Individual Choice and Risk Aversion in the Laboratory: A Reconsideration

by

Glenn W. Harrison, Eric Johnson, Melayne M. McInnes & E. Elisabet Rutström†

May 2005

Abstract. Experimental methods have been used to examine individual choices for consistency with expected utility theory (EUT). However, previous tests of EUT have not recognized the confounding influence of indifference. Properly identifying indifference requires knowledge of the risk attitudes of each subject. We independently estimate the risk attitudes of a sample of subjects, and then use those estimates to condition our analysis of their responses to two popular tests of EUT. We find that the distribution of risk attitudes places many of these subjects at precisely the place at which indifference makes tests of EUT meaningless. The argument we make here is that previous dismissals of EUT are premature. The most important conclusion we draw is that EUT is hard to test, and that its main weakness as a theory may be the difficulty of undertaking operationally meaningful tests of it. We also show that, once the risk attitudes of the subjects are known, it is possible to design tests of EUT that are meaningful, in the sense of not being confounded by indifference. Based on such a test we find significant evidence of inconsistent choices.

Key words: expected utility theory, experimental economics, risk aversion

† Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström); Department of Economics, Moore School of Business, University of South Carolina, USA (Johnson and McInnes). E-mail contacts: GHARRISON@BUS.UCF.EDU, EJOHNSON@MOORE.SC.EDU, MCINNES@MOORE.SC.EDU and ERUTSTROM@BUS.UCF.EDU. We thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/MRI 9871019 and NSF/POWRE 9973669 to Rutström and grant NSF/SES 0213974 to McInnes. We are grateful to Maribeth Coller, Graham Loomes, Chris Starmer, Robert Sugden and Melonie Sullivan for comments, as well as participants at the 11th Conference on the Foundations & Applications of Utility, Risk & Decision Theory in Paris. Supporting data and instructions are available at the ExLab Digital Archive at http://exlab.bus.ucf.edu.

The position of Expected Utility Theory (EUT) as the leading theory of economic choice has been undermined by experimental tests that have produced contradictory and anomalous results. In these experiments, pairs of lotteries have been carefully constructed so that only specific patterns of choice are consistent with EUT. However, other patterns of choices have been frequently observed. We show that previous research has not properly accounted for a potentially important confound, namely that subjects may have been indifferent when making their choices.

EUT does not make predictions for how choices are made when decision makers are indifferent. Data generated based on indifference choices are therefore not only uninformative, as they would be if they are randomly made, but may also confound inferences about preference orderings if they are made in some systematic way. Risk attitudes affect the certainty equivalents of the lotteries in typical tests of EUT, and plausible levels of risk aversion may result in *any* choice being consistent with the EUT hypothesis due to indifference.

Having demonstrated the *potential* problem theoretically, we then show that the empirically relevant distribution of risk aversion of subjects makes this criticism substantively significant for a wide class of tests of EUT. The distribution of risk attitudes elicited from our subject pool demonstrates that certain tests of theory are difficult to undertake with our sample. Since the distribution of risk attitudes in our particular sample is common across many other samples, this difficulty is likely a general one.

However, since indifference is a consequence of the interaction between a subjects' risk attitude and the parameters of the lottery task, we show that it is quite easy to devise specific tests of EUT that do not suffer from the indifference confound *once the risk attitudes are known*. Performing such a test we find evidence of violations of EUT. We do not offer these conclusions as arguments in favor of existing alternative decision theories. Any theory that allows for individual variation in risk attitude, or any other parameter of the theory that affects the propensity to be indifferent, must meet the same challenges for designing operationally meaningful tests.[1] We propose a

[1] For example, the same points that we make about the role of risk aversion and tests of EUT can be made about the specification of the probability weighting function and tests of prospect theory. Furthermore, since prospect

"complementary slack" experimental design: if one test of EUT is confounded by indifference, then the other will almost certainly not be.[2]

Our approach realistically presumes that the observer does not know the risk attitude of the subject precisely, and must estimate it somehow by observing the behavior of the subject. Since that process of inference is not perfect, there will be some error in the observer's estimate of the subject's risk attitudes. If that error is sufficiently large, EUT is virtually impossible to test because it is virtually impossible to refute. Thus we do not rely on the *subject* making errors, as has been the focus of much important work seeking to account for violations of EUT.[3] Instead, our approach is logically prior to such concerns, since one cannot identify that a violation has occurred unless one knows that the subject is not indifferent. On the other hand, we see our approach as eventually complementary to the "EUT with errors" approach once the observer has been able to correctly identify the extent of violations.

