as the payment protocol. We do *not infer any violations*

[1] The validity of ROCL over objective probabilities has also been identified as a potential indicator of attitudes towards uncertainty and ambiguity. Smith [1969] conjectured that people might have similar, source-dependent preferences over compound lotteries defined over objective probabilities *and* over ambiguous lotteries where the probabilities are not well-defined. Halevy [2007] provides experimental evidence that attitudes towards ambiguity and compound objective lotteries are indeed tightly associated. Abdellaoui, Klibanoff and Placido [2014] find that the latter relationship is weaker in their experiment.

[2] One alternative is to present the decision maker with several tasks at once and evaluate the portfolio chosen, or to present the decision maker with several tasks in sequence and account for wealth effects. Neither is attractive, since they each raise a number of (fascinating) theoretical confounds to the interpretation of observed behavior. One uninteresting alternative is not to pay the decision maker for the outcomes of the task.

of ROCL when subjects are each given only *one* decision to make. Section 4 draws conclusions for modeling, experimental design, and inference about decision making.

1. Theory

We start with a statement of some basic axioms used in models of decision-making under risk, and then discuss their implications for the experimental design. Our primary conclusion is the existence of an interaction of usual experimental payment protocols and the validity of ROCL. To understand how one can design theoretically clean tests of ROCL that do not run into confounds, we must state the axioms precisely.

A. Basic Axioms

Following Segal [1988][1990][1992], we distinguish between three axioms: the **Reduction of Compound Lotteries Axiom** (ROCL), the **Compound Independence Axiom** (CIA) and the **Mixture Independence Axiom** (MIA).

The ROCL states that a decision-maker is indifferent between a two-stage compound lottery and the actuarially-equivalent simple lottery in which the probabilities of the two stages of the compound lottery have been multiplied out. With notation to be used to state all axioms, let X, Y and Z denote simple lotteries, A and B denote two-stage compound lotteries, $\succ$ 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 in a first stage \$100 if a coin flip is a head and \$50 if the coin flip is a tail and in a second stage pays "double or nothing" of each possible outcome of the first stage with a 50:50 chance, then X would be the lottery that pays \$200 with probability $\frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$, \$100 with probability $\frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$, and nothing with probability ½.[3] To use the language of Samuelson [1952; p.671], a compound lottery generates a *compound income-probability-situation*, and its corresponding actuarially equivalent single-stage lottery defines an *associated income-probability-situation*, and that "...only algebra, not human behavior, is involved in this definition." From an observational perspective, one must then see choices between compound lotteries and actuarially-equivalent simple lotteries to test ROCL.

The CIA states that two compound lotteries, each formed from a simple lottery by adding a positive common lottery with the same probability, will exhibit the same preference ordering as the simple lotteries. In other words, the CIA states that if A is the compound lottery giving the simple lottery X with probability α and the simple lottery Z with probability $(1-\alpha)$, and B is the compound lottery giving the simple lottery Y with probability α and the simple lottery Z with probability $(1-\alpha)$, then $A \succ B$ iff $X \succ Y$ $\forall\ \alpha \in (0,1)$. 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.[4]

Finally, the MIA says that the preference ordering of two simple lotteries must be the same as the actuarially-equivalent simple lottery formed by adding a common outcome in a compound lottery of each of the simple lotteries, where the common outcome has the same value and the same (compound lottery) probability. More formally, the MIA says that $X \succ Y$ iff the actuarially-equivalent simple lottery of $\alpha X + (1-\alpha)Z$ is strictly preferred to the actuarially-equivalent simple lottery of $\alpha Y + (1-\alpha)Z$, $\forall\ \alpha \in (0,1)$. Stated so, it is clear the MIA strengthens the CIA by making a definite statement that the

---

[3] Formally, compound lottery A pays either $100 or $50 with equal chance in the first stage; in the second "double or nothing" stage it pays $200 or nothing with equal chance if the outcome of the first stage is $100, and pays $100 or nothing with equal chance if the outcome of the first stage is $50. This compound lottery reduces to a single-stage lottery X that pays $200, $100 or $0 with 25%, 25% and 50%, respectively.

[4] Segal [1992; p.170] defines the CIA by assuming that the second-stage lotteries are replaced by their certainty-equivalent, "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, by construction, since the informational bridge to that calculation has been burnt. The certainty-equivalent could have been generated by any model of decision making under risk, such as RDU or Prospect Theory.

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.

The reason these three axioms are important is that the failure of MIA does not imply the failure of the 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 the CIA might fail, only ROCL might fail, or both might fail. Because specific types of failures of ROCL lie at the heart of many important models of decision-making under uncertainty and ambiguity, it is critical to keep the axioms distinct as a theoretical *and* experimental matter.

B. Payment Protocols and Experimental Design

The choice of the payment protocol is critical to test ROCL. The RLIM payment protocol is the most popular payment protocol for individual choice experiments, and it assumes the validity of the CIA. RLIM entails the subject making K choices and then one of the K choices is 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 is made with a random draw from a 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 immediately the RLIM can be 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]).

If the K objects of choice in the experiment include any compound lotteries directly or indirectly, then RLIM requires the stronger MIA instead of just the CIA. Indeed, this was the setting for the classic discussions of Holt [1986], Karni and Safra [1987] and Segal [1988] on the interaction of the independence axiom with RLIM, which were motivated by the "preference reversal" findings of Grether and Plott [1979]. 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 CIA and ROCL, hence MIA. Holt [1986] and Karni and Safra [1987] showed that if CIA 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 CIA and transitivity was assumed, that one might also still observe choices that suggest "preference reversals." Again, the only reason that ROCL was implicated in these discussions is because the experimental task implicitly included choices over compound lotteries. In our experiment, we consider choices over simple lotteries and compound lotteries, so the validity of RLIM in the latter rests on the validity of the CIA and ROCL.

The need to assume the CIA or MIA can be avoided by setting K=1 and asking each subject to answer one binary choice task for payment, as advocated by Harrison and Swarthout [2014] and Cox, Sadiraj and Schmidt [2015]. Unfortunately, this comes at the cost of another 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 relying on randomization that will vary with sample size. Sadly, plausible estimates of the degree of heterogeneity in the typical population imply massive sample sizes for reasonable power, well beyond those of most experiments.

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. Lottery Parameters

We designed our battery of lotteries to allow for specific types of comparisons needed for testing ROCL. Beginning with a given *simple* (S) lottery and *compound* (C) lottery, we next create an *actuarially-equivalent* (AE) lottery from the C lottery, and then we construct three pairs of lotteries: a S-C pair, a S-AE pair, and an AE-C pair. By repeating this process 15 times, we create a battery of lotteries consisting of 15 S-C pairs shown in Table A2, 15 S-AE pairs shown in Table A3, and 10 AE-C pairs[5] shown in Table A4. Appendix A explains the logic behind the selection of these lotteries.

Figure 1 displays the coverage of lotteries in the Marschak-Machina triangle, combining all of the contexts used. Probabilities were drawn from {0, ¼, ½, ¾, 1}, and the final prizes from {$0, $10, $20, $35, $70}. We use a "double or nothing" (DON) procedure for creating compound lotteries. So, the first-stage prizes displayed in a compound lottery were drawn from {$5, $10, $17.50, $35}, and then the second-stage DON procedure yields the set of final prizes given above, which is either $0 or double the stakes of the first stage.

The majority of our compound lotteries use a conditional version of DON in the sense that the initial lottery will trigger the DON procedure only if a particular outcome is realized in the initial lottery. For example, consider the compound lottery formed by an initial lottery that pays $10 and $20 with equal probability and a subsequent DON lottery if the outcome of the initial lottery is $10, implying a payoff of $20 or $0 with equal chance *if the DON stage is reached*. If the initial lottery outcome is $20, there is no subsequent DON stage. The right panel of Figure 2 shows a tree representation of this compound lottery where the initial lottery is depicted in the first stage and the DON lottery is depicted in the second stage of the compound lottery. The left panel of Figure 2 shows the corresponding

[5] The lottery battery contains only 10 AE-C lottery pairs because some of the 15 S-C lottery pairs shared the same compound lottery.

actuarially-equivalent simple lottery which offers $20 with probability ¾ and $0 with probability ¼.

The conditional DON lottery allows us to obtain better coverage in the Marshak-Machina triangle than unconditional DON in terms of probabilities. If we used only the unconditional DON option, we would impose an *a priori* restriction within the Marschak-Machina triangle to lotteries that always assign 50% chance to getting $0.[6] We avoid this restriction by using conditional DON with 50:50 odds in the second stage, which allows us to construct lotteries that assign probabilities of getting $0 that need not necessarily be 50%.[7] The main reason for this design choice is that we want to allow for variation both in prizes and probability distributions so we can potentially identify source-dependent preferences that take into account attitudes towards variability in prizes and towards probabilities. As an example, one can potentially identify if an individual is an Expected Utility maximizer when faced with single-stage lotteries but distort probabilities optimistically when faced with a compound lottery, therefore increasing the relative attractiveness of compound lotteries over their actuarially-equivalent lotteries. If we restrict the choice lotteries to a smaller portion of the Marschak-Machina triangle, for instance, we might miss source-dependent attitudes that occur at low probability levels but not at higher probability levels.

B. Experimental Procedures

We implement two between-subjects treatments. We call one treatment "Pay 1-in-40" (1-in-40) and the other "Pay 1-in-1" (1-in-1). Table 1 summarizes our experimental design and the sample size of subjects and choices in each treatment.

---

[6] For instance, suppose a compound prospect with an initial lottery that pays positive amounts $X and $Y with probability p and (1-p), respectively, and offers DON in the second stage for either outcome in the first stage. The corresponding actuarially-equivalent lottery pays $2X, $2Y and $0 with probabilities p/2, (1-p)/2 and ½, respectively.

[7] For instance, suppose now a compound prospect with an initial lottery that pays positive amounts $X and $Y with probability p and (1-p), respectively, and offers DON if the outcome of the first stage is $X. The corresponding actuarially-equivalent lottery pays $2X, $2Y and $0 with probabilities p/2, 1-p and p/2, respectively.

In the 1-in-40 treatment, each subject faces choices over all 40 lottery pairs, with the order of the pairs randomly shuffled for each subject. After all choices have been made, one choice is randomly selected for payment using RLIM, with each choice having a 1-in-40 chance of being selected. The selected choice is then played out and the subject receives the realized monetary outcome, with no other rewarded tasks. This treatment is potentially different from the 1-in-1 treatment in the absence of ROCL, since RLIM induces a compound lottery consisting of a 1-in-40 chance for each of the 40 chosen lotteries to be selected for payment.

In the 1-in-1 treatment, each subject faces a single choice over two lotteries. The lottery pair presented to each subject is randomly selected from the battery of 40 lottery pairs. The lottery chosen by the subject is then played out and the subject receives the realized monetary outcome. There are no other rewarded tasks, before or after a subject's binary choice, that affect earnings. Further, there is no other activity that may contribute to learning about decision making in this context.

The general procedures during an experiment session were as follows. Upon arrival at the laboratory, each subject drew a number from a box which determined random seating position within the laboratory. After being seated and signing the informed consent document, subjects were given printed instructions and allowed sufficient time to read these instructions[8]. Once subjects had finished reading the instructions, an experimenter at the front of the room read aloud the instructions, word for word. Then the randomizing devices[9] were explained and projected onto the front screen and three large flat-screen televisions spread throughout the laboratory. The subjects were then presented with lottery choices, followed by a demographic questionnaire that did not affect final payoffs. Next, each subject was approached by an experimenter who provided dice for the subject to roll and determine her own

[8] Appendix C of the Working Paper provides complete subject instructions.

[9] Only physical randomizing devices were used, and these devices were demonstrated prior to any decisions. In the 1-in-40 treatment, two 10-sided dice were rolled by each subject until a number between 1 and 40 came up to select the relevant choice for payment. Subjects in both treatments would roll the two 10-sided dice (a second roll in the case of the 1-in-40 treatment) to determine the outcome of the chosen lottery.

payoff. If a DON stage was reached, a subject would flip a U.S. quarter dollar coin to determine the final outcome of the lottery. Finally, subjects left the laboratory and were privately paid their earnings: a $7.50 participation payment in addition to the monetary outcome of the realized lottery.

We used software to present lotteries to subjects and record their choices. Figure 3 shows an example of the subject display of an AE-C lottery pair.[10] The pie chart on the right of Figure 3 displays the first and second stages of the compound lottery as an initial lottery that has a DON stage identified by text. The pie chart on the left of Figure 3 shows the paired AE lottery. Figure 4 shows an example of the subject display of a S-C lottery pair, and Figure 5 shows an example of the subject display of a S-AE lottery pair.

C. Evaluation of Hypotheses

The 1-in-40 treatment adds an additional layer of compounding of choices that subjects do not face in the 1-in-1 treatment. If subjects in both treatments have the same risk preferences and behavior is consistent with ROCL, we should see the same pattern of decisions for comparable lottery pairs across the two treatments in spite of the additional layer of compounding in the 1-in-40 treatment. The same pattern should also be observed as one characterizes heterogeneity of individual preferences towards risk, although these inferences depend on the validity of the manner in which heterogeneity is modeled.

Nothing here assumes that behavior is characterized by EUT. The validity of EUT requires both ROCL and CIA. So when we say that risk preferences should be the same in the two treatments under ROCL, these are simply statements about the Arrow-Pratt risk premium, and not about how that is

[10] Decision screens were presented to subjects in color. Black borders were added to each pie slice in Figures 3, 4 and 5 to facilitate black-and-white viewing. In black and white these displays might make it appear that the $0 prize was shown in bold and the others not. This is an illusion; all prizes and probabilities were displayed equally, but with distinct colors.

decomposed into explanations that rely on diminishing marginal utility or probability weighting. For instance, the Rank-Dependant Utility model assumes ROCL but the risk premium of a compound lottery depends both on aversion towards variation in prizes (utility function) and attitudes towards probabilities. We later analyze the decomposition of the risk premium as well as the nature of any violation of ROCL.

Our method of evaluation is twofold. First, we estimate structural models of risk preferences and test if the risk preference parameters depend on whether a C or an AE lottery is being evaluated. This method does not assume EUT, and indeed we allow non-EUT specifications. We specify a source-dependent form of utility and probability weighting, and test for violations of ROCL by determining if the subjects evaluate simple and compound lotteries differently. We use a similar approach to Abdellaoui, Baillon, Placido and Wakker [2011], who studied source functions to model preferences towards different sources of uncertainty. They concluded that different probability weighting functions are used when subjects face risky processes with known probabilities versus uncertain processes with unknown probabilities. They call this "source dependence," where the notion of a source is relatively easy to identify in the context of a laboratory experiment, and hence provides the tightest test of this proposition. In our case, simple one-stage objective lotteries are one source of risk while objective compound lotteries constitute another source of risk. If individuals do perceive both as two different sources of risk then we should find evidence of source dependence and this is to be interpreted as a violation of ROCL.[11] We chose source-dependent models to study attitudes towards compound lotteries for strong theoretical reasons. Smith [1969] conjectured that attitudes towards ambiguity were connected to attitudes towards compound lotteries if subjects perceived ambiguous lotteries as

[11] Harrison [2011] shows that the conclusions in Abdellaoui, Baillon, Placido and Wakker [2011] are an artefact of estimation procedures that do not take account of sampling errors. A correct statistical analysis that does account for sampling errors provides no evidence for source dependence using their data. Of course, failure to reject a null hypothesis could just be due to samples that are too small.

compound lotteries. Recently, Halevy [2007] and Abdellaoui, Klibanoff and Placido [2014] have studied this relationship in experimental settings. Given this well-documented relationship, we believe that source-dependent models are natural candidates to study the validity of ROCL since they can accommodate distinct preferences for compound *and* simple single-stage lotteries.

Our second method of evaluation of ROCL uses non-parametric tests to evaluate the choice patterns of subjects. These constitute a robustness check of our parametric tests. Our experimental design allows us to evaluate ROCL using choice patterns in two ways: we examine choice patterns across the linked S-C and S-AE lottery pairs where ROCL predicts *consistent choices*, and we examine choice patterns in AE-C lottery pairs where ROCL predicts *indifference*.

In both of our methods of evaluation of ROCL, we use data from the 1-in-1 treatment and the 1-in-40 treatment which uses RLIM as the payment protocol. Of course, analysis of the data from the 1-in-40 treatment requires us to assume that CIA holds. However, by also analyzing choices from the 1-in-1 treatment we can test if the RLIM itself creates distortions that could be confounded with violations of ROCL, and in fact this turns out to be critical to the validity of ROCL.

D. Different Sample Sizes

One difference in the treatments is that every subject in the 1-in-40 task makes 40 times the number of choices of each subject in the 1-in-1 task. This is tautological, from our design, but might raise concerns when estimating treatment effects. We conduct some simple statistical and econometric checks to address these concerns.[12]

The simplest check is to sample *from* the sample of 1-in-40 choices to mimic the sample of 1-in-1 choices. Each bootstrap simulation draws one choice at random from the 40 choices of each subject in

[12] We certainly allow for correlated errors within the choices by each individual in the 1-in-40 tasks, which is a separate statistical issue.

the 1-in-40 treatment and estimates a structural model of the treatment effect for those data and the 1-in-1 choices. Since the order of lottery pairs presented to each subject in the 1-in-40 treatment was randomized, we can simply bootstrap from each of the 40 choices in sequence. In addition, we add in 460 further bootstrap replications in which one choice from each of the 1-in-40 subjects is sampled at random. The distribution of *p*-values over these 500 bootstrap draws will then reveal if the treatment effect is significant or not, without any concerns about differential sample sizes.

3. Evidence

We evaluate the evidence by estimating preferences as well as by examining choice patterns. Each approach has strengths and weaknesses. Evaluating choice patterns has the advantage of remaining agnostic about the particular model of decision making under risk. However, it has the disadvantage of not using all information embedded in the difference between the two lotteries. If one assumes an RDU model for illustration, it is intuitively clear that a deviation from RDU maximization should be more serious if the RDU difference is large than when it is minuscule. Simply counting the number of violations of predicted choice patterns, and ignoring the size of the deviation, ignores this information. Of course, to use that information one has to make some assumptions about what determines the probability of any predicted choice, and hence offer a metric for comparing the importance of deviations from risk neutrality.

A structural model of behavior, again using RDU for example, allows a more rigorous use of information on the size of errors from the perspective of the null hypothesis that ROCL is valid. For example, choices that are inconsistent with the null hypothesis but that involve statistically insignificant errors from the perspective of that hypothesis are not treated with the same weight as statistically significant errors.

An additional advantage of a structural model is that it is relatively easy to extend it to allow for varying degrees of heterogeneity of preferences, which is critical for between-subject tests with 1-in-1 data unless one is willing to maintain the unattractive assumption of homogeneous risk preferences. Given the importance of our treatment in which we study just one choice per subject, this ability to compare behavior from pooled choices across subjects, while still conditioning on some differences in subjects, is essential.

Again, we see these two ways of evaluating results as complementary. This is true even when they both come to the same general conclusion.

A. Estimated Risk Preferences

We estimate risk preferences assuming a RDU model of decision-making under risk.[13] The key issue is whether the structural risk parameters of the model differ when subjects evaluate simple lotteries or compound lotteries.

Assuming a Constant Relative Risk Aversion (CRRA) utility function, we allow the CRRA parameter r for simple lotteries and the parameter (r + rc) for compound lotteries, where rc captures the additive effect of evaluating a compound lottery. Hence the decision maker employs the utility function

$$U(x \mid \text{simple lottery}) = x^{(1-r)}/(1-r) \qquad (1)$$
$$U(x \mid \text{compound lottery}) = x^{(1-r-rc)}/(1-r-rc) \qquad (1')$$

where x is the monetary outcome of lotteries. The RDU model extends the EUT model by allowing for decision weights on lottery outcomes. Decision weights are calculated for each lottery outcome, using differences between rank-ordered cumulative probabilities generated from a probability weighting function. We adopt the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ for simple lotteries and γ+γc for compound lotteries:

$$\omega(p \mid \text{simple lottery}) = p^{\gamma} \qquad (2)$$
$$\omega(p \mid \text{compound lottery}) = p^{\gamma+\gamma c} \qquad (2')$$

[13] Appendix F of the Working Paper documents the basic RDU model.

where p is the is the probability of a given outcome of a lottery. EUT is the special case in which $\gamma = \gamma + \gamma c = 1$. Under RDU the hypothesis of source-independence, which is consistent with ROCL, is that $\gamma c = 0$ *and* $rc = 0$. We also consider the inverse-S probability weighting function given by:

$$\omega(\, p \mid \text{simple lottery}\,) = p^{\gamma}/(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma} \tag{3}$$

$$\omega(\, p \mid \text{compound lottery}\,) = p^{\gamma+\gamma c}/(p^{\gamma+\gamma c}+(1-p)^{\gamma+\gamma c})^{1/(\gamma+\gamma c)} \tag{3'}$$

We undertake non-nested specification tests to evaluate which probability weighting function is the best.

Specifying preferences in this manner provides us with a structural test for ROCL under an EUT source-dependent model, since the EUT source-dependent model is a special case of the RDU source-dependent model when there is no probability weighting. In this case, if $rc = 0$ then compound lotteries are evaluated identically to simple lotteries, which is consistent with ROCL. However, if $rc \neq 0$, then decision-makers violate ROCL in a certain source-dependent manner, where the "source" here is whether the lottery is simple or compound. This specification follows from Smith [1969], who proposed a similar source-dependent relationship between objective and subjective compound lotteries as an explanation for the Ellsberg Paradox. Of course, the linear specification $r + rc$ is a parametric convenience, but the obvious one to examine initially.

One of the reasons for wanting to estimate a structural econometric model is to have some controls for heterogeneity of preferences. We include the effects of allowing a series of binary demographic variables on a linear specification for each structural parameter: **female** is 1 for women, and 0 otherwise; **senior** is 1 for whether that was the current stage of undergraduate education, and 0 otherwise; **white** is 1 based on self-reported ethnic status; and **gpaHI** is 1 for those reporting a cumulative grade point average between 3.25 and 4.0, and 0 otherwise. Thus the structural parameters r, rc, γ and γc are each estimated as a linear function of a constant and these observable characteristics.

The complete econometric model is otherwise conventional, and written out in detail in Appendix D of the Working Paper.

In general we find that the EUT model is rejected in favor of the RDU model, whether one allows source-dependence or not. For that reason we will focus our evaluation of hypotheses on the RDU source-dependent model, with allowance for observable heterogeneity of risk preferences. However, one noteworthy result is that if we *incorrectly* assumed an EUT source-dependent model we would reject the ROCL assumption in the 1-in-1 treatment, with a *p*-value of 0.008. Of course, that assumption is invalid, demonstrating the importance of finding the correct specification of decision-making under risk, since rejections of the null hypothesis can be confounded with the wrong choice of preference representation in the parametric tests.

We do not find evidence of the source-dependence hypothesis with the 1-in-1 data, hence we cannot reject the ROCL hypothesis in that setting. Figure 6 shows the point estimates for utility and probability weighting functions, conditional on either the Power or Inverse-S specifications, with both performing comparably from an explanatory perspective. Formal hypothesis tests do *not* allow us to reject the hypothesis of the same risk preferences for simple and compound lotteries, with *p*-values of 0.72 and 0.95 for the Power and Inverse-S specifications respectively.[14]

However, when we turn to the 1-in-40 data we estimate very different risk preferences for simple and compound lotteries. Figure 7 shows the point estimates, and these are statistically significantly different with *p*-values less than 0.001. These estimates imply that an average subject exhibits different diminishing marginal utility and different probability weighting depending on whether he is evaluating a simple versus compound lottery. To illustrate the magnitude of this difference, consider the AE-C pair #37. Assume a hypothetical subject characterized by the average parameters estimated for the source-dependent RDU model.[15] Assuming the Power (Inverse-S) probability

[14] We draw the same qualitative conclusion if we assume homogeneous preferences, with *p*-values of 0.28 and 0.76 respectively.

[15] In these calculations we use the estimates of the homogenous preferences specification described in Appendix D of the Working Paper. The results are virtually the same if we used the heterogeneous preferences specification and the unconditional average of utility and probability weighting function parameter estimates.

weighting function, this subject would attach a certainty equivalent of \$29.6 (\$31.3) to the C lottery and of \$34 (\$37.1) to its corresponding AE lottery, a 14.7% (18.5%) difference which implies compound risk aversion.[16] However, heterogeneity implies that there could be different combinations of parameter values for utility and probability weighting functions at the individual level, and thus *revealed* attitudes can be of the compound risk loving type. In the choice pattern section below we explore in more detail the nature of the violations of ROCL in the 1-in-40 treatment.

Finally, we undertake 500 bootstrap simulations from the 1-in-40 data to check if using just one observation from the 40 choices of each subject makes any difference to our conclusions. As explained earlier, each bootstrap simulation draws one choice at random from the 40 that each subject in the 1-in-40 treatment made, estimates the implied test of ROCL using those 62 selected choices, and calculates the *p*-value for the test of ROCL. We use the RDU model for this purpose, since it generalizes the EUT model and is a better characterization for these data. Figure 8 displays the resulting bootstrap distributions, and confirms that our conclusions are not an artefact of unequal sample sizes in the 1-in-1 and 1-in-40 treatments. The figure shows the distribution of the *p*-values of the joint test of $\gamma c = rc = 0$ for both the Power and Inverse-S probability weighting functions. We note that most of the probability mass is peaked at the far left of each panel, conveying that most every iteration of the bootstrapping exercise resulted in a *p*-value near zero.[17] The median *p*-value for the Power case is less than 0.001, and median *p*-value for the Inverse-S case is 0.073. These simulations corroborate our earlier finding of source dependence in the 1-in-40 data, and hence a rejection of the ROCL hypothesis. If our earlier finding was indeed an artefact of unequal sample sizes, then the bootstrap exercise would have resulted

---

[16] The certainty equivalents are calculated, as usual, as the certain amount of money that makes an individual indifferent between receiving this certain amount and playing the lottery. We evaluate these certainty equivalents using the utility function for simple lotteries, since a sure amount of money is a simple single-stage lottery with no risk. Finally, the RDU of the AE and the C lotteries are estimated by using the utility and probability weighting functions for simple lotteries and for compound lotteries, respectively.

[17] For each of the two probability weighting functions, around 90 percent of the bootstrap simulations resulted in a *p*-value less than 0.1.

in much more diffuse graphs. In the next section we conduct a robustness check of our parametric results with non-parametric tests of ROCL and obtain similar results.

B. Evidence from Choice Patterns

To analyze the choice patterns, we test two hypotheses predicted by ROCL: consistency of choices when the compound lottery is replaced by its actuarially-equivalent simple lottery, and indifference between a compound lottery and an actuarially-equivalent simple lottery. This section summarizes our hypothesis tests of choice patterns. More detailed discussion of our statistical tests are presented in Appendix E of the Working Paper.

Beginning with the consistency hypothesis, we consider the following scenario. Suppose a subject is presented with a given S-C lottery pair, and further assume that she prefers the C lottery over the S lottery. If the subject satisfies ROCL and is also presented with a second pair of lotteries consisting of the same S lottery and the AE lottery of the previously-presented C lottery, then she would prefer and should choose the AE lottery. Similarly, of course, if she instead prefers the S lottery when presented separately with a given S-C lottery pair, then she should choose the S lottery when presented with the corresponding S-AE lottery pair. Therefore, ROCL is violated if we observe unequal proportions of S lottery choices across a S-C pair and its linked S-AE pair.

We *do* find evidence of violations of ROCL in the 1-in-40 treatment, while we do *not* find evidence in the choice data to reject the consistency hypothesis in the 1-in-1 treatment. Table E1 presents results of the Cochran Q test coupled with the Bonferroni-Dunn correction procedure to evaluate consistency of choices in the 1-in-40 treatment, and we see statistically significant evidence of inconsistent choices. Table E2 presents a Fisher Exact test for each of the comparisons in the 1-in-1 treatment with sufficient data for the test. Only 1 of the 11 tests results in statistically significant evidence of inconsistency. Further, we conduct a Cochran-Mantel-Haenszel test to jointly evaluate whether choices in the 1-in-1 treatment are consistent over all linked pairs, and we find that ROCL cannot be rejected (*p*-value = 0.122).[18]

Moving on to the indifference hypothesis, we *do* find statistical evidence of violations of ROCL in the 1-in-40 treatment, although we do *not* find statistical evidence to reject the ROCL prediction of indifference in the 1-in-1 treatment that controls for potential confounds.[19] Table E4 presents a Cochran Q test of the AE-C choices in the 1-in-40 treatment, and equiprobable choice is resoundingly rejected (*p*-value < 0.0001). In contrast, Table E5 presents choice data for all AE-C lottery pairs in the 1-in-1 treatment. Roughly 59% of subjects chose the C lottery, and a Fisher Exact test fails to reject the hypothesis of indifferent (i.e., equiprobable) choices (*p*-value = 0.342). Further, Table E6 reports an individual Binomial test of equiprobable choices for each AE-C pair in the 1-in-1 treatment, and every *p*-value is insignificant at any reasonable level of confidence.[20] Of course, the sample size is an issue here but we have already addressed this issue with the parametric tests.

We are also interested in studying the patterns of violations of ROCL and we can do that in the 1-in-40 treatment. A pattern inconsistent with ROCL would be when a subject chooses the S lottery

---

[18] One referee suggested running additional sessions where one would focus on pairs of two compound lotteries, say A and B. Then one could compare A with the actuarially-equivalent lottery of B, and B with the actuarially-equivalent lottery of A. This is an attractive extension of our design.

[19] We use an indifference test of ROCL since it is a natural one given that the definition of ROCL itself requires the indifference between a compound lottery and its actuarially-equivalent lottery. However, it is not easy to identify empirically when a subject is truly indifferent between two options. Thus we follow Starmer and Sugden [1991] and use an equiprobable non-parametric test for the basic indifference prediction of ROCL. According to Starmer and Sugden [1991, p. 976] if subjects are offered two lotteries that are equivalent in the ROCL sense then "there seems to be no reason to expect either of these responses [i.e., choosing one or the other lottery] to be more frequent than the other... we should expect the choice between these two responses to be made at random; as a result, these responses should have the same expected frequency. If, then, we were to find a significantly greater frequency of ...[one of the responses over the other], we should have found a pattern that was inconsistent with the reduction principle [i.e., ROCL]." Thus the prediction is that we should observe the compound lottery and its actuarially-equivalent lottery being chosen with equal proportions when pooling 1-in-1 choices across subjects. Only 10 of the 40 lottery pairs in our battery of lotteries is of the compound-actuarially equivalent type.

[20] Evidently, it is possible that we fail to reject the null hypothesis given that we have a small sample size in the indifference test of the 1-in-1 case. That is the reason why we also tested ROCL with the consistency test in which the sample size is bigger and with the structural estimation where we pool responses from both the consistency and the indifference tests.

when presented with a given S-C lottery pair, but switching to choose the AE lottery when presented with the matched S-AE pair. We construct a 2 × 2 contingency table for each given set of two matched lottery pairs that shows the number of subjects who exhibit each of the four possible choice patterns: (i) always choosing the S lottery; (ii) choosing the S lottery when presented with a S-C pair and switching to choose the AE lottery when presented with the matched S-AE pair; (iii) choosing the C lottery when presented with a S-C pair and switching to choose the S lottery when presented with the matched S-AE pair; and (iv) choosing the C lottery when presented with the S-C lottery and choosing the AE lottery when presented with the matched S-AE pair.

Since we have paired observations, we use the McNemar test to evaluate the null hypothesis of equiprobable occurrences of discordant choice patterns (ii) and (iii) within each set of matched pairs. We find a statistically significant difference in the number of (ii) and (iii) choice patterns within 4 of the 15 matched pairs. Table E3 reports the exact *p*-values for the McNemar test. The McNemar test results in *p*-values less than 0.05 in four comparisons: Pair 1 vs. Pair 16, Pair 3 vs. Pair 18, Pair 10 vs. Pair 25 and Pair 13 vs. Pair 28.[21] Moreover, the odds ratios of the McNemar tests suggest that the predominant switching pattern is choice pattern (iii): subjects tend to switch from the S lottery in the S-AE pair to the C lottery in the S-C pair.

To summarize, we find consistent evidence from the choice patterns, whether we look at predictions of indifference or predictions of consistent choice. The evidence implies a failure of ROCL for binary choice when one embeds these choices in a payment protocol that induces a further level of compounding.

C. Nature of the Violations of ROCL in the 1-in-40 Treatment

[21] These violations of ROCL are also supported by the B-D procedure if the family-wise error rate is set to 10%.

There are two possible types of violations of ROCL observable in our consistency tests: *compound risk loving* and *compound risk aversion*. For completeness, we define the former (latter) by revealed behavior of people choosing (avoiding) the compound lottery over a simple lottery when offered this binary choice, and choosing (avoiding) the same simple lottery over the AE of the compound lottery when offered this binary option.[22] Our battery of lotteries have 30 lottery pairs that comprise 15 tests of ROCL consistency applied to 62 subjects in the 1-in-40 treatment, for a total of 930 ROCL consistency tests in our experiment. Of those, 279 of those tests (30% of the total) revealed behavior inconsistent with ROCL , and we see that compound risk loving is the most common form of violation: 100 tests revealed compound risk aversion and 179 tests revealed compound risk loving.

Both the non-parametric and the parametric tests of ROCL provided evidence consistent with our definition of compound risk loving, although compound risk aversion is still present to a lesser extent. The McNemar test indicates that subjects violating ROCL in the 1-in-40 treatment tend to do so more frequently by choosing the S lottery in the S-AE pair and then switching to the C lottery when offered the S-C pair. Additionally, as depicted in Figure 7, the source-dependent RDU models with CRRA utility function and Power probability weighting function contains elements of both compound risk loving *and* compound risk aversion. It does this by assigning to compound lotteries a utility function that is more concave than the utility function used for simple single-stage lotteries (hence a tendency towards greater compound risk aversion), and by assigning to compound lotteries a probability weighting function that is consistent with probability optimism (hence a tendency towards greater compound risk loving). A similar pattern, although less obvious, can be seen with the source-dependent model that uses an Inverse-S probability weighting function.

---

[22] Compound risk aversion is consistent with discordant choice pattern (ii) of the McNemar Test, while compound risk loving is consistent with discordant choice pattern (iii). Under our definition, a person who satisfies ROCL would be compound risk neutral and should make consistent choices as defined in choice patterns (i) and (iv) of the McNemar test.

In order to examine transparently the *strength* of preferences of the typical subject in favor of or against compound risk in our experiment, we used econometric models that assume preference homogeneity to estimate the implied certainty equivalent (CE) of compound and actuarially-equivalent lotteries. Figure 9 shows the CE for the compound lottery and its actuarially-equivalent lottery for the four ROCL consistency tests for which the non-parametric tests indicated ROCL violations.

The CE calculations in Figure 9 show a pattern consistent with compound risk loving: the CE of the compound lottery in each test is greater than the CE of its paired actuarially-equivalent lottery. This can explain the frequent switching behavior of subjects that exhibited compound risk loving and reveals a non-trivial preference for compound lotteries. For instance, in the ROCL test that compares Pair 10 versus Pair 25, the source-dependent RDU model with Power probability weighting function estimates a CE of approximately US $15.5 for the compound lottery and a CE of US $12 for the respective actuarially equivalent lottery.[23] A compound risk loving individual chose the simple lottery in Pair 25 (with EV of US $43.75) over the actuarially-equivalent lottery (with EV of US $35), and switches in Pair 10 to choose the compound lottery over the simple lottery (each with the same EV values as in Pair 25). This implies that such a subject attaches additional value to the compound lottery that makes the subject violate ROCL: the subject is switching to choose a compound lottery that pays, after multiplying and reducing probabilities to a single stage, either US $70 or nothing with a 50:50 chance when the subject had previously chosen a lottery that pays either US $70 or US $35 with 25% and 75% probabilities, respectively, a less risky lottery with higher expected value. This reveals a non-trivial preference for compound lotteries that induces the subject to forego US $8.75 of expected value to choose the compound lottery over the simple lottery.