In Section 1 we review the need for estimates of risk attitudes in tests of EUT. We revisit two classical rational choice problems, and show how the *observer's inferences* from them are affected by how the assumed risk attitude of the subject may make him indifferent. In section 2 we examine some careful experimental tests of these problems by Cubitt, Starmer and Sugden [1998] and Grether and Plott [1979], and experimental procedures for characterizing risk attitudes due to Holt and Laury [2002]. In Section 3 we discuss how variants of these three experiments can be conducted on the same subject pool in order to provide meaningful tests of EUT. We then present the

---

theory does not constrain the form of the "value function," that function can also have the free parameter of risk aversion considered here. Thus prospect theory has at least two free parameters, necessarily making it harder to refute than EUT, which only has a subset of those two. More generally, we expect that any "horse race" of the predictive power of alternative theories will suffer from serious problems of being unable to discern a winner from the fog of noise generated by imprecise measurements of risk attitudes.

[2] The "almost certainly" qualifier is needed because it is possible that we have such imprecise estimates of risk attitudes that the confound applies to both tests. This is an empirical matter, and we show that the qualifier is not needed for our sample and the methods we use to elicit risk attitudes.

[3] See Harless and Camerer [1994], Hey and Orme [1994] and Loomes and Sugden [1995] for the first wave of empirical studies including some formal stochastic specification in the version of EUT tested. There are, in fact, several species of "errors" in use, reviewed by Hey [1995], Loomes and Sugden [1995], Ballinger and Wilcox [1997], and Loomes, Moffatt and Sugden [2002]. Some place the error at the final choice between one lottery or the other after the subject has decided deterministically 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.

distribution of estimated risk attitudes of our subject pool and examine its implications for subjects' preferences over lottery choices taken from the choice experiments. In Section 4 we briefly review comparable distributions of elicited risk attitudes in other samples, to assess how empirically general our conclusions are.

1. Risk Aversion and Tests of EUT

A. Risk Aversion and the Common Ratio Tests of EUT

The most popular tests of the axioms of traditional EUT rely on observed choices by individuals over pairs of lotteries. By designing choice situations involving a comparison of choices across lotteries, the choice in one could be used to infer the risk attitude of the subject and the choice in the second pair would then provide a test of EUT, conditioned on the revealed risk attitude. Consider the Common Ratio (CR) test, for example, which would seem to be independent of risk aversion. Subjects are first asked to choose between two lotteries, A and B. Then one constructs two additional compound lotteries, A* and B*, by adding a front end probability q of winning zero to lotteries A and B. That is, A* offers a $(1-q)$ chance to play lottery A and a q chance of winning zero. Subjects choosing A over B and B* over A*, or choosing B over A and A* over B*, are said to violate EUT.

For a specific example, suppose Lottery A consists of prizes \$0 and \$30 with probabilities 0.2 and 0.8. Lottery A has expected utility $[0.2 \times U(\$0)] + [0.8 \times U(\$30)]$, where $U(m)$ denotes the utility of final monetary prize m. Lottery B, consisting of prizes \$0 and \$20 with probabilities 0 and 1, has expected utility $[0 \times U(\$0)] + [1 \times U(\$20)] = U(\$20)$. Let $q = 0.25$, and generate lotteries A* and B* from A and B as defined above. This example is said to provide a test of EUT independently of the risk attitude of the subject, since under EUT the risk attitude only affects the specification of the utility function. Thus, risk aversion will determine whether one prefers A to B, but it cannot change the preference as one moves from lottery pair AB to lottery pair A*B*.