[23] As a point of reference, the expected value of the compound lottery in Pair 10 is US $35. The CE of the compound lottery implies a risk premium of more than 50% according to the models where homogeneity is assumed. However, this premium is also capturing the behavior of more than half of the subjects in the experiment that avoided any of the compound lotteries.

4. Conclusions and Discussion

Because of the attention paid to violations of the Independence Axiom, it is noteworthy that early formal concerns with the possibility of a "utility or disutility for gambling" centered around the Reduction of Compound Lotteries (ROCL) axiom.[24] Von Neumann and Morgenstern [1953, p. 28] commented on the possibility of allowing for a (dis)utility of gambling component in their preference representation:[25]

> Do not our postulates introduce, in some oblique way, the hypotheses which bring in the mathematical expectation [of utility]? More specifically: May there not exist in an individual a (positive or negative) utility of the mere act of 'taking a chance,' of gambling, which the use of the mathematical expectation obliterates? How did our axioms (3:A)-(3:C) get around this possibility? As far as we can see, our postulates (3:A)-(3:C) do not attempt to avoid it. Even the one that gets closest to excluding the 'utility of gambling' - (3:C:b)- seems to be plausible and legitimate - unless a much more refined system of psychology is used than the one now available for the purposes of economics [...] Since (3:A)-(3:C) secure that the necessary construction [of utility] can be carried out, concepts like a 'specific utility of gambling' cannot be formulated free of contradiction on this level.

On the very last page of their *magnus opus*, von Neumann and Morgenstern [1953; p. 632] propose that if their postulate (3:C:b), which is the ROCL, is relaxed, one could indeed allow for a specific utility for the act of gambling:

> It seems probable, that the really critical group of axioms is (3:C) - or, more specifically, the axiom (3:C:b). This axiom expresses the combination rule for multiple chance alternatives, and it is plausible, that a specific utility or disutility of gambling can only

[24] The issue of the (dis)utility of gambling goes back at least as far as Pascal, who argued in his *Pensées* that "people distinguish between the pleasure or displeasure of chance (uncertainty) and the objective evaluation of the worth of the gamble from the perspective of its consequences" (see Luce and Marley [2000; p. 102]). Referring to the ability of bets to elicit beliefs, Ramsey [1926] claims that "[t]his method I regard as fundamentally sound; but it suffers from being insufficiently general, and from being necessarily inexact. It is inexact partly [...] because the person may have a special eagerness or reluctance to bet, because he either enjoys or dislikes excitement or for any other reason, e.g. to make a book. The difficulty is like that of separating two different cooperating forces" (from the reprint in Kyburg and Smokler [1964; p. 73]).

[25] To understand this quote, the intuitive meaning of the von Neumann-Morgensten axioms are as follows: axiom (3:A:a) is a completeness-of-preferences assumption, axiom (3:A:b) is a transitivity axiom, axioms (3:B:a) and (3:B:b) are in the spirit of an independence axiom, axioms (3:B:c) and (3:B:d) reflect continuity assumptions, and axioms (3:C:a) and (3:C:b) are those that deal with compound lotteries.

> exist if this simple combination rule is abandoned. Some change of the system [of axioms] (3:A)-(3:B), at any rate involving the abandonment or at least a radical modification of (3:C:b), may perhaps lead to a mathematically complete and satisfactory calculus of utilities which allows for the possibility of a specific utility or disutility of gambling. It is hoped that a way will be found to achieve this, but the mathematical difficulties seem to be considerable.

Thus, the relaxation of ROCL opens the door to the possibility of having a distinct (dis)utility for the act of gambling on compound lotteries with objective probabilities. This implies that people would have preferences over compound lotteries that differ from preferences over single-stage lotteries.

Our primary goal is to test this hypothesis for objective probabilities. Our conclusions are influenced by the experiment payment protocols used and the assumptions about how to characterize risk attitudes and heterogeneity across subjects. We find evidence of violations of ROCL, but only when subjects are presented with choices in which the binary choices involve compound lotteries *and* the payment protocol is itself generates an additional layer of compounding. When subjects are only presented with one binary choice, and there is no additional compounding required by the payment protocol, behavior is consistent with ROCL. These results are obtained consistently whether we use structural econometrics to estimate preferences or non-parametric statistics to analyze choice patterns. It is important to realize that testing ROCL using payment procedures that might assume ROCL itself, or a weaker axiom of choice over compound lotteries, such as CIA, might introduce potential confounds. For instance, our results imply that the violations of ROCL that we observe are a product of using the RLIM payment procedure in the experiment. In this sense, the payment protocol is contaminating the hypothesis testing of ROCL.

We do not test any specific theories to explain why ROCL does well in one setting compared to the other. Several can be conjectured and we leave for future research the systematic analysis of these conjectures.

First, it is possible that subjects pay more attention to the properties of the compound lotteries, for example, by calculating the expected value of the lottery when they are offered only one decision to make, and thereby satisfy ROCL with greater frequency. However, when they are offered compound lotteries embedded in a 1-in-K payment protocol design, subjects might have less time to rationalize their choices and tend to succumb to heuristics, such as compound risk aversion or loving, to make fast choices and compensate for the lack of time to analyze in detail the properties of the compound lotteries.

Another explanation could be that subjects in the 1-in-40 treatment see themselves as facing one foreground risk that is well specified at choice k, and view the remaining 40-k choices as akin to a "background risk." This is plausible since they do not know the specific risks, but could guess at their general form. Even in the case of zero-mean background risks, positive and negative effects on foreground risk aversion can be predicted (Eeckhoudt, Gollier and Schlesinger [1996], Gollier and Pratt [1996], Quiggin [2003] and Harrison, List and Towe [2009]). Similarly, it is also possible that subjects are attempting to form a portfolio in the 1-in-40 treatment, whereas this is not possible by construction in the 1-in-1 treatment. Again, this explanation is complicated by the fact that subjects typically do not know the specific lotteries to come. Additionally, there could be "learning effects" over time as subjects gain experience, or exhibit fatigue, with evaluating lotteries.

Additionally, there is a strand of literature that relates attitudes towards compound lotteries and gradual resolution of risk over time. For example, Dillenberger [2010] studies the effect of time on preferences by distinguishing between uncertainty that is resolved over time, which creates a compounded representation of uncertainty, and one-shot uncertainty.

Finally, there could be a threshold level of compounding above which subjects have trouble satisfying ROCL[26] due to a cognitive inability to reduce complex compound probabilities or due to

---

[26] In our case, the payment protocol in the 1-in-40 treatment is inducing an additional layer of risk that might be triggering the ROCL violations. This implies a very simple hypothesis that can be tested in future research: subjects satisfy ROCL until a certain number of layers of risks. A natural test of this hypothesis would be to analyze the propensity to violate ROCL in compound lotteries with two layers of risks, three layers of risk and so on.

subjects simply finding pleasure in facing several layers of risk. This is related to a much simpler explanation for ROCL violations, implied by Smith [1969], where people might derive utility or disutility of gambling. If compound lotteries are subjectively perceived as closer to such gambling experiences than single-stage lotteries, then ROCL might fail. For instance, a gambling lover will always derive more utility from a compound lottery than its actuarially-equivalent lottery. For this person, there is more of a thrill playing a gambling game that involves facing several layers of chance, compared to playing a single shot gamble with the same odds of winning where uncertainty is resolved in only one stage.[27]

However, it is still puzzling why violations are observed in the 1-in-40 treatment but not in the 1-in-1 treatment.

ROCL is central to the evaluation of behavior towards risk, uncertainty and ambiguity. We present experimental evidence on the validity of ROCL in a specific domain defined over objective probabilities. We caution against any experimental evaluation of ROCL over subjective beliefs that assumes no interaction with the payment protocol.

[27] These explanations, and others, could be examined with extended designs. For example, one could test other values of K. We considered K=40 since this is a plausible level for studies estimating risk attitudes and testing the axioms of EUT, yet in different settings a smaller or larger K is of interest. For instance, the popular Holt and Laury [2002] method for eliciting risk attitudes uses K=10. Harrison and Swarthout [2014] show that behavior over 1-in-30 choices differs from behavior over 1-in-1 choices, although they did not test the interaction of the ROCL axiom with those payment protocols.

Figure 1: Probability Coverage of Battery of 40 Lotteries Pairs

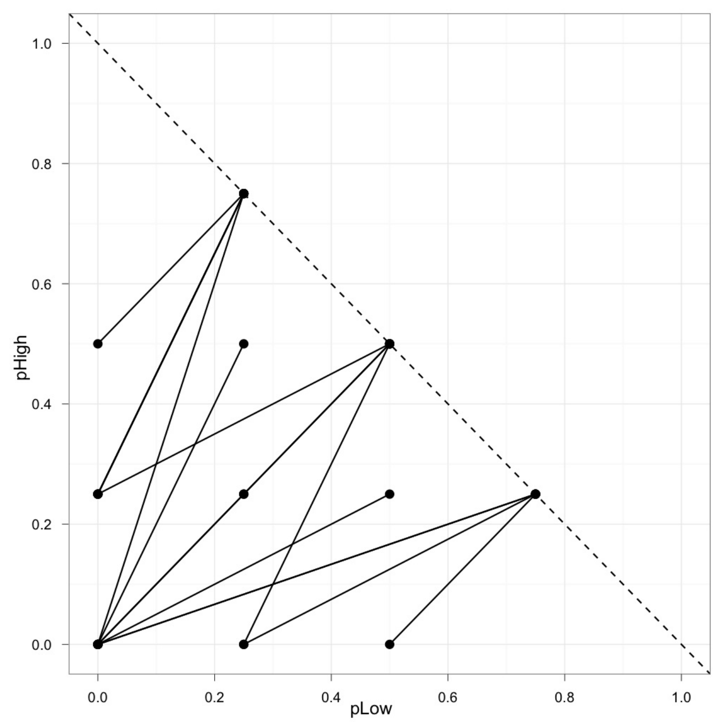


Figure 2: Tree Representation of a Compound Lottery and its Corresponding Actuarially-Equivalent Simple Lottery