To show precisely how risk aversion does matter, assume that risk attitudes can be characterized by the popular Constant Relative Risk Aversion (CRRA) function, $U(m) = (m^{1-r})/(1-r)$, where r is the CRRA coefficient. With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(m) = \ln(m)$. The certainty equivalents of the lottery pairs AB and A*B* as a function of r are shown in the left and right upper panels respectively of Figure 1.[4] The CRRA coefficient ranges from -0.5 (moderately risk loving) up to 1.25 (very risk averse), with a risk-neutral subject at $r = 0$. The certainty equivalent of lottery B, which offers $20 for sure, is the horizontal line in the left panel. The certainty equivalents of A, A* and B* all decline as risk aversion increases.

Problem areas for testing EUT in this manner are immediately apparent by inspection of Figure 1. The first problem is caused by the fact that the preferred choice changes at CRRA coefficients "close" to 0.45, and that at or around this point the subject would be indifferent between A and B *and* between A* and B*. Thus an *outside observer* has to be able to claim with some level of statistical confidence that the subject does *not* have a CRRA coefficient of 0.45 in order for any choice pattern to be useable as a test of the validity of EUT. If this claim cannot be rejected, then one cannot logically conclude that an observed choice pattern violates EUT. Whether we can reject this claim or not is an empirical and inferential matter, not a matter of *a priori* logic since it pertains to the subjective preferences of the subject over risk.[5] The possibility that we cannot reject the claim is not an idle one, since this value is close to the average CRRA estimate for many samples

---

[4] Other specifications of the utility function imply different values for the certainty equivalents of the lotteries. Our general point is not affected, however, since the certainty equivalent for a lottery is decreasing in the risk attitude of the subject for all well-behaved functional forms. Therefore our arguments hold generally, and the use of CRRA only serves as an illustration.

[5] Rabin [2000] examines the theoretical role of risk aversion and EUT, and argues that EUT must be rejected for individuals who are risk averse at low monetary stakes. If true, then further tests of EUT are not needed for those individuals who are found to be risk averse in these low stake lottery choices. He proves a calibration theorem showing that if individuals are risk averse over low stakes lotteries then there are absurd implications about the bets those individuals will accept at higher stakes. Following the interpretation of these arguments by Cox and Sadiraj [2004] and Rubinstein [2002], a problem for EUT does indeed arise if (a) subjects exhibit risk aversion at low stake levels, *and* (b) one assumes that utility is defined in terms of terminal wealth. If, on the other hand, one assumes utility is defined over income, this critique does not apply. A close reading of Rabin [2000; p. 1288] is consistent with this perspective, as is the model proposed by Charness and Rabin [2002] to account for experimental data they collect. Whether or not one models utility as a function of terminal wealth (EUTw) or income (EUTi) depends on the setting. Both specifications have been popular. The EUTw specification was widely employed in the seminal papers defining risk aversion and its application to portfolio choice. The EUTi specification has been widely employed by auction theorists and experimental economists testing EUT, and it is the specification we employ here.

of experimental subjects.[6] Of course, there is nothing magical about the value 0.45: it just happens to be the CRRA value *in the context of these lottery choices* that makes the subject indifferent between the two choices.

The second problem, a relaxed version of the first problem, is that small perturbations either side of 0.45 cause similarly small perturbations in the certainty-equivalent differences Since it is reasonable to assume that subjects have some perceptual threshold of payoff differences below which they are indifferent, many subjects with risk attitudes either side of $r = 0.45$ may still be indifferent between the two options.[7] Further, the certainty-equivalents of A* and B* remain small for *all* values of r above 0.45. This problem is illustrated in the lower panels of Figure 1. For example, if $r = 0.75$ the certainty equivalent of A* is only $0.29 and the certainty-equivalent of B* is only $0.37. There is a formal difference in favor of B*, but it is only 8 cents. Again, this particular value of the CRRA coefficient is not extraordinary, and is observed in experimental samples. At high levels of risk aversion, the difference in certainty equivalents between the lotteries A* and B* effectively collapses to zero.[8]

The presence of indifference has a confounding influence on tests of EUT in two ways. First, if there is an asymmetry in the proportion of indifferent subjects across tasks, observations may yield different choice proportions if subjects simply employ some random device for decision making when indifferent. Second, when economic factors do not hold sway, it is perfectly plausible that other factors dominate (e.g., left-right or top-down biases, well known in the field of survery design), and we would not necessarily expect to see indifference reflected in random choices. We prefer to remain open-minded about the presence of non-monetary motivations, arguing instead that the lack of saliency in the monetary instrument may lead to *either* noise *or* to systematic choice patterns.

The implication of Figure 1 is that observations of risk-loving or risk-neutral subjects would

---

[6] See Holt and Laury [2002] for reviews.

[7] Hey and Orme [1994] estimate such a threshold parameter (τ).

[8] It does not collapse to zero as a formal matter, but the numerical differences become so small that they are not worth arguing over. In this instance "small" means really small: less than one-hundredth of a penny.

be most informative on the validity of EUT *in the context of these lottery choices* since the difference between certainty equivalents is more likely to be salient for both paired choices and indifference will be less of a problem. For example, for $r = -0.26$, the difference between certainty equivalents is $5.13 for lottery pair AB and $1.71 for lottery pair A*B*.

Figure 1 also shows that the difference in certainty equivalents between lotteries A* and B* is tiny by comparison to the difference in certainty equivalents between lotteries A and B, complicating comparisons of choice patterns between the two even when it can be assumed that indifferent subjects choose randomly. Thus, the strength of the incentives to reveal the preferred alternative varies across the choice pairs, giving rise to indifference in one but not the other. For a sample drawn from a risk averse population we should then expect to see a fairly low proportion of subjects selecting the risky lottery A. If they are risk averse enough the proportion will fall well below 50%. A sample from the same population that faces the A*B* choice, however, would be indifferent and, using a random choice device, for example, would exhibit a 50/50 choice. Thus, without accounting for the effect of indifference we would have an apparent EUT violation in which the EUT inconsistent combination of B then A* is as likely as the consistent B then B*.

Since the most common finding in tests of the CR effect and the Allais paradox[9] is that subjects prefer lottery B in the first lottery choice pair, we can infer that most of the subjects have risk attitudes in the range from 0.45 and above. Given our earlier discussion, it should then come as no surprise that about 50% of those who chose B in the first pair chose A* in the second. In the experiments of Cubitt, Starmer and Sugden [1998], for example, 71% of subjects choose B in a problem of the first kind, and in the second lottery pair only 38% chose B*.[10]

B. Risk Aversion and the Preference Reversal Tests of EUT

The effects of risk aversion on our ability to test EUT depend on the particular lottery pairs

[9] There are several variants of each of these tests. Most differ in terms of how the choices are presented to subjects, but are formally identical under the null hypothesis of EUT. Those differences can lead to differences in behavior under alternative hypotheses of choice behavior.

[10] These numbers are for their problems 2 and 4, respectively.

used. Consider the certainty-equivalents of the paired lottery choices in the classic Preference Reversal (PR) experiments of Grether and Plott [1979], shown in Figure 2. In this experiment, subjects are asked to make direct choices over 6 lottery pairs, and separately to value each of the 12 lotteries underlying these lottery pairs. The "reversal" phenomenon refers to the inconsistency between the direct preferences elicited from the 6 lottery pairs and the implied preferences obtained by comparing the cardinal valuations of each lottery in the pair.

In the literature on preference reversals the term "P-bets" refers to lotteries that have a relatively high *probability* of winning a relatively low monetary prize. The alternative bets are called "$-bets," because they have a lower probability of winning than the paired P-bet, but a higher *dollar* prize if the subject wins. Choice pair #1, for example, consists of the P-bet offering $4 with probability 35/36 and a loss of $1 with probability 1/36, and the $-bet offering $16 with probability 11/36 and a loss of $1 with probability 25/36. In this case the expected values are $3.86 and $3.85, respectively. The other five lottery pairs are similar.[11]

Figure 2 shows again that there are ranges of the CRRA coefficient for which individuals are likely to be indifferent between the two choices offered in each of the paired lottery choices.[12] Because the lottery pairs in these experiments have virtually identical expected values, the difference in certainty equivalents is zero if the CRRA coefficient is close to $r = 0$.[13] Thus we get a dramatic difference in the effect of risk aversion between the CR choices shown in Figure 1 and these "Grether & Plott" lottery choices: in one case the CRRA coefficient generating indifference is 0.45, and in another case it is 0.