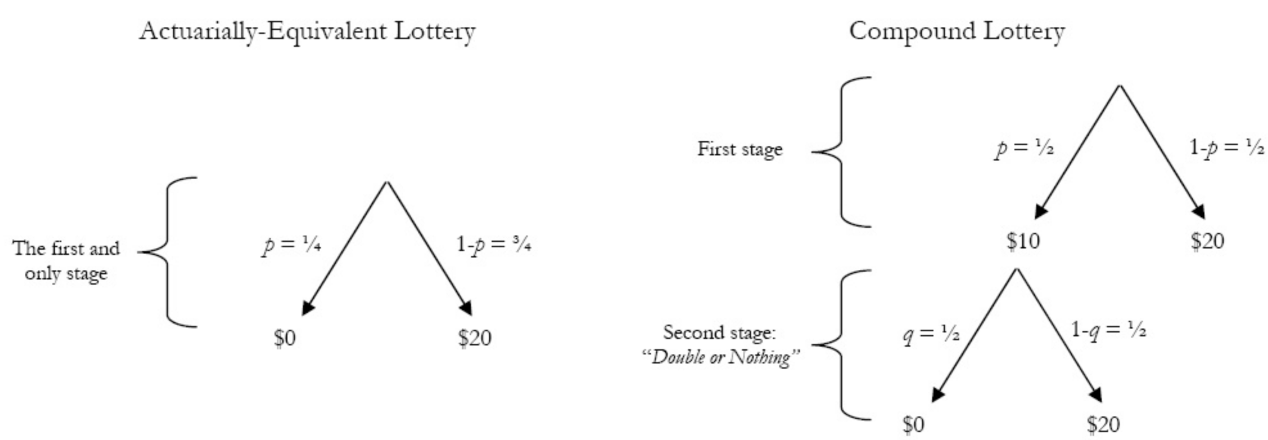


Table 1: Experimental Design

Treatment	Subjects	Choices
1. Pay-1-in-1	133	133
2. Pay-1-in-40	62	2480

Figure 3: Choices Over Compound and Actuarially-Equivalent Lotteries

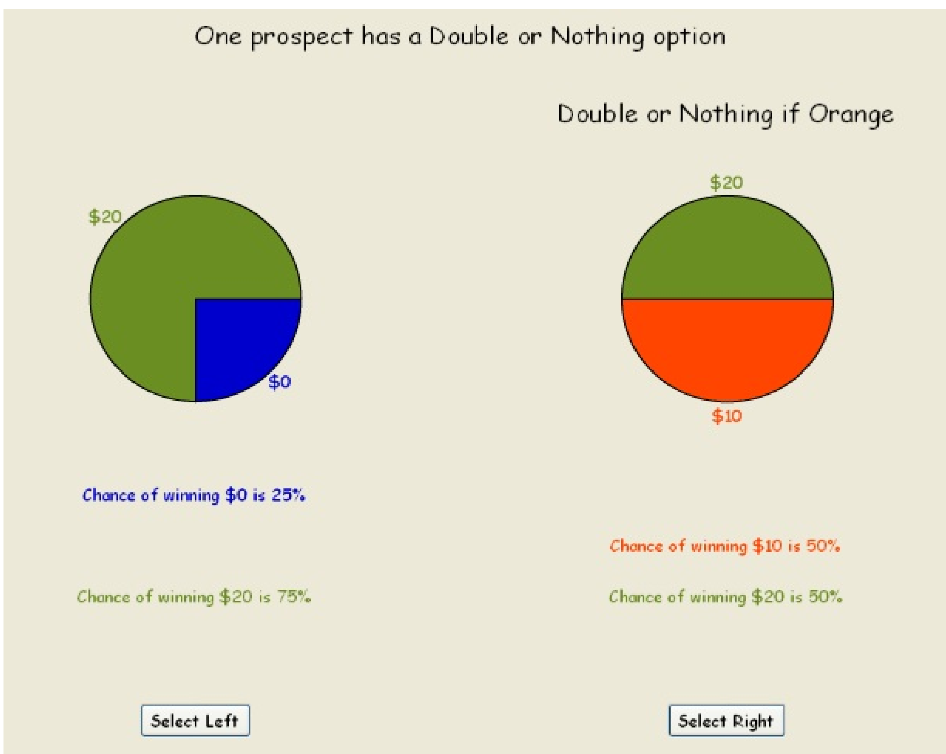


Figure 4: Choices Over Simple and Compound Lotteries

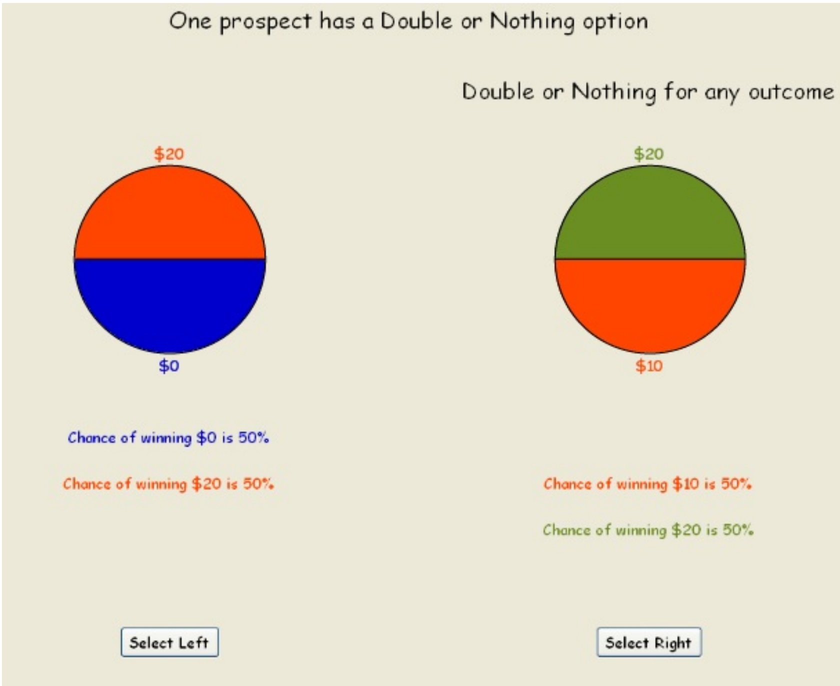


Figure 5: Choices Over Simple and Actuarially-Equivalent Lotteries

Figure 6: Estimated RDU Models for 1-in-1 Data

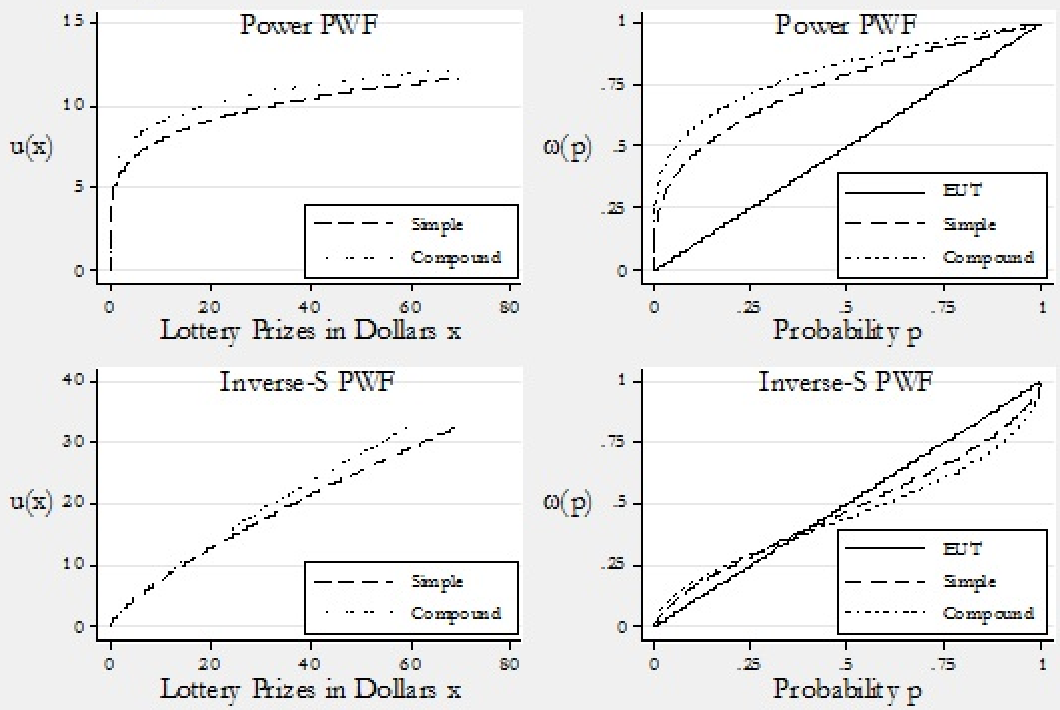


Figure 7: Estimated RDU Models for 1-in-40 Data

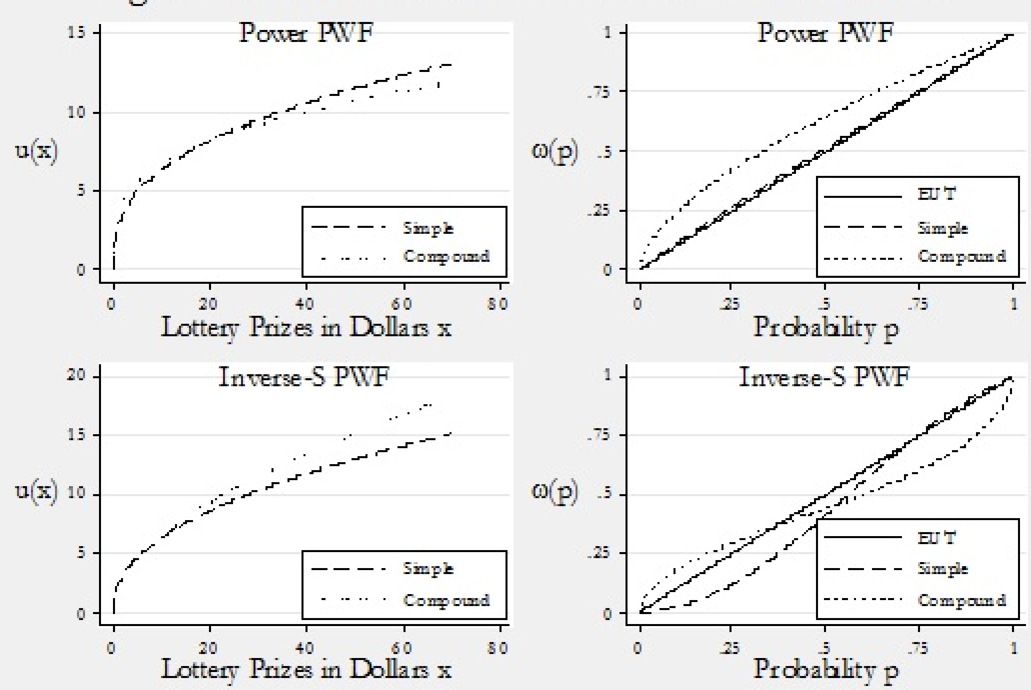
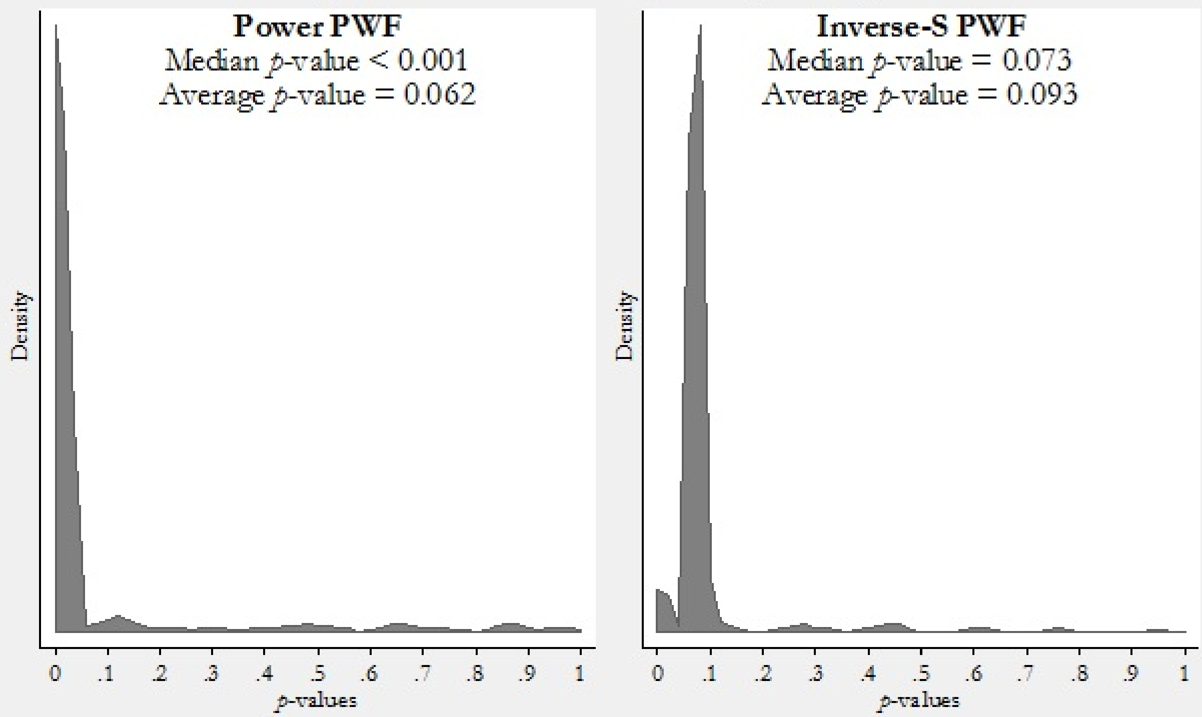
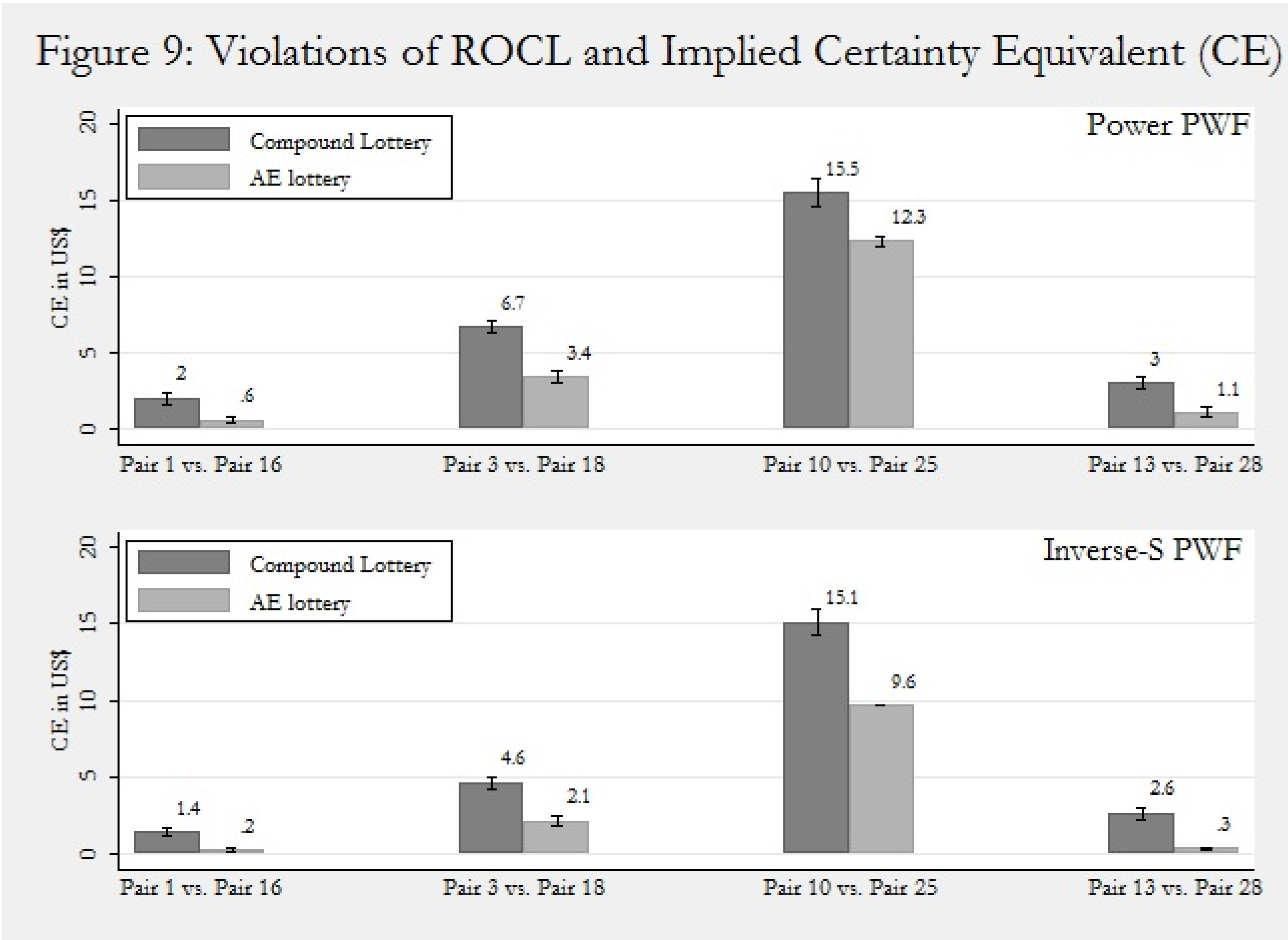


Figure 8: Bootstrap Evaluation of Different Sample Sizes

500 bootstrap evaluations using the RDU source-dependent model assuming homogeneous preferences





Note 1: Standard errors for the estimated CE are represented by the vertical lines at the top of the bars.

Note 2: As a point of reference, the expected values for the compound lottery and its actuarially equivalent lottery in each pairwise comparison are (from left to right in the figure): US $ 5, US $ 11.3 US $ 35 and US $ 8.8.

References

Abdellaoui, Mohammed; Baillon, Aurélien; Placido, Lætitia and Wakker, Peter P., "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review*, 101, April 2011, 695-723.

Abdellaoui, Mohammed; Klibanoff, Peter, and Placido, Lætitia, "Ambiguity and Compound Risk Attitudes: An Experiment," *Working Paper*, MEDS Department, Kellogg School of Management, Northwestern University, 2014; *Management Science*, forthcoming.

Cox, James C.; Sadiraj, Vjollca, and Schmidt, Ulrich, "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics*, 18, 2015, 215-250.

Dillenberger, David, "Preferences for One-shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica*, 78(6), 2010, 1973-2004.

Eeckhoudt, Louis; Gollier, Christian, and Schlesinger, Harris, "Changes in Background Risk and Risk Taking Behavior," *Econometrica*, 64, 1996, 683-689.

Ellsberg, Daniel, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 1961, 643-669.

Fellner, William, "Distortion of Subjective Probabilities as Reaction to Uncertainty," *Quarterly Journal of Economics*, 48(5), November 1961, 670-689.

Fellner, William, "Slanted Subjective Probabilities and Randomization: Reply to Howard Raiffa and K. R. W. Brewer," *Quarterly Journal of Economics*, 77(4), November 1963, 676-690.

Gollier, Christian, and Pratt, John W., "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica*, 64, 1996, 1109-1123.

Halevy, Yoram, "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75, 2007, 503-536.

Harrison, Glenn W., "The Rich Domain of Uncertainty: Comment," *Working Paper 2011-13*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2011.

Harrison, Glenn W.; List, John A., and Towe, Chris, "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75(2), March 2007, 433-458.

Harrison, Glenn W., and Swarthout, J. Todd, "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision*, 77(3), 2014, 423-438.

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.

Kyburg, Henry E. and Smokler, Howard E., *Studies in Subjective Probability* (New York: Wiley and Sons, 1964).

Luce, R. Duncan, and Marley, A.A.J., "On Elements of Chance," *Theory and Decision*, 49, 2000, 97-126.

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

Quiggin, John, "Background Risk in Generalized Expected Utility Theory," *Economic Theory*, 22, 2003, 607-611.

Ramsey, Frank P., *The Foundations of Mathematics and Other Logical Essays* (New York: Harcourt Brace and Co, 1926).

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

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).

Smith, Vernon L., "Measuring Nonmonetary Utilities in Uncertain Choices: the Ellsberg Urn," *Quarterly Journal of Economics*, 83(2), May 1969, 324-329.

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

von Neumann, John, and Morgensten, Oskar, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1953; Third Edition; Princeton University Paperback Printing, 1980).

Appendix A: Parameters

To construct our battery of 40 lottery pairs, we used several criteria to choose the compound lotteries and their actuarially-equivalent lotteries used in our experiment:

1. The lottery compounding task should be as simple as possible. The instructions used by Halevy [2007] are a model in this respect, with careful picture illustrations of the manner in which the stages would be drawn. We wanted to avoid having physical displays, since we had many lotteries. We also wanted to be able to have the computer interface vary the order for us on a between-subject basis, so we opted for a simpler procedure that was as comparable as possible in terms of information as our simple lottery choice interface.
2. The lottery pairs should offer reasonable coverage of the Marschak-Machina (MM) triangle and prizes.
3. There should be choices/chords that assume parallel indifference curves, as expected under EUT, but the slope of the indifference curve should vary, so that the battery of lotteries can be used to test for a wide range of risk attitudes under the EUT null hypothesis.
4. There should be a number of compound lotteries with their actuarially-equivalent counterparts in the interior of the triangle. Experimental evidence suggests that people tend to comply with the implications of EUT in the interior of the triangle and to violate it on the borders (Conlisk [1989], Camerer [1992], Harless [1992], Gigliotti and Sopher [1993] and Starmer [2000]).
5. We were careful to choose lottery pairs with stakes and expected payoff per individual that are comparable to those in the original battery of 69 simple lotteries, since these had been used extensively in other samples from this population.

Our starting point was the battery of 69 lotteries in Table A1 used in Harrison and Swarthout [2014], which in turn were derived from Wilcox [2010]. The lotteries were originally designed in part to satisfy the second and third criteria given above. Our strategy was then to "reverse engineer" the initial lotteries needed to obtain compound lotteries that would yield actuarially-equivalent prospects which already existed in the set of 69 pairs. For instance, the first pair in our battery of 40 lotteries was derived from pair 4 in the battery of 69 (contrast pair 1 in Table A2 with pair 4 in Table A1). We want the distribution of the "risky" lottery in the latter pair to be the actuarially-equivalent prospect of our compound lottery. To achieve this, we have an initial lottery that pays $10 and $0 with 50% probability each, and offering "double or nothing" if the outcome of the latter prospect is $10. Hence it offers equal chances of $20 or $0 if the DON stage is reached. The $5 stake was changed to $0 because DON requires this prize to be among the possible outcomes of the compound lotteries.[28] The actuarially-equivalent lottery of this compound prospect pays $0 with 75% probability and $20 with 25% probability, which is precisely the risky lottery in pair 4 of the default battery of 69 pairs. Except for the compound lottery in pair 9 in our set of lotteries, the actuarially-equivalent lotteries play the role of the "risky" lotteries.

Figure A1 shows the coverage of these lottery pairs in terms of the Marschak-Machina triangle. Each prize context defines a different triangle, but the patterns of choice overlap

[28] We contemplated using "double or $5," but this did not have the familiarity of DON.

considerably. Figure A1 shows that there are many choices/chords that assume parallel indifference curves, as expected under EUT, but that the slope of the indifference curve can vary, so that the tests of EUT have reasonable power for a wide range of risk attitudes under the EUT null hypothesis (Loomes and Sugden [1998] and Harrison, Johnson, McInnes and Rutström [2007]). 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.

The majority of our compound lotteries use a conditional version of the DON device because it allows to obtain good coverage of prizes and probabilities and keeps the compounding representation simple. As noted in the text, one can construct diverse compound lotteries with only two simple components: initial lotteries that either pay two outcomes with 50:50 odds or pay a given stake with certainty, and a conditional DON which pays double a predetermined amount with 50% probability or nothing with equal chance.

In our design, if the subject has to play the DON option she will toss a coin to decide if she gets double the stated amount. One could use randomization devices that allow for probability distributions different from these 50:50 odds, but we want to keep the lottery compounding simple and familiar. Therefore, if one commits to 50:50 odds in the DON option, using exclusively unconditional DON will only allow one to generate compound lotteries with actuarially-equivalent prospects that assign 50% chance to getting nothing. For instance, suppose a compound prospect with an initial lottery that pays positive amounts \$X and \$Y with probability p and $(1-p)$, respectively, and offers DON for any outcome. The corresponding actuarially-equivalent lottery pays \$2X, \$2Y and \$0 with probabilities $p/2$, $(1-p)/2$ and ½, respectively.

The original 69 pairs use 10 contexts defined by three outcomes drawn from \$5, \$10, \$20, \$35 and \$70. For example, the first context consists of prospects defined over prizes \$5, \$10 and \$20, and the tenth context consists of lotteries defined over stakes \$20, \$35 and \$70. As a result of using the DON device, we have to introduce \$0 to the set of stakes from which the contexts are drawn. However, some of the initial lotteries used prizes in contexts different from the ones used for final prizes, so that we could ensure that the stakes for the compounded lottery matched those of the simple lotteries. For example, pair 3 in Table A2 is defined over a context with stakes \$0, \$10 and \$35. The compound lottery of this pair offers an initial lottery that pays \$5 and \$17.50 with 50% chance each and a DON option for any outcome. This allows us to have as final prizes \$0, \$10 and \$35.