[11] The loss probability is the complement of the win probability. Choice pair #2 pays $2 with probability 29/36 and -$1.50 otherwise, or $9 with probability 7/36 and -$0.50 otherwise; choice pair #3 pays $3 with probability 34/36 and -$2 otherwise, or $6.50 with probability 18/36 and -$1 otherwise; choice pair #4 pays $4 with probability 32/36 and -$0.50 otherwise, or $40 with probability 4/36 and -$1 otherwise; choice pair #5 pays $2.50 with probability 34/36 and -$0.50 otherwise, or $8.50 with probability 14/36 and -$1.50 otherwise; and choice pair #6 pays $2 with probability 33/36 and -$2 otherwise, or $5.00 with probability 18/36 and -$1.50 otherwise.

[12] Figure 2 is drawn based on the inflation-adjusted Grether and Plott [1979] lottery values used in our experiments, described later. Lottery pair #6 differs slightly from the original due to an inadvertent change in the instructions. For this lottery pair, the P-bet is preferred to the $-bet if $r > 1.11$, while for the original Grether and Plott lottery values for pair #6 the P-bet is preferred to the $-bet when $r > 0$. Our references in the text are to the first 5 pairs in Figure 2, unless otherwise noted.

[13] For pair #6 the point of indifference is $r = 1.11$, as explained above.

Allowing for the more relaxed definition of indifference based on small perturbations in the certainty-equivalent differences, Figure 2 shows that the sensitivity of the test of EUT again varies across the six lottery choices, just as it varied with the two lottery choices in Figure 1. The flatter the line showing the difference in certainty-equivalents, the larger is the range of CRRA coefficients that would result in indifference.

The fact that indifference is induced by different risk attitudes across lotteries is actually good news for testing EUT, since it means that one can, in principle, design lottery pairs that provide informative tests. In practice this requires both that we devise some instrument to measure risk attitudes that is deemed reliable and robust, and that this instrument provides empirical estimates of risk aversion that are precise enough so that one is able to rule out the problematic values. Nevertheless, our main point does not depend on the ability to elicit precise and reliable risk attitude coefficients. The presence of indifference as a confound in tests of EUT is not a function of the researcher's ability to correctly identify the observations that correspond to indifference. The ability to test EUT in meaningful ways does depend on such identification, however.

The rationale for our experimental design can now be easily stated: in order to test EUT we must find out more precisely where the subjects are in terms of this possible array of CRRA coefficients. We propose a design, following Holt and Laury [2002], that will allow the researcher to infer a the risk attitude coefficient for each subject. Given the distribution of risk attitudes, we can ascertain which lottery tasks provide meaningful tests of EUT on our sample. Conditioning on the risk attitudes of our experimental subjects we redo both the CR test of CSS and the PR choice tests of GP. We then evaluate EUT for those subjects that are *not* predicted to be indifferent in these tests.

2. Previous Experiments

A. Common Ratio Tests of EUT

Previous experiments have dealt with several design problems, apart from the one that concerns us. These problems arise when asking a subject to give two "real responses" in the lab. First, there might be wealth effects, or expected wealth effects, when the earnings from one lottery affect valuations for the second lottery. Second, if one alternatively picks out one choice at random to pay the subject, one is assuming that one of the axioms of EUT (independence) is correct. If it is not, then this random payoff device can generate inconsistent preferences even if the underlying preferences are consistent. Third, there could be order effects, such that responses to the first task are systematically different from responses to the second task, irrespective of preference differences.[14]

The most popular solution, independently developed by Conlisk [1989] and Cubitt, Starmer and Sugden [1998] (CSS), is to have each subject give only one choice response, and to then *assume that different samples have on average the same preference structures*.[15] This solution requires that one recruit relatively large samples, so that randomization over the characteristics that presumably drive differences in preferences is sufficient to ensure that on average there is no sampling bias in favor of one lottery or another. This approach nicely solves the problems of wealth effects and having to assume the validity of the random lottery device.

An elegant design for testing CR problems is provided in CSS.[16] The first lottery problem is

[14] Harrison, Johnson, McInnes and Rutström [2005b] and Holt and Laury [2005] show that order effects are significant in the elicitation of risk attitudes, which is immediately relevant for our overall design.

[15] Another solution has been to bite the bullet of wealth effects and test statistically for their presence. This can be difficult to do, since it relies on the estimation of effects from lagged endogenous variables, and specific parameterizations of the way in which accumulated wealth affects the utility function. A related solution is to exogenously vary the initial wealth of subjects and see if it affects decisions. Yet another solution is to exploit the structure of specific alternatives to traditional EUT, and devise tests that use the random payment device but that are still valid under that specific alternative theory.

[16] They present subjects with five different problems in order to test specific theories as alternatives to EUT. For our purposes we seek the "toughest test" of EUT, and therefore look to the set of responses in all five problems of CSS that provide the biggest difference in response. This suggests that we use their Problems 2 and 4; however, their Problem 2 is an extremely expensive task to implement, given that one loses three subjects on average for every subject that actually provides a useable response. We therefore employed Problem 1 instead of Problem 2 for the AB choice shown in Figure 1, and compared responses from it to Problem 4, which is the A*B* choice in Figure 1. In their Problem 1 CSS observed only 38% of subjects choosing the risky option, A, while in their Problem 4 nearly 68% chose A*. The difference between responses in Problems 1 and 4 is indeed statistically significant using the CSS samples of 50

phrased in our instructions as follows:

> There are 100 numbers on the die, numbered 1-100. We will roll the die whenever a number is needed. Choose either option A or option B:
>
> **Option A**: We will pick a number by rolling the die. If it is numbered 1-20, you will receive nothing. If it is numbered 21-100, you will receive $30.
>
> **Option B**: No number will be drawn from the die. You will receive $20.

The second lottery problem, presented to a different set of subjects, is as follows:

> **Option A***: We will pick a number by rolling the die. If it is numbered 1-75, you will receive nothing. If it is numbered 76-100, we will pick another number by rolling the die again. If it is numbered 1-20, you will receive nothing. If it is numbered 21-100, you will receive $30.
>
> **Option B***: We will pick a number by rolling the die. If it is numbered 1-75, you will receive nothing. If it is numbered 76-100, you will receive $20.

Comparing choices across these two treatments CSS find evidence of choice inconsistencies. The proportion of subjects who choose the safe lottery in the A*B* task (38%) is much smaller than the same proportion in the AB task (71%). As long as indifference can be ruled out, this evidence is clearly unfavorable to EUT. Unfortunately, their data do not allow one to rule out indifference: that provides us with the agenda for our new experiments.

B. Preference Reversal Tests of EUT

The results from the original Grether and Plott [1979] experiments with real monetary incentives are reported in the bottom panel of Figure 3.[17] Each subject could make up to 6 preference reversals, that is, choosing the P-bet over the $-bet but valuing the $-bet more than the P-bet or the reverse. The modal number of reversals was only 1, but many subjects made 2 or 3 reversals, such that the average is 2.08 per subject with a standard deviation of 1.4 reversals per subject. EUT strictly predicts zero reversals, of course. But if all subjects were risk-neutral, then any reversal literally involves zero opportunity cost in terms of expected utility. Without knowledge that these subjects are not risk neutral, one cannot say more about the validity of EUT. Of course, that is

and 48 (p = 0.005 in a two-tailed Fisher Exact test).

[17] The raw data were obtained from Charles Plott, and differ slightly from those reported in Grether and Plott [1979]. There appear to be more subjects in the data files than are reported in the original study. However, the results are completely consistent in terms of the conclusions.

the objective of our new experiments.