Our battery of 40 lotteries uses 6 of the original 10 contexts, but substitute the \$5 stake for \$0. We do not use the other 4 contexts: for them to be distinct from our 6 contexts they would have to have 4 outcomes, the original 3 outcomes plus the \$0 stake required by the DON option. We chose to use only compound lotteries with no more than 3 final outcomes, which in turn requires initial lotteries with no more than 2 outcomes. Accordingly, the initial lotteries of compound prospects are defined over distributions that offer either 50:50 odds of getting any of 2 outcomes or certainty of getting a particular outcome which makes our design simple. It is worth noting that there are compound lotteries composed of initial prospects that offer an amount \$X with 100% probability and a DON option that pays \$2X and \$0 with 50% chance each. By including this type of "trivial" compound lottery, we provide the basis for ROCL to be tested in its simplest form.

Finally, we included compound lotteries with actuarially-equivalent counterparts in the interior and on the border of the MM triangle, since previous experimental evidence suggests that this is relevant to test the implications of EUT.

Additional References

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

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

Gigliotti, Gary, and Sopher, Barry, "A Test of Generalized Expected Utility Theory," *Theory and Decision*, 35, 1993, 75-106.

Harless, David W, "Predictions about Indifference Curves Inside the Unit Triangle: a Test of Variants of Expected Utility," *Journal of Economic Behavior and Organization*, 18, 1992, 391-414.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Measurement With Experimental Controls," in M. Boumans (ed.), *Measurement in Economics: A Handbook* (San Diego, CA: Elsevier, 2007).

Starmer, Chris, "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38, June 2000, 332-382.

Table A1: Default Simple Lotteries

		Prizes			"Safe" Lottery Probabilities			"Risky" Lottery Probabilities				
Pair	Context	Low	Middle	High	Low	Middle	High	Low	Middle	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

Figure A1: Default Simple Lotteries

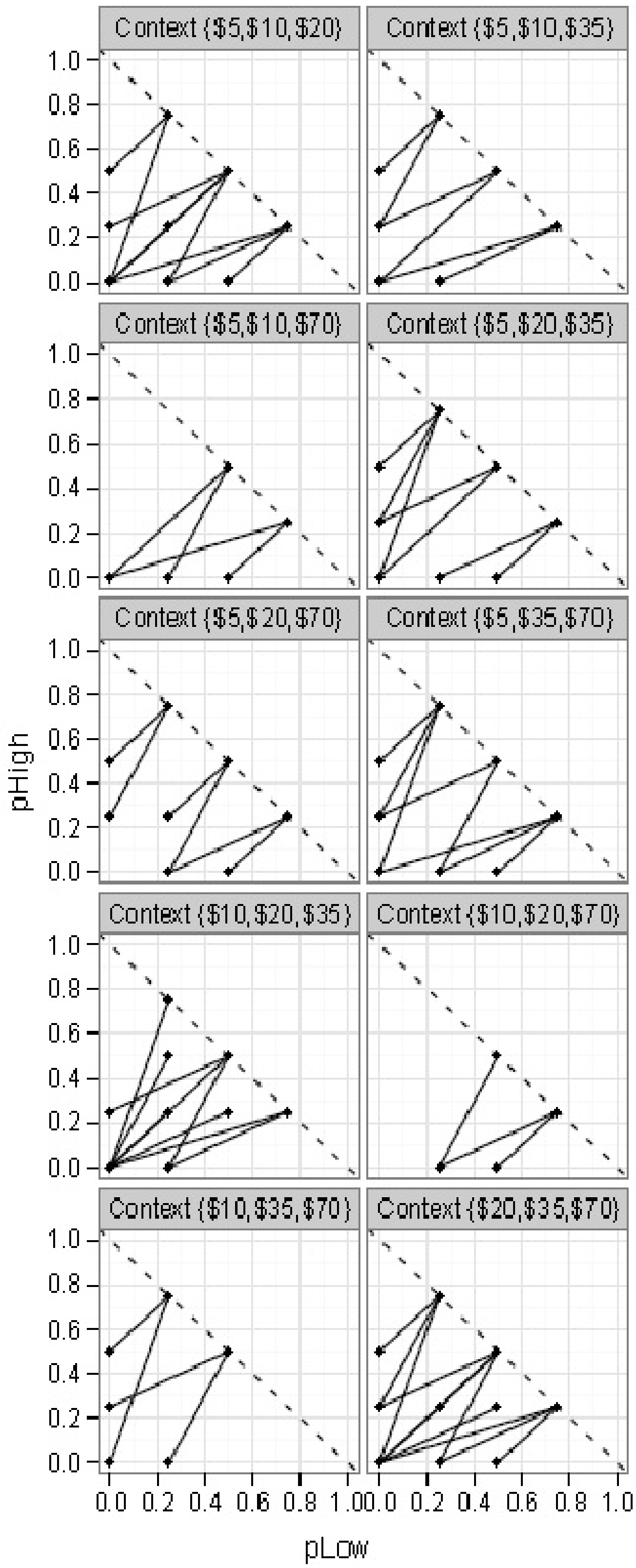


Table A2: Simple Lotteries vs. Compound Lotteries (Pairs 1-15)

								Compound Lottery								
		Final Prizes			Simple Lottery Probabilities			Initial Lottery Prizes			Initial Lottery Probabilities			"Double or Nothing"	EV	EV
Pair	Context	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	option	Simple	Compound
1	1	$0	$10	$20	0.5	0.5	0	$0	$10	$20	0.5	0.5	0	DON if middle	$5.00	$5.00
2	1	$0	$10	$20	0	1	0	$0	$10	$20	0.5	0.5	0	DON if middle	$10.00	$5.00
3	2	$0	$10	$35	0	1	0	$0	$5	$17.50	0	0.5	0.5	DON for any outcome	$10.00	$11.25
4	3	$0	$10	$70	0.25	0.75	0	$0	$35	$70	0	1	0	DON for any outcome	$7.50	$35.00
5	3	$0	$10	$70	0	1	0	$0	$35	$70	0	1	0	DON for any outcome	$10.00	$35.00
6	4	$0	$20	$35	0	1	0	$0	$10	$35	0	0.5	0.5	DON if middle	$20.00	$22.50
7	5	$0	$20	$70	0	0.5	0.5	$0	$35	$70	0	0.5	0.5	DON if middle	$45.00	$52.50
8	6	$0	$35	$70	0	1	0	$0	$35	$70	0	0.5	0.5	DON if middle	$35.00	$52.50
9	5	$0	$20	$70	0.5	0	0.5	$0	$20	$35	0	0.5	0.5	DON if high	$35.00	$27.50
10	6	$0	$35	$70	0	0.75	0.25	$0	$35	$70	0	1	0	DON for any outcome	$43.75	$35.00
11	5	$0	$20	$70	0	1	0	$0	$20	$35	0	0.5	0.5	DON if high	$20.00	$27.50
12	6	$0	$35	$70	0	0.75	0.25	$0	$35	$70	0	0.5	0.5	DON if middle	$43.75	$52.50
13	2	$0	$10	$35	0.25	0.75	0	$0	$17.50	$35	0.5	0.5	0	DON if middle	$7.50	$8.75
14	4	$0	$20	$35	0	0.75	0.25	$0	$17.50	$35	0	0.5	0.5	DON if middle	$23.75	$26.25
15	5	$0	$20	$70	0	0.75	0.25	$0	$35	$70	0	0.5	0.5	DON if middle	$32.50	$52.50

Table A3: Simple Lotteries vs. Actuarially-Equivalent Lotteries (Pairs 16-30)

		Final Prizes			Simple Lottery Probabilities			Actuarially-Equivalent Lottery Probabilities			EV Simple	EV Actuarially-Equivalent
Pair	Context	Low	Middle	High	Low	Middle	High	Low	Middle	High		
16	1	$0	$10	$20	0.5	0.5	0	0.75	0	0.25	$5.00	$5.00
17	1	$0	$10	$20	0	1	0	0.75	0	0.25	$10.00	$5.00
18	2	$0	$10	$35	0	1	0	0.5	0.25	0.25	$10.00	$11.25
19	3	$0	$10	$70	0.25	0.75	0	0.5	0	0.5	$7.50	$35.00
20	3	$0	$10	$70	0	1	0	0.5	0	0.5	$10.00	$35.00
21	4	$0	$20	$35	0	1	0	0.25	0.25	0.5	$20.00	$22.50
22	5	$0	$20	$70	0	0.5	0.5	0.25	0	0.75	$45.00	$52.50
23	6	$0	$35	$70	0	1	0	0.25	0	0.75	$35.00	$52.50
24	5	$0	$20	$70	0.5	0	0.5	0.25	0.5	0.25	$35.00	$27.50
25	6	$0	$35	$70	0	0.75	0.25	0.5	0	0.5	$43.75	$35.00
26	5	$0	$20	$70	0	1	0	0.25	0.5	0.25	$20.00	$27.50
27	6	$0	$35	$70	0	0.75	0.25	0.25	0	0.75	$43.75	$52.50
28	2	$0	$10	$35	0.25	0.75	0	0.75	0	0.25	$7.50	$8.75
29	4	$0	$20	$35	0	0.75	0.25	0.25	0	0.75	$23.75	$26.25
30	5	$0	$20	$70	0	0.75	0.25	0.25	0	0.75	$32.50	$52.50

Table A4: Actuarially-Equivalent Lotteries vs. Compound Lotteries (Pairs 31-40)

								Compound Lottery								
		Final Prizes			Actuarially-Equivalent Lottery Probabilities			Initial Lottery Prizes			Initial Lottery Probabilities			"Double or Nothing" option	EV Actuarially-Equivalent	EV Compound
Pair	Context	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High			
31	1	$0	$10	$20	0.75	0	0.25	$0	$10	$20	0.5	0.5	0	DON if middle	$5.00	$5.00
32	4	$0	$20	$35	0.25	0	0.75	$0	$17.50	$35	0	0.5	0.5	DON if middle	$26.25	$26.25
33	2	$0	$10	$35	0.5	0.25	0.25	$0	$5	$17.50	0	0.5	0.5	DON for any outcome	$11.25	$11.25
34	3	$0	$10	$70	0.5	0	0.5	$0	$35	$70	0	1	0	DON for any outcome	$35.00	$35.00
35	4	$0	$20	$35	0.25	0.25	0.5	$0	$10	$35	0	0.5	0.5	DON if middle	$22.50	$22.50
36	5	$0	$20	$70	0.25	0	0.75	$0	$35	$70	0	0.5	0.5	DON if middle	$52.50	$52.50
37	6	$0	$35	$70	0.25	0	0.75	$0	$35	$70	0	0.5	0.5	DON if middle	$52.50	$52.50
38	5	$0	$20	$70	0.25	0.5	0.25	$0	$20	$35	0	0.5	0.5	DON if high	$27.50	$27.50
39	6	$0	$35	$70	0.5	0	0.5	$0	$35	$70	0	1	0	DON for any outcome	$35.00	$35.00
40	2	$0	$10	$35	0.75	0	0.25	$0	$17.50	$35	0.5	0.5	0	DON if middle	$8.75	$8.75

Appendix B: Related Literature

Cubitt, Starmer and Sugden [1998a] studied the Reduction of Compound Lotteries Axiom (ROCL) in a 1-in-1 design that gave each subject one and only one problem for real stakes and was conceived to test principles of dynamic choice. Also, Starmer and Sugden [1991], Beattie and Loomes [1997] and Cubitt, Starmer and Sugden [1998] have studied the Random Lottery Incentive Method (RLIM), and as a by-product have tested the ROCL axiom. We focus on the results related to the ROCL axiom: Harrison and Swarthout [2014] review the results related to RLIM.

Cubitt, Starmer and Sugden [1998a] gave to one group of subjects one problem that involved compound lotteries and gave to another group the reduced compound version of the same problem. If ROCL is satisfied, one should see the same pattern of choice in both groups. They cannot find statistically significant violations of ROCL in their design.

Starmer and Sugden [1991] gave their subjects two pairs of lotteries that were designed to test "common consequence" violations of EUT. In each pair i there was a risky (Ri) option and a safe (Si) option. They recruited 160 subjects that were divided into four groups of equal number. Two groups faced one of the two pairs in 1-in-1 treatments, while the other two groups were given both pairs to make a choice over using the RLIM to choose the pair for final payoff. We focus on the latter two groups since RLIM induces four possible compound lotteries: i) (0.5, R1; 0.5, R2), ii) (0.5, R1; 0.5, S2), iii) (0.5, R2; 0.5, S1) and iv) (0.5, S1; 0.5, S2). The lottery parameters were chosen to make compound lotteries ii) and iii) have equal actuarially-equivalent prospects.

They hypothesize that if a *reduction principle* holds, and if any of the induced compound lotteries ii) and iii) above is preferred by a subject, then the other one must be preferred.[29] The rejection of this hypothesis is a violation of ROCL, since this axiom implies that two compound lotteries with the same actuarially-equivalent prospects should be equally preferred. Therefore, the null hypothesis in Starmer and Sugden [1991; p. 976] is that "the choice between these two responses to be made at random; as a result, these responses should have the same expected frequency." From the 80 subjects that faced the 1-in-2 treatments, 32.5% of the individuals chose (0.5, R1; 0.5, S2) and 15% chose (0.5, R2; 0.5, S1), thus Starmer and Sugden reject the null hypothesis of equal frequency in choices based on a one-tail test with a binomial distribution and *p*-value=0.017. This pattern is very similar in each of the 1-in-2 treatments; in one of them the proportions are 30% and 15%, whereas in the other they are 35% and 15%. A two-sided Fisher Exact test yields a *p*-value of 0.934, which suggest that these patterns of choices are very similar in both 1-in-2 treatments. Therefore, there is no statistical evidence to support ROCL in their experiment.

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, two of which were compound lotteries.[30] They recruited 289 subjects that were randomly

---

[29] Following Holt [1986], Starmer and Sugden [1991, p. 972] define the reduction principle to be when "compound lotteries are reduced to simple ones by the calculus of probabilities and that choices are determined by the subject's preferences over such reduced lotteries." This is what we call ROCL, in *addition* to some axioms that are maintained for present purpose.

[30] Beattie and Loomes [1997] use nine prospects: A= (0.2, £15; 0.8, £0), B= (0.25, £10; 0.75, £0), C= (0.8, £0; 0.2, £30), D= (0.8, £5; 0.2, £0), E=(0.8, £15; 0.2, £0), F = (1, £10), G=(1, £4), H=(0.5,£10; 0.5,£0),

assigned to six groups. The first group faced a hypothetical treatment and was paid a flat fee for completing all four tasks. The second group was given a 1-in-4 treatment, and each of the other four groups faced one of the four tasks in 1-in-1 treatments. Sample sizes were 49 for the hypothetical treatment and the 1-in-4 treatment, and a total of 191 in the four 1-in-1 treatments.

Beattie and Loomes [1997; p. 164] find that "there is no support for the idea that two problems involving the same 'reduced form' alternatives – and therefore involving the same difference between expected values – will be treated equivalently." On this basis, their Question 3 in the 1-in-4 treatment would be actuarially-equivalent to their Question 1 in the 1-in-1 treatment. They found that the pattern of choices in both treatments are so different "that a chi-square test rejects with a very great confidence (p<.001) the hypothesis that they are treated equivalently" (p. 164). The p-value<0.001 of the Fisher Exact test provides further support for this violation of ROCL.

Their Question 4 is a task that is similar to the method developed by Binswanger [1980]: subjects are offered an ordered set of choices that increase the average payoff while increasing variance. The difference with the Binswanger procedure is that two of the four choices involved compound lotteries: one paid a given amount of money if two Heads in a row were flipped, and the other paid a higher amount if three Heads in a row were flipped. For responses in Question 4, Beattie and Loomes [1997; p.162]

> ...conjecture that the REAL [1-in-1] treatment might stimulate the greatest effort to picture the full sequential process [of coin flipping in the compound prospects] and, as a part of that, to anticipate feelings at each stage in the sequence; whereas the HYPO [hypothetical] treatment would be most conducive to thinking of the alternatives in their reduced form as a set of simple lotteries... The RPSP [1-in-4 treatment] might then, both formally and psychologically, represent an intermediate position, making the process less readily imaginable by adding a further stage (the random selection of the problem) to the beginning of the sequence, and reducing but not eliminating the incentive to expend the necessary imaginative effort.

On this basis, they predict that, when answering Question 4, subjects in the hypothetical treatment are more likely to think in *reduced form probability distributions*. Beattie and Loomes consider that this might enhance the salience of the high-payoff option, and thus the compound lotteries are expected to be chosen more frequently in the hypothetical treatment than in the 1-in-1 and 1-in-4 treatments.

Beattie and Loomes [1997; p.165] found support for these conjectures: Their subjects tend to choose the compound lotteries more often in the hypothetical treatment than in the ones with economic incentives (i.e., 1-in-1 and 1-in-4 treatments). They found that under the hypothetical treatment more than 1 in 3 of the sample opted for the compound lotteries; this proportion was reduced in the 1-in-4 treatment to just over 1 in 5; and in the 1-in-1 treatment the proportion fell to 1 in 12. A chi-square test rejects (p-value < 0.01) the hypothesis that there is no difference in patterns

---

I=(£25 if two Heads in a row are flipped; otherwise nothing) and J=(£62.50 if three Heads in a row are flipped; otherwise nothing). Questions 1 through 3 are binary choices that offer, respectively, A or B, C or D and E or F. In Question 4, the subject must choose the prospect that she prefers the most among G, H, I or J. Options I and J are compound lotteries with 2 and 3 stages, respectively.

across treatments. The Fisher Exact test is consistent with this result.[31]

Cubitt, Starmer and Sugden [1998] use common consequence and common ratio pairs of pairs in three experiments. We focus in the first two since the third experiment has no treatments relevant to test ROCL. In the first experiment they compare 1-in-1 choices with 1-in-2 choices. Their comparison rests on subjects not having extreme risk-loving preferences over the pairs of lotteries in the 1-in-2 treatment designed to capture this behavior. Given that this *a priori* assumption is true, and it is generally supported by their data, the lottery pairs in each of the 1-in-2 treatments were chosen to generate compound prospects with actuarially-equivalent lotteries equal to the prospects in each of the 1-in-1 treatments.