One concern with the PR test of EUT is that it relies on the ability of the auction procedure to elicit selling prices. Harrison [1992] argues that this procedure provides poor incentives for truthful revelation of values, and that this problem is independent of risk attitudes. We need not rely on the selling prices to test EUT, becasue PR responses provide an inherent test which relies only on the direct binary choices. Given the risk attitude of the subject, and assuming that the subject is not indifferent, the subject should choose consistently across the direct binary choices.[18] Thus, any given subject should choose the P-bet *or* the $-bet in all cases.[19] Because each subject is observed to make 6 P-bet versus $-bet choices, the subject can have at most 6 consistent choices and no less than 3 consistent choices.[20] The top panel of Figure 4 displays the results of examining the data of Grether and Plott [1979] using this metric. EUT predicts that the subject should make 6 consistent choices, assuming strict preference. The worst outcome for EUT is if the subject only makes 3 choices, again assuming the subject is not indifferent. We observe a poor outcome for EUT, taking these results at face value, with the vast majority of subjects making only 3 or 4 consistent choices. However, as noted in the preceding section we need to be able to additionally say that these are not risk neutral subjects before we can make a claim that these data violate EUT. This is the issue we address with our new experiments.

To reiterate, the overall design calls for both CR and PR tests. Unless subjects' risk attitudes are so imprecisely estimated that we cannot rule out the hypothesis that they are indifferent to both, one of these tests will provide an operationally meaningful test of EUT precisely when the other cannot. And if we cannot estimate risk attitudes precisely enough to say whether a subject is risk-neutral or risk averse, specifically $r = 0$ or $r = 0.45$, then we have no business calling on these lottery

[18] Since we know the risk attitude of the subject, we can further test if the subject chooses consistently *and* if they choose the predicted choice. We can also evaluate the cost of any inconsistencies to that subject, measured in terms of the foregone certainty equivalent of an inconsistent choice. We examine these issues in Harrison, Johnson, McInnes and Rutström [2005c].

[19] In this test we definitely exclude pair #6, for reasons explained above.

[20] The minimum number of consistent choices occurs if the subject chooses 3 P-bets and 3 $-bets. Either choosing one additional P-bet or one additional $-bet results leads to a pattern of {4 P-bets, 2 $-bets} or {2 P-bets, 4 $-bets} which would be tallied as 4 consistent choices.

choice tasks to reliably discriminate between alternative theories.

C. Measures of Risk Aversion

In order to remove the confounding influence of indifference it is important to have a good instrument for assessing subjects' risk attitudes. Holt and Laury [2002] (HL) devise a simple experimental measure for risk aversion using a multiple price list design. Each subject is presented with a choice between two lotteries, which we can call A or B. Table 1 illustrates the basic payoff matrix presented to subjects. The first row shows that lottery A offered a 10% chance of receiving $2 and a 90% chance of receiving $1.60. The expected value of this lottery, EV^A, is shown in the third panel as $1.64.[21] Similarly, lottery B in the first row has chances of payoffs of $3.85 and $0.10, for an expected value of $0.48. Thus the two lotteries have a relatively large difference in expected values, in this case $1.17. As one proceeds down the matrix, the expected value of both lotteries increases, but the expected value of lottery B becomes greater than the expected value of lottery A.

The subject is asked to choose A or B in each row. One row is later selected at random for payout for that subject. The logic behind this test for risk aversion is that only risk-loving subjects would take lottery B in the first row, and only risk-averse subjects would take lottery A in the second-to-last row. Arguably, the last row is simply a test that the subject understood the instructions, and has no relevance for risk aversion at all. A risk neutral subject should switch from choosing A to B when the EV of each is about the same, so a risk-neutral subject would choose A for the first four rows and B thereafter.

HL examine two main treatments with 212 subjects. The first is designed to measure the effect of varying incentives. They vary the scale of the payoffs in the matrix shown in Table 1 by multiplying the payoffs by 20, 50, or 90. Thus, Table 1 shows the scale of 1. The second treatment in the HL design is the effect of hypothetical payoffs. We adapt the HL design for our purposes as described below. We only vary the stakes by a factor of 10, thereby matching the stakes in the EUT

[21] The EV columns of Table 1 were not presented to subjects.

lotteries used, and we do not include a hypothetical treatment.[22]

3. New Experiments

Our experimental design is described in Figure 5. We conducted experiments to elicit each subject's attitudes toward risk. We also conducted experiments with the same subjects that will allow us to test for violations of EUT using CR tests and PR tests after controlling for risk aversion. Because we have some knowledge of the risk attitudes of our subjects we can test EUT both on our entire sample and then on the sample remaining after we remove subjects suspected of being indifferent.

To avoid possible intra-session effects, only one experiment was run in each session. Then the same subjects were contacted again by e-mail and invited to participate in subsequent experiments that were separated by at least one week. Students were recruited from the University of South Carolina in late 2002 and early 2003.[23] In total, 178 subjects participated in a risk aversion experiment (RA) and 176 participated in the CR experiment. The order of these two experiments was varied. Overall, there were 152 subjects for whom we can match results from the risk aversion test to the CR tests. For 90 subjects we can match responses for all three experiments: the RA test, the CR test, and the PR test. No attempt was made to screen subjects for subsequent experiments based on their choices in earlier experiments.

Our subjects received a fixed show-up fee of $5 in each of the three experiments, consistent with our standard procedures.[24] This is a constant across all subjects, and does not vary with the decisions the subjects faced. No subject faced losses.

---

[22] For a risk neutral person, the difference in certainty equivalents are the expected value differences shown in Table 1multiplied by a factor of ten. For a more risk averse person, for example a person with a CRRA coefficient of 0.45, the HL choice pairs still provide sizeable differences in certainty equivalent values. The certainty equivalent differences range from $13.90 for the first choice pair to -$18.50 for the last choice pair.

[23] Some subjects participated in two sessions over this time period. Harrison, Johnson, McInnes and Rutström [2005a] report tests with this population that risk attitudes were temporally stable over this time horizon.

[24] The subjects in CSS were recruited in lectures for experiments run during the usual lecture time and they received no show-up fee. In our case, subjects were recruited via ExLab (http://exlab.bus.ucf.edu), consisting of a combination of e-mail alerts and online registration schedules from a subject pool database.

A. Characterizing Risk Attitudes

We adapt the HL procedure for our risk experiment by scaling it appropriately for our purposes. Multiplying by 10 the original payoff scale of 1 shown in Table 1, we had prizes ranging between $0.10 and $3.85, provides responses that span prizes between $1.00 and $38.50. These two payoffs scales are referred to as 1x and 10x hereafter. The 10x payoffs comfortably covers the range of prizes needed to apply the measures of risk aversion to our CR experiments, which have prizes of $0, $20 and $30. All subjects were given the 10x test, but some were also given a 1x test prior to the 10x (referred to as the 1x10x treatment) to test for order effects in elicitation of risk attitudes (see Harrison, Johnson, McInnes and Rutström [2005b] and Holt and Laury [2005]).

Figure 6 displays estimates of risk aversion for the 152 subjects that were in both the RA and CR experiments. We employ the CRRA utility function introduced earlier to define the CRRA intervals represented by each row in the payoff matrix faced by the subject. We then use an interval regression statistical model in which each subject's choice is the CRRA interval at which they "switch" from choosing lottery A to choosing lottery B.[25] For subjects that participated in the 1x10x experiments, the data constitute a panel consisting of two observations for that subject, so we estimate a panel interval regression models for the entire sample. In each case we include a standard list of socio-demographic characteristics, as well as using a random effects specification to allow for unobserved individual characteristics.[26] We also included a binary indicator to control for order effects when subjects did both 1x and 10x tasks, and dummy variables for each experimental session. The estimates shown in Figure 6 are then obtained as predictions from this estimated model, setting each individual's characteristics equal to their actual values.[27] For comparability with the EUT

---

[25] Several subjects switched two or more times. In this case we use the first and last switch points to define a relatively "fat" interval for that subject. Andersen, Harrison, Lau