If ROCL is satisfied, the distribution of responses between risky and safe lotteries should be the same in both treatments. The *p*-value from the Fisher Exact test in one of the 1-in-1 and 1-in-2 treatment comparisons[32] is 0.14, which suggests that ROCL is most likely violated.[33]

Similarly, in the second experiment the 1-in-2 treatment induced compound lotteries with actuarially-equivalent prospects equal to the lottery choices in one of their 1-in-1 treatment. In the latter, 52% of the 46 subjects chose the risky lottery, whereas 38% of the 53 subjects in the 1-in-2 treatment chose the risky prospect. These choice patterns suggest that ROCL does not hold in the second experiment.[34]

Bernasconi and Loomes [1992] devise exactly the right sorts of tests of ROCL with objective probabilities, motivated we were by the theoretical claims of Segal [1987] about the role of ROCL with respect to the modeling of ambiguity aversion. Unfortunately their experiments were hypothetical, and conducted with pseudo-volunteers from classes. Although the authors claim that "... we can confidently say that we found no obvious signs of widespread confusion, misunderstanding or carelessness," there is no way to ascertain *a priori* the reliability of the responses.

Abdellaoui, Klibanoff and Placido [2014] elicit certainty equivalents (CE) from simple risk and compound risk lotteries on a within-subjects basis. Each subject made a number of binary choices to provide a certainty-equivalent interval for each of 32 gambles, with one selected at the end to play out for money. Thus their design *assumes* the RLIM procedure which we evaluate in our design, and which we find to interact with inferences about ROCL over objective probabilities. They also included some

---

[31] We test the similarity between treatments of the proportions of subjects that chose each of the four prospects in Question 4. The two-sided Fisher Exact test applied to the hypothetical and the 1-in-1 treatments rejects the hypothesis of no difference in choice patterns (*p*-value = 0.001). The *p*-value for the comparison of the same four choices between the hypothetical and the 1-in-4 treatments is 0.473.

[32] Groups 1.1 and 1.3 in their notation.

[33] The proportions of subjects that chose the risky prospect in the other 1-in-1 and 1-in-2 treatments (groups 1.2 and 1.4 in their notation) are close: 50% and 55%, respectively. However, we cannot perform the Fisher Exact test for this 1-in-1 and 1-in-2 comparison, since the compound lotteries induced by the 1-in-2 treatment have actuarially-equivalent prospects equal to the ones in the 1-in-1 treatment only if the subjects do not exhibit extreme risk-loving preferences. Since 8% of the subjects in this 1-in-2 treatment exhibited risk-loving preferences, one cannot perform the Fisher test because this contaminates the comparison between the compound lotteries and their actuarially-equivalent counterparts.

[34] Since one of the subjects in the 1-in-2 treatment (group 2.3) exhibited risk-loving preferences, we cannot perform the Fisher Exact test for the reasons explained earlier.

ambiguous lotteries within the 32 gambles, to test whether attitudes towards compound objective risk, as compared to simple objective risk, were correlated with attitudes towards ambiguity.

Focusing on their results for simple and compound objective risk, they compare certainty equivalents over the 64 subjects for tasks with the same expected value (EV). They further break the comparisons down into one set of tasks for which the winning probability was 1/12, another set of tasks for which it was 6/12, and a third set of tasks for which it was 11/12. Within each set the EV was the same: €4.17, €25 and €45.83. Thus any inferences about the effect of the probability of winning on the risk premium are perfectly confounded by inferences about the effect of the scale of the prizes. They report (p.9) that subjects were, on average, risk-seeking for low probabilities, risk-neutral for probabilities of one-half, and risk-averse for high probabilities. Of course, this is consistent with increasing relative risk aversion for prizes deriving solely from the utility function, and need not be due to risk premia induced by probability weighting.

The critical test is whether one observed differences in behavior between the CE for simple lotteries and the CE for compound lotteries. For the 1/12 probability, they report (Table 3, p.10) an average CE of 7.81 with a standard deviation of 5.22 for the simple lottery, and an average CE of 6.78 with a standard deviation of 5.12 for one of the representative compound risks (called "hypergeometric compound risk" by them). In this case the corresponding *t*-test, assuming unequal variances, has a *t* statistic of 1.13 and a two-sided *p*-value on the null hypothesis of no difference in means equal to 0.26. So one would not conclude that there is any statistically significant difference. The same tests for the other four compound risk types of the probability 1/12 result in *p*-values of 0.06 ("compound risk high"), 0.07 ("compound risk high with explicit degenerate urn"), 0.29 ("compound risk low") and 0.32 ("compound risk low with explicit degenerate urn"). So 2 of the 5 tests have *p*-values that might be viewed as indicating statistically significant differences in risk premia between simple and compound lotteries. The correct test would recognize that these are matched samples, since the same subject gave the CE in each case; we do not have access to the raw data at the individual level to conduct these tests. The correct test would also recognize that the choices across the pairs are not independent, so that the five *p*-values reported above are not independently generated.

For the set of compound lotteries with a winning probability of 6/12, the *p*-values for these tests are 0.13, 0.91, 0.03, 0.82, 0.33, 0.72 and 0.22. The first 2 tests refer to "diverse uniform compound risk" and "degenerate uniform compound risk," respectively, and the remaining 5 tests are the same as for the winning probability of 1/12. Only 1 of the 7 tests have a *p*-value approaching statistical significance.

Abdellaoui, Klibanoff and Placido [2014] draw conclusions about the validity of ROCL for objective risks from the tabulations of CE differences in their Table 4. These tabulations show, for each comparison type, the fraction of the 64 subjects that had a CE for the simple risk below the CE for the compound risk, the fraction that had *exactly the same CE*, and the fraction that has a E for the simple risk above the CE for the compound risk. They refer to the middle fraction as those exhibiting behavior consistent with ROCL, but obviously this ignores the usual sampling variability one expects to see in such behavior. They report ROCL-consistent fractions that are generally between 20% and 26%, but it is hard to know how to interpret those fractions in any statistically meaningful way. This is unfortunate, since the inferences they draw from them correspond to our own when one assumes

RLIM.

Finally, for the set of compound lotteries with a winning probability of 11/12, the *p*-values for these tests are 0.19, <0.001, 0.002, <0.001, and 0.01 respectively. The type of compound risk in each case is the same as for the 1/12 case. Here we see much more significant evidence of differences in risk attitudes, with the evidence in all cases in favor of there being a larger risk premium for the compound risk than for the EV-comparable simple risk.

Kaivanto and Kroll [2012] undertake a strange experiment in which subjects are asked to choose between two ways of generating a 1/10 probability of winning €100 or nothing. The first way is to roll a 100-sided die, and receive the prize if the same two digits appear (i.e., 11, 22, 33, etc.). the other way is to roll a 10-sided die twice, where the first roll generates the first digit and the second roll generates the second digit. The first randomization method is interpreted as a simple lottery, and the second randomization method as a compound lottery. The majority of subjects preferred the 100-sided randomization method. It is not at all apparent why one would view this as a test of ROCL in any interesting sense.

Additional References

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

Bernasconi, Michele, and Loomes, Graham, "Failures of the Reduction Principle in an Ellbserg-type Problem," *Theory and Decision*, 32(1), 1992, 77-100.

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

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

Cubitt, Robin P.; Starmer, Chris, and Sugden, Robert, "Dynamic Choice and Common Ration Effect: An Experimental Investigation ," *The Economic Journal*, 108, September 1998a, 1362-1380.

Kaivanto, Kim, and Kroll, Eike B., "Negative Recency, Randomization Device Choice, and Reduction of Compound Lotteries," *Economics Letters*, 115, 2012, 263-267.

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

Segal, Uzi, "The Ellsberg Paradox and Risk: An Anticipated Utility Approach," *International Economic Review*, 28, 1987, 175-202.

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

Appendix C: Instructions (WORKING PAPER)

B.1. Instructions for Treatment 1-in-1

1p

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.

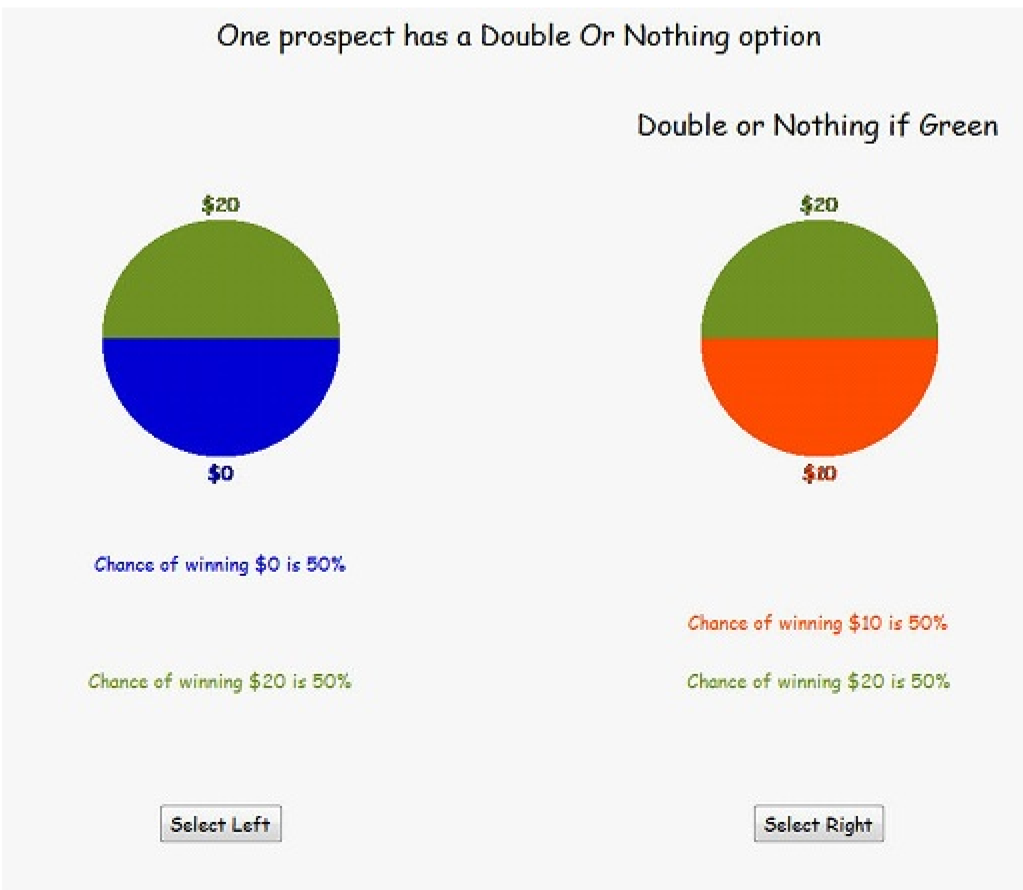
Here is an example of what the computer display of a pair of prospects might look like.

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.

You could also get a pair of prospects in which one of the prospects will give you the chance to play “Double or Nothing.” For instance, the right prospect in the following screen image pays



“Double or Nothing” if the Green area is selected, which happens if the number drawn is between 51 and 100. The right pie chart indicates that if the number is between 1 and 50 you get $10. However, if the number is between 51 and 100 a coin will be tossed to determine if you get double the amount. If it comes up Heads you get $40, otherwise you get nothing. The prizes listed underneath each pie refer to the amounts before any "Double or Nothing" coin toss.

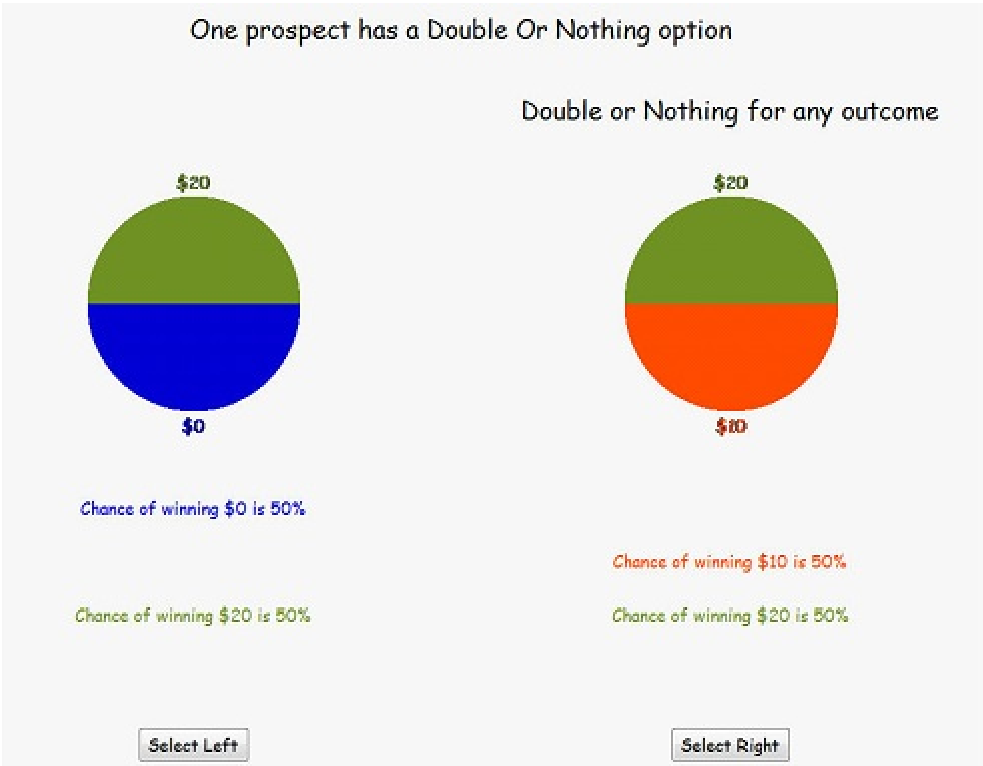
The pair of prospects you choose from is shown on a screen on the computer. On that screen, you should indicate which prospect you prefer to play by clicking on one of the buttons beneath the prospects.

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, and if necessary you will then toss a coin to determine if you get "Double or Nothing."

For instance, suppose you picked the prospect on the left in the last example. If the random number was 37, you would win $0; if it was 93, you would get $20.

If you picked the prospect on the right and drew the number 37, you would get $10; if it was 93, you would have to toss a coin to determine if you get "Double or Nothing." If the coin comes up Heads then you get $40. However, if it comes up Tails you get nothing from your chosen prospect.

It is also possible that you will be given a prospect in which there is a "Double or Nothing" option no matter what the outcome of the random number. The screen image below illustrates this possibility.



Therefore, your payoff is determined by three things:

- by which prospect you selected, the left or the right;
- by the outcome of that prospect when you roll the two 10-sided dice; and
- by the outcome of a coin toss if the chosen prospect outcome is of the "Double or Nothing" type.

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.

B.2 Instructions for Treatment 1-in-40

40p

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 40 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 might look like.

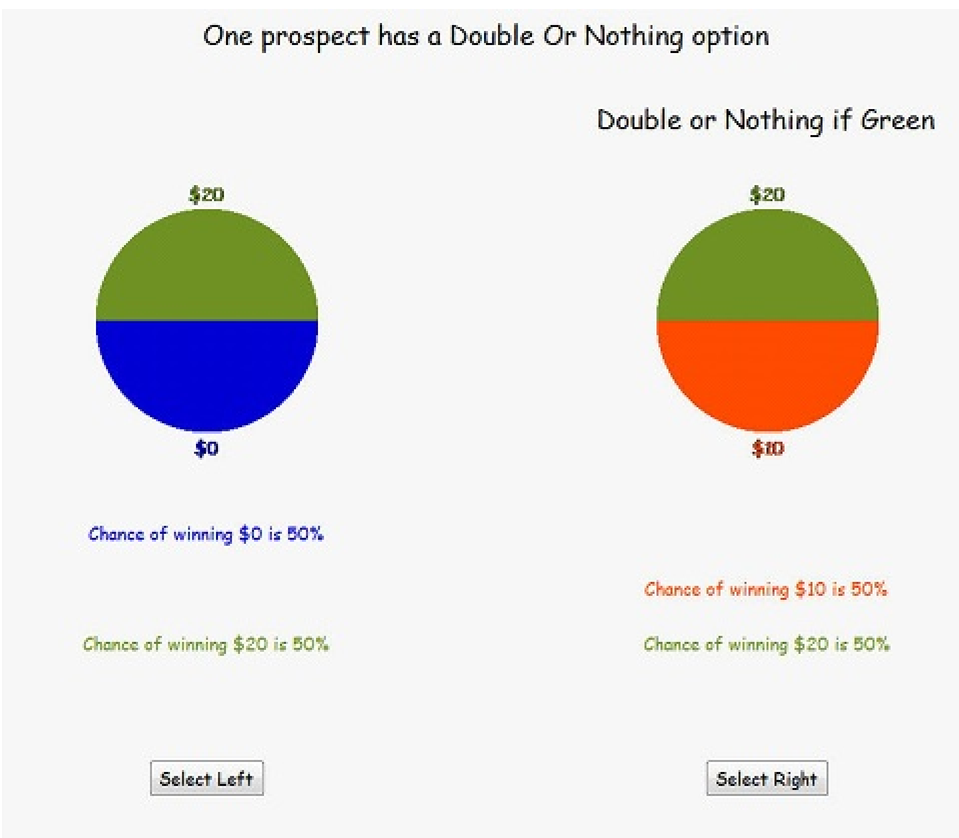
The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

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.

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.

You could also get a pair of prospects in which one of the prospects will give you the chance to play "Double or Nothing." For instance, the right prospect in the following screen image pays "Double or Nothing" if the Green area is selected, which happens if the number drawn is between 51



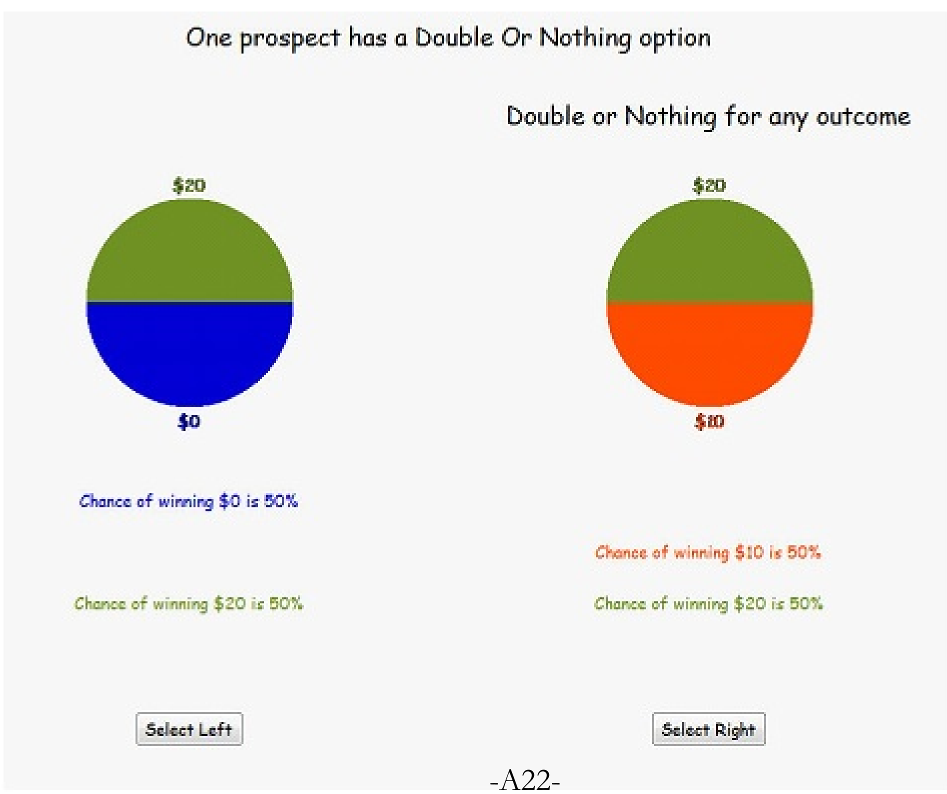
and 100. The right pie chart indicates that if the number is between 1 and 50 you get $10. However, if the number is between 51 and 100 a coin will be tossed to determine if you get double the amount. If it comes up Heads you get $40, otherwise you get nothing. The prizes listed underneath each pie refer to the amounts before any "Double or Nothing" coin toss.

After you have worked through all of the 40 pairs of prospects, raise your hand and an experimenter will come over. You will then roll two 10-sided dice until a number between 1 and 40 comes up to determine which pair of prospects will be played out. Since there is a chance that any of your 40 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, and if necessary you will then toss a coin to determine if you get "Double or Nothing."

For instance, suppose you picked the prospect on the left in the last example. If the random number was 37, you would win $0; if it was 93, you would get $20.

If you picked the prospect on the right and drew the number 37, you would get $10; if it was 93, you would have to toss a coin to determine if you get "Double or Nothing." If the coin comes up Heads then you get $40. However, if it comes up Tails you get nothing from your chosen prospect.

It is also possible that you will be given a prospect in which there is a "Double or Nothing" option no matter what the outcome of the random number. The screen image below illustrates this possibility.



Therefore, your payoff is determined by four things:

- by which prospect you selected, the left or the right, for each of these 40 pairs;
- by which prospect pair is chosen to be played out in the series of 40 such pairs using the two 10-sided dice;
- by the outcome of that prospect when you roll the two 10-sided dice; and
- by the outcome of a coin toss if the chosen prospect outcome is of the "Double or Nothing" type.

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

Appendix D: Structural Econometric Analysis (WORKING PAPER)

Assume that utility of income is defined by

$$U(x) = x^{(1-r)}/(1-r) \tag{D1}$$

where x is the lottery prize and $r \neq 1$ is a parameter to be estimated. For r=1 assume U(x)=ln(x) if needed. Thus r is the coefficient of CRRA: r=0 corresponds to risk neutrality, r<0 to risk loving, and r>0 to risk aversion. Let there be J possible outcomes in a lottery, and denote outcome $j \in J$ as x_j. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i:

$$EU_i = \sum_{j=1,J} [\ p(x_j) \times U(x_j)\ ]. \tag{D2}$$

The EU for each lottery pair is calculated for a candidate estimate of r, and the index

$$\nabla EU = EU_R - EU_L \tag{D3}$$

is calculated, where EU_L is the "left" lottery and EU_R is the "right" lottery as presented to subjects. This latent index, 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,

$$\text{prob(choose lottery R)} = \Phi(\nabla EU) \tag{D4}$$

Even though this "link function" is common in econometrics texts, it forms the critical statistical link between observed binary choices, the latent structure generating the index ∇EU, and the probability of that index being observed. The index defined by (D3) is linked to the observed choices by specifying that the R lottery is chosen when $\Phi(\nabla EU) > \frac{1}{2}$, which is implied by (D4).

The likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of r 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

$$\ln L(r; y, \mathbf{X}) = \sum_i [\ (\ln \Phi(\nabla EU) \times \mathbf{I}(y_i = 1)) + (\ln (1-\Phi(\nabla EU)) \times \mathbf{I}(y_i = -1))\ ] \tag{D5}$$

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

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, with the goal of illustrating how to write explicit maximum likelihood (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"), if needed.

It is also a simple matter to generalize this ML analysis to allow the core parameter r to be a linear function of observable characteristics of the individual or task. We extend the model to be $r = r_0 + R \times \mathbf{X}$, where r_0 is a fixed parameter and R is a vector of effects associated with each characteristic in the variable vector **X**. In effect, the unconditional model assumes $r = r_0$ and estimates r_0. This extension significantly enhances the attraction of structural ML estimation, particularly for responses pooled over different subjects and treatments, since one can condition estimates on observable characteristics of the task or subject.

In our case we also extend the structural parameter to take on different values for the lotteries presented as compound lotteries. That is, (D1) applies to the evaluation of utility for all simple lotteries and a different CRRA risk aversion coefficient r + rc applies to compound lotteries, where rc captures the additive effect of evaluating a compound lottery. Hence, for compound lotteries, the decision maker employs the utility function

$$U(x \mid \text{compound lottery}) = x^{(1-r-rc)}/(1-r-rc) \tag{D1$'$}$$

instead of (D1), and we would restate (D1) as

$$U(x \mid \text{simple lottery}) = x^{(1-r)}/(1-r) \tag{D1$''$}$$

for completeness. Specifying preferences in this manner provide us with a structural test for ROCL. If rc = 0 then this implies that compound lotteries are evaluated identically to simple lotteries, which is consistent with ROCL. However, if rc ≠ 0, as conjectured by Smith [1969] for objective and subjective compound lotteries, then, decision-makers violate ROCL in a certain source-dependent manner, where the "source" here is whether the lottery is simple or compound. As stressed by Smith [1969], rc ≠ 0 for *subjective* lotteries provides a direct explanation for the Ellsberg Paradox, but is much more readily tested on the domain of objective lotteries. Of course, the linear specification r + rc is a parametric convenience, but the obvious one to examine initially.

An important extension of the core model is to allow for subjects to make some *behavioral* errors. The notion of error is one that has already been encountered in the form of the statistical assumption that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery. This assumption is clear in the use of a non-degenerate link function between the latent index ∇EU and the probability of picking a specific lottery as given in (D4). If there were no errors from the perspective of EUT, this function would be a step function: zero for all values of $\nabla EU<0$, anywhere between 0 and 1 for $\nabla EU=0$, and 1 for all values of $\nabla EU>0$.

We employ the error specification originally due to Fechner and popularized by Hey and Orme [1994]. This error specification posits the latent index

$$\nabla EU = (EU_R - EU_L)/\mu \tag{D3$'$}$$

instead of (D3), where μ is a structural "noise parameter" used to allow some errors from the perspective of the deterministic EUT model. This is just one of several different types of error story that could be used, and Wilcox [2008] provides a masterful review of the implications of the alternatives.[35] As $\mu \rightarrow 0$ 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 (D3), where the probability of picking one lottery is given by the ratio of the EU of one lottery to the sum of the EU of both lotteries. Thus μ can be viewed as a parameter that flattens out the link functions as it 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 posits the latent index

$$\nabla EU = ((EU_R - EU_L)/\nu)/\mu \qquad (D3'')$$

instead of (D3′), where ν is a new, 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 pair to lottery choice pair: hence it is said to be "contextual." If the pair contains a compound and a simple lottery, and the subject exhibits different utility functions for each, then the normalizing term ν can potentially depend on the utility used over compound lotteries *and* on the one used for simple lotteries. For the Fechner specification, dividing by ν ensures that the *normalized* EU difference $[(EU_R - EU_L)/\nu]$ remains in the unit interval for each lottery pair. The term ν does not need to be estimated in addition to the utility function parameters and the parameter for the behavioral error term, since it is given by the data and the assumed values of those estimated parameters.

The specification employed here is the source-dependent CRRA utility function from (D1′) and (D1″), the Fechner error specification using contextual utility from (3″), and the link function using the normal CDF from (D4). The log-likelihood is then

$$\ln L(r, rc, \mu; y, \mathbf{X}) = \textstyle\sum_i [(\ln \Phi(\nabla EU)\times \mathbf{I}(y_i = 1)) + (\ln (1-\Phi(\nabla EU))\times \mathbf{I}(y_i = -1))] \qquad (D5'')$$

and the parameters to be estimated are r, rc and μ given observed data on the binary choices y and the lottery parameters in **X**.

It is possible to consider more flexible utility functions than the CRRA specification in (D1), but that is not essential for present purposes. We do, however, consider extensions of the EUT model to allow for rank-dependent decision-making under Rank-Dependent Utility (RDU) models.

The RDU model extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same parametric specification (D1′) and (D1″) considered for source-dependent EUT. To calculate decision weights under RDU one replaces expected utility defined by (D2) with RDU

[35] 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.

$$\text{RDU}_i = \sum_{j=1,J} [\, w(p(M_j)) \times U(M_j) \,] = \sum_{j=1,J} [\, w_j \times U(M_j) \,] \tag{D2′}$$

where

$$w_j = \omega(p_j + \ldots + p_J) - \omega(p_{j+1} + \ldots + p_J) \tag{D6a}$$

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

$$w_j = \omega(p_j) \tag{EDb}$$

for j=J, with the subscript j ranking outcomes from worst to best, and $\omega(\cdot)$ is some probability weighting function.

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

$$\omega(p) = p^{\gamma} \tag{D7}$$

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" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \;\; \forall p$ and hence the "RDU EV" weighted by $\omega(p)$ instead of p has to be less than the EV weighted by p. The rest of the ML specification for the RDU model is identical to the specification for the EUT model, but with different parameters to estimate.

It is obvious that one can extend the probability weighting specification to be source-dependent, just as we did for the utility function. Hence we extend (D7) to be

$$\omega(\, p \mid \text{compound lottery} \,) = p^{\gamma+\gamma c} \tag{D7′}$$

for compound lotteries, and

$$\omega(\, p \mid \text{simple lottery} \,) = p^{\gamma} \tag{D7″}$$

for simple lotteries. The hypothesis of source-independence, which is consistent with ROCL, in this case is that $\gamma c = 0$ and $rc = 0$. We also estimate our models using the inverse-S probability weighting function given by:

$$\omega(p) = p^{\gamma}/(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma} \tag{D8}$$

The source-dependent specification trivially allows for the inverse-S probability weighting function as follows

$$\omega(\, p \mid \text{compound lottery} \,) = p^{\gamma+\gamma c}/(p^{\gamma+\gamma c}+(1-p)^{\gamma+\gamma c})^{1/(\gamma+\gamma c)} \tag{D8′}$$

for compound lotteries, and

$$\omega(p \mid \text{simple lottery}) = p^{\gamma}/(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma} \qquad \text{(D8″)}$$

for simple lotteries.

Analysis of Data from the 1-in-1 Treatment

Assuming homogeneity in preferences, we find that the EUT source-dependent model is rejected in favor of the RDU source-dependent model. Conditional on the RDU source-dependent model we find *no violations of ROCL*. Table D1 shows the parameter estimates of the models we consider. Panel A shows the estimates for the model assuming a power weighting function: the estimates for parameters r, rc, γ and γc are 0.62, 0.14, 0.77 and -0.20, respectively. A test of the joint null hypothesis that rc = γc = 0 results in a *p*-value of 0.25. The same test results in a *p*-value of 0.76 for the case of the inverse-S weighting function. A test of the joint hypothesis that γ=1 and γc=0 results in a *p*-value less than 0.001, whichever probability weighting function is assumed. Thus we reject the hypothesis of that these data are best characterized by the EUT source-dependent model, and conclude that the evidence favors the RDU source-dependent model.

Analysis of Data from the 1-in-40 Treatment.

Assuming homogeneity in preferences, we find that the EUT source-dependent model is rejected in favor of the RDU source-dependent model. Conditional on the RDU source-dependent model we find *violations of ROCL*. Table D2 shows estimates for the parameters of the models of interest. A statistical test for the joint null hypothesis that rc = γc=0 results in a *p*-value less than 0.001 whether or not we assume the Power or Inverse-S probability weighting functions. Finally, a test for the joint hypothesis that γ=1 and γc=0 results in a *p*-value less than 0.001, whichever probability weighting function is assumed. Thus we reject the hypothesis of that these data are best characterized by the EUT source-dependent model, and conclude that the evidence favors the RDU source-dependent model.

Additional References

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).

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., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 162(1), May 2011, 89-104.

Table D1: Estimates of Source-Dependent Models Allowing with 1-in-1 Data

Data from the 1-in-1 treatment (133 observations from 133 subjects × 1 choice).
Estimates from the Fechner error parameter omitted.

A. Source-Dependent RDU with Power Weighting Function (LL=-81.82)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.6182909	.0669022	9.24	0.000	.487165	.7494169
rc	.1410226	.0922965	1.53	0.127	-.0398752	.3219204
γ	.7699848	.193429	3.98	0.000	.3908709	1.149099
γc	-.1955013	.1819329	-1.07	0.283	-.5520832	.1610806

(Ho: rc = γc = 0; p-value = 0.2848)

(Ho: γc = 0 and γ = 1; p-value < 0.001)

B. Source-Dependent RDU with Inverted-S Weighting Function (LL=-81.36)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.4209517	.1168385	3.60	0.000	.1919524	.649951
rc	-.0927005	.1540249	-0.60	0.547	-.3945838	.2091829
γ	.6768297	.1122897	6.03	0.000	.4567461	.8969134
γc	-.1470867	.2065464	-0.71	0.476	-.5519103	.2577369

(Ho: rc = γc = 0; p-value = 0.7634)

(Ho: γc = 0 and γ = 1; p-value = 0.0088)

C. Source-Dependent Version of EUT (LL=-84.24)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.5799425	.073165	7.93	0.000	.4365416	.7233433
rc	.2718091	.1027934	2.64	0.008	.0703378	.4732804

Table D2: Estimates of Source-Dependent Models with 1-in-40 Data

Data from the 1-in-40 treatment (2480 observations from 62 subjects × 40 choices).
Estimates from the Fechner error parameter omitted.

A. Source-Dependent RDU with Power Weighting Function (LL=-1460.57)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.5670508	.028896	19.62	0.000	.5104156	.623686
rc	.1090157	.0175052	6.23	0.000	.0747062	.1433252
γc	1.088315	.080862	13.46	0.000	.9298286	1.246802
γc	-.3957716	.0628587	-6.30	0.000	-.5189724	-.2725707

(Ho: rc = γc = 0; p-value<0.0001)
(Ho: γc = 0 and γ = 1; p-value < 0.0001)

B. Source-Dependent RDU with Inverted-S Weighting Function (LL=-1462.80)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.5594736	.0271264	20.62	0.000	.5063069	.6126404
rc	-.0567367	.0410043	-1.38	0.166	-.1371036	.0236301
γ	1.655437	.1490058	11.11	0.000	1.363391	1.947483
γc	-.9713041	.1425871	-6.81	0.000	-1.25077	-.6918386

(Ho: rc = γc = 0; p-value<0.0001)
(Ho: γc = 0 and γ = 1; p-value < 0.0001)

C. Version of EUT (LL=-1512.45)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
r	.6200649	.0143099	43.33	0.000	.592018	.6481118
rc	.2238311	.0193132	11.59	0.000	.185978	.2616843

Appendix E: Non-parametric Tests (WORKING PAPER)

E1. Choice Patterns Where ROCL Predicts Consistent Choices

Suppose a subject is presented with a given S-C lottery pair, and further assume that she prefers the C lottery over the S lottery. If the subject satisfies ROCL and is separately presented with a second pair of lotteries consisting of the same S lottery and the AE lottery of the previously-presented C lottery, then she would prefer and should choose the AE lottery. Similarly, of course, if she instead prefers the S lottery when presented separately with a given S-C lottery pair, then she should choose the S lottery when presented with the corresponding S-AE lottery pair.

We can construct 15 comparisons of lottery pairs that constitute 15 consistency tests of ROCL. In the 1-in-40 treatment we again must assume that the RLIM is incentive compatible, and we again use data from the 1-in-1 treatment to control for possible confounds created by the RLIM. We must now assume homogeneity in risk preferences for the analysis of behavior in the 1-in-1 treatment, since we are making across-subject comparisons. However, in the next section we will present econometric analysis which allows for heterogeneity in risk preferences and test if a violation of the homogeneity assumption is confounded with a violation of ROCL.

Our hypothesis is that a given subject chooses the S lottery when presented with the S-C lottery pair if and only if the same subject also chooses the S lottery when presented with the corresponding S-AE lottery pair.[36] Therefore, ROCL is satisfied if we observe that the proportion of subjects who choose the S lottery when presented with a S-C pair is equal to the proportion of subjects who choose the S lottery when presented with the corresponding S-AE pair. Conversely, ROCL is violated if we observe unequal proportions of choosing the S lottery across a S-C pair and linked S-AE pair. We do not find evidence to reject the consistency in patterns implied by ROCL in the 1-in-1 treatment, while we do find evidence of violations of ROCL in the 1-in-40 treatment.

Analysis of Data from the 1-in-40 Treatment We use the Cochran Q test coupled with the Bonferroni-Dunn (B-D) correction procedure[37] to test the hypothesis that subjects choose the S lottery in the same proportion when presented with linked S-C and S-AE lottery pairs. The B-D

---

[36] Notice that this is equivalent to stating the null hypothesis using the C and AE lotteries. We chose to work with the S lottery for simplicity.

[37] The B-D method is a *post-hoc* procedure that is conducted after calculating the Cochran Q test. The first step is to conduct the Cochran Q test to evaluate the null hypothesis that the proportions of individuals who choose the S lottery is the same in all 15 S-C and 15 S-AE linked lottery pairs. If this null is rejected the B-D method involves calculating a critical value d that takes into account all the information of the 30 lottery pairs. The B-D method allows us to test the statistical significance of the observed difference between proportions of subjects who choose the S lottery in any given paired comparison. Define p_1 as the proportion of subjects who choose the S lottery when presented with a given S-AE lottery pair. Similarly, define p_2 as the proportion of subjects who chose the S lottery in the paired S-C lottery pair. The B-D method rejects the null hypothesis that $p_1=p_2$ if $|p_1-p_2| > d$. In this case we would conclude that the observed difference is statistically significant. This is a more powerful test than conducting individual tests for each paired comparison because the critical value d takes into account the information of all 15 comparisons. See Sheskin [2004; p. 871] for further details of the B-D method.

procedure takes into account repeated comparisons and allows us to maintain a familywise error rate across the 15 paired comparisons of S-C and S-AE lottery pairs.

We find evidence to reject the ROCL consistency prediction. Table E1 shows the results of the B-D method[38] for each of the 15 paired comparisons. Table E1 provides evidence that with a 5% familywise error rate, subjects choose the S lottery in different proportions across linked S-C lottery pairs and S-AE lottery pairs in two comparisons: Pair 1 vs. Pair 16 and Pair 3 vs. Pair 18. This implies that the ROCL prediction of consistency is rejected in 2 of our 15 consistency comparisons.

We are also interested in studying the patterns of violations of ROCL. A pattern inconsistent with ROCL would be subjects choosing the S lottery when presented with a given S-C lottery pair, but switching to prefer the AE lottery when presented with the matched S-AE pair. We construct 2 × 2 contingency tables that show the number of subjects in any given matched pair who exhibit each of the four possible choice patterns: (i) always choosing the S lottery; (ii) choosing the S lottery when presented with a S-C pair and switching to prefer the AE lottery when presented with the matched S-AE pair; (iii) choosing the C lottery when presented with a S-C pair and switching to prefer the S lottery when presented with the matched S-AE; and (iv) choosing the C lottery when presented with the S-C lottery and preferring the AE lottery when presented with the matched S-AE.

Since we have paired observations, we use the McNemar test to evaluate the null hypothesis of equiprobable occurrences of discordant choice patterns (ii) and (iii) within each set of matched pairs. We find a statistically significant difference in the number of (ii) and (iii) choice patterns within 4 of the 15 matched pairs. Table E3 reports the exact *p*-values for the McNemar test. The McNemar test results in *p*-values less than 0.05 in four comparisons: Pair 1 vs. Pair 16, Pair 3 vs. Pair 18, Pair 10 vs. Pair 25 and Pair 13 vs. Pair 28.[39] Moreover, the odds ratios of the McNemar tests suggest that the predominant switching pattern is choice pattern (iii): subjects tend to switch from the S lottery in the S-AE pair to the C lottery in the S-C pair. The detailed contingency tables for these 4 matched pairs show that the number of choices consistent with pattern (iii) is considerably greater than the number of choices consistent with (ii).

<u>Analysis of Data from the 1-in-1 Treatment</u> We use the Cochran-Mantel-Haenszel (CMH) test to test the joint hypothesis that in all of the 15 paired comparisons, subjects choose in the same proportion the S lottery when presented with the S-C lottery pair and its linked S-AE lottery pair (see Mantel and Haenszel [1959]).[40] If the CMH test rejects the null hypothesis, then we interpret this as evidence of overall ROCL-inconsistent observed behavior. We also use the Fisher Exact test to

---

[38] The Cochran Q test rejected its statistical null hypothesis χ^2 statistic 448.55, 29 degrees of freedom and *p*-value<0.0001.

[39] These violations of ROCL are also supported by the B-D procedure if the familywise error rate is set to 10%.

[40] The proportion of subjects who choose the S lottery when presented with a S-C pair, or its paired S-AE lottery pair, has to be equal within each paired comparison, but can differ across comparisons. More formally, the CMH test evaluates the null hypothesis that the odds ratio of each of the 15 contingency tables constructed from the 15 paired comparisons are jointly equal to 1.

evaluate individually the consistency predicted by ROCL in each of the 15 linked comparisons of S-C pairs and S-AE pairs for which we have enough data to conduct the test.

We cannot reject the ROCL consistency prediction. The CMH test does not reject the joint null hypothesis that the proportion of subjects chose the S lottery when they were presented with any given S-C pair is equal to the proportion of subjects that chose the S lottery when they were presented with the corresponding S-AE pair. The χ^2-statistic for the CMH test with the continuity correction[41] is equal to 2.393 with a corresponding *p*-value of 0.122. Similarly, the Fisher Exact tests presented in Table E2 show only in one comparison the *p*-value is less than 0.05. Therefore, we cannot reject the ROCL consistency prediction in the 1-in-1 treatment. However, as we mentioned previously, this conclusion relies on the assumption of homogeneity in preferences which we can control for in the econometric models.

E2. Choice Patterns Where ROCL Predicts Indifferent Choices

Analysis of Data from the 1-in-40 Treatment The strategy to test the ROCL prediction of indifference in this treatment is different from the one used in the 1-in-1 treatment, given the repeated measures we have for each subject in the 1-in-40 treatment. We now use the Cochran Q test to evaluate whether the proportion of subjects who choose the C lottery is the same in each of the 10 AE-C lottery pairs.[42] A significant difference of proportions identified by this test is sufficient to reject the null prediction of indifference.[43] Of course, an insignificant difference of proportions would require us to additionally verify that the common proportion across pairs the pairs is indeed 50% before we fail to reject the null hypothesis of indifference.

We observe an overall violation of the ROCL indifference prediction in the 1-in-40 treatment. Table E4 reports the results of the Cochran Q test, as well as summary statistics of the information used to conduct the test. The Cochran Q test yields a *p*-value of less than 0.0001, which strongly suggests rejection of the null hypothesis of equiprobable proportions. We conclude that, for at least for one of the AE-C lottery pairs, the proportion of subjects who chose the C lottery is not equal to 50%. This result is a violation of ROCL and we cannot claim that subjects satisfy ROCL and choose at random in all of the 10 AE-C lottery pairs in the 1-in-40 treatment.

Analysis of Data from the 1-in-1 Treatment. We use a generalized version of the Fisher Exact test to jointly test the null hypothesis that the proportion of subjects who chose the C lottery over the

[41] We follow Li, Simon and Gart [1979] and use the continuity correction to avoid possible misleading conclusions from the test in small samples.

[42] The Binomial Probability test is inappropriate in this setting, as it assumes independent observations. Obviously, observations are not independent when each subject makes 40 choices.

[43] For example, suppose there were only 2 AE-C lottery pairs. If the Cochran Q test finds a significant difference, we conclude that the proportion of subjects choosing the C lottery is not the same in the two lottery pairs. Therefore, even if the proportion for one of the pairs was truly equal to 50%, the test result would imply that the other proportion is not statistically equal to 50%, and thus indifference fails.

AE lottery in each of the AE-C lottery pairs are the same, as well as the Binomial Probability test to evaluate our null hypothesis of equiprobable choice in each of the AE-C lottery pairs.

We do not observe statistically significant violations of the ROCL indifference prediction in the 1-in-1 treatment. Roughly 59% of the subjects (19 out of 32 subjects) presented with a AE-C lottery pair chose the C lottery and we cannot reject that this proportion is statistically different from 50%. Table E5 presents the generalized Fisher Exact test for all AE-C lottery pair choices, and the test's *p*-value of 0.342 provides no support for the hypothesis of violations of the ROCL indifference prediction. We see from this test that the proportions are the same across pairs. We now use the Binomial Probability test to see if the proportions are different from 50%. Pooling over all AE-C lottery pairs we find a *p*-value of 18.9% for the null hypothesis that subjects chose the C and the AE lotteries in equal proportions Table E6 shows the Binomial Probability test applied individually to each of the AE-C lottery pairs for which we have observations. We cannot reject the null hypothesis that subjects chose the C and the AE lotteries in equal proportions, as all *p*-values are insignificant at any reasonable level of confidence.

Additional References

Li, Shou-Hua; Simon, Richard M., and Gart, John J., "Small Sample Properties of the Mantel-Haenszel Test," *Biometrika*, 66, 1979, 181-183.

Mantel, Nathan, and Haenszel, William, "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," *Journal of the National Cancer Institute*, 22, 1959, 719-748.

Sheskin, David J., *Handbook of Parametric and Nonparametric Statistical Procedures* (Boca Raton: Chapman & Hall/CRC, 2004; Third Edition).

Table E1: Bonferroni-Dunn Method on Matched Simple-Compound and Simple-Actuarially-Equivalent Pairs

Treatment: 1-in-40

Matching	Proportion of subjects that chose the S lottery in the S-AE pair (p_1)	Proportion of subjects that chose the S lottery in the S-C pair (p_2)	$\lvert p_1 - p_2 \rvert$
Pair 1 vs. Pair 16	0.871	0.387	0.484
Pair 2 vs. Pair 17	0.984	0.952	0.032
Pair 3 vs. Pair 18	0.887	0.629	0.258
Pair 4 vs. Pair 19	0.226	0.21	0.016
Pair 5 vs. Pair 20	0.403	0.290	0.113
Pair 6 vs. Pair 21	0.742	0.661	0.081
Pair 7 vs. Pair 22	0.677	0.548	0.129
Pair 8 vs. Pair 23	0.548	0.548	0
Pair 9 vs. Pair 24	0.258	0.306	0.048
Pair 10 vs. Pair 25	0.919	0.774	0.145
Pair 11 vs. Pair 26	0.581	0.613	0.032
Pair 12 vs. Pair 27	0.565	0.645	0.081
Pair 13 vs. Pair 28	0.871	0.726	0.145
Pair 14 vs. Pair 29	0.742	0.677	0.065
Pair 15 vs. Pair 30	0.387	0.419	0.032

Note: the test rejects the null hypothesis of $p_1=p_2$ if $\lvert p_1-p_2 \rvert > d$. The calculation of the critical value d requires that one first define *ex ante* a familywise Type I error rate (α_{FW}). For α_{FW} = 10% the corresponding critical value is 0.133, and for α_{FW} = 5% the critical value is 0.159.

Table E2: Fisher Exact Test on Matched Simple-Compound and Simple-Actuarially-Equivalent Pairs

Treatment: 1-in-1

Comparison	Total # of subjects in		Proportion of subjects that chose the S lottery in the S-AE pair (π_1)	Proportion of subjects that chose the S lottery in the S-C pair (π_2)	*p*-value for $H_0: \pi_1 = \pi_2$
	S-AE Pair	S-C Pair			
Pair 1 vs. Pair 16	3	2	1	0.5	0.4
Pair 3 vs. Pair 18	6	2	0.5	0	0.464
Pair 5 vs. Pair 20	1	2	0	1	0.333
Pair 6 vs. Pair 21	3	4	0.67	0.5	1
Pair 7 vs. Pair 22	4	9	1	0.56	0.228
Pair 8 vs. Pair 23	3	4	0.33	0.5	1
Pair 9 vs. Pair 24	3	6	0	0.83	0.048
Pair 11 vs. Pair 26	5	9	0.6	0.56	1
Pair 12 vs. Pair 27	5	2	0.8	1	1
Pair 13 vs. Pair 28	4	1	0.5	1	1
Pair 15 vs. Pair 30	3	1	1	0	0.25

Note: Due to the randomization assignment of lottery pairs to subjects, the table only shows the Fisher Exact test for 11 S-AE/S-C comparisons for which there are sufficient data to conduct the test. Setting a critical *p*-value of 0.05, the Bonferroni correction for the eleven pairwise comparisons yields a *p*-value threshold of 0.005 (=0.05/11) for each individual test. As can be seen in the table, none of the individual tests rejects the null hypothesis with this corrected *p*-value threshold.

Table E3: McNemar Test on Matched Simple-Compound and Simple-Actuarially-Equivalent Pairs
Treatment: 1-in-40

Matching	Exact *p*-value	Odds Ratio
Pair 1 vs. Pair 16	<0.0001	0.0625
Pair 2 vs. Pair 17	0.625	0.3333
Pair 3 vs. Pair 18	0.0001	0.0588
Pair 4 vs. Pair 19	1	0.8571
Pair 5 vs. Pair 20	0.1671	0.4615
Pair 6 vs. Pair 21	0.3323	0.5454
Pair 7 vs. Pair 22	0.1516	0.5
Pair 8 vs. Pair 23	1	1
Pair 9 vs. Pair 24	0.6072	0.6667
Pair 10 vs. Pair 25	0.0352	0.25
Pair 11 vs. Pair 26	0.8238	1.222
Pair 12 vs. Pair 27	0.4049	1.555
Pair 13 vs. Pair 28	0.0117	0.1
Pair 14 vs. Pair 29	0.5034	0.6667
Pair 15 vs. Pair 30	0.8388	1.1818

Table E4: Cochran Q Test on the Actuarially-Equivalent Lottery vs. Compound Lottery Pairs

Treatment: 1-in-40

Cochran's χ^2 statistic (9 d.f) = 86.090

p-value < 0.0001

Data

AE-C Lottery Pair	Observed # of choices of C lotteries (out of 62 observations)
31	48
32	46
33	34
34	26
35	24
36	18
37	34
38	35
39	49
40	28

Table E5: Generalized Fisher Exact Test on the Actuarially-Equivalent Lottery vs. Compound Lottery Pairs

Treatment: 1-in-1

Fisher Exact *p*-value = 0.342

AE-C Lottery Pair	Observed # of choices of AE lotteries	Observed # of choices of C lotteries	Total
31	0	1	1
32	0	3	3
33	2	5	7
36	4	1	5
37	1	1	2
38	2	1	3
39	1	4	5
40	3	3	6
Total	13	19	32

Note: due to the randomization assignment of lottery pairs to subjects, there were no observations for pairs 34 and 35.

Table E6: Binomial Probability Tests on Actuarially-Equivalent Lottery vs. Compound Lottery Pairs

Treatment: 1-in-1

AE-C Lottery Pair	Total # of observations	Observed # of choices of C lotteries	Observed proportion of choices of C lotteries (p)	*p*-value for H_0: p = 0.5
32	3	3	1	0.25
33	7	5	0.714	0.453
36	5	1	0.2	0.375
37	2	1	0.5	1
38	3	1	0.333	1
39	5	4	0.8	0.375
40	6	3	0.5	1

Note: due to the randomization assignment of lottery pairs to subjects there were no observations for pairs 34 and 35 and only 1 observation for pair 31.

Appendix F: The Rank-Dependent Utility Model (WORKING PAPER)

Assume that utility of income in an elicitation is defined by

$$U(x) = x^{(1-\rho)}/(1-\rho) \tag{F1}$$

where x is the lottery prize and r≠1 is a parameter to be estimated. For ρ=1 assume U(x)=ln(x) if needed. Thus ρ is the coefficient of CRRA for an EUT individual: ρ=0 corresponds to risk neutrality, ρ<0 to risk loving, and ρ>0 to risk aversion.[44]

Let there be J possible outcomes in a lottery defined over objective probabilities commonly implemented in experiments. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i:

$$EU_i = \sum_{j=1,J} [\, p(x_j) \times U(x_j) \,]. \tag{F2}$$

The RDU model of Quiggin [1982] extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same parametric specification (F1) considered for EUT.[45] To calculate decision weights under RDU one replaces expected utility defined by (F2) with RDU

$$RDU_i = \sum_{j=1,J} [\, w(p(M_j)) \times U(M_j) \,] = \sum_{j=1,J} [\, w_j \times U(M_j) \,] \tag{F3}$$

where

$$w_j = \omega(p_j + ... + p_J) - \omega(p_{j+1} + ... + p_J) \tag{F4a}$$

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

$$w_j = \omega(p_j) \tag{F4b}$$

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

We consider two popular probability weighting functions. The first is the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ:

$$\omega(p) = p^\gamma \tag{F5}$$

So γ≠1 is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \;\; \forall p$ and hence the "RDU EV" weighted by ω(p) instead of p has to be less than the EV weighted by p.

[44] Of course, risk attitudes under RDU depend on more than the curvature of the utility function.

[45] To ease notation we use the same parameter ρ because the context always make it clear if this refers to an EUT model or a RDU model.

The second probability weighting function is the "inverse-S" function popularized by Tversky and Kahneman [1992]:

$$\omega(p) = p^{\gamma} / (p^{\gamma} + (1-p)^{\gamma})^{1/\gamma} \qquad \text{(F6)}$$

This function exhibits inverse-S probability weighting (optimism for small p, and pessimism for large p) for $\gamma<1$, and S-shaped probability weighting (pessimism for small p, and optimism for large p) for $\gamma>1$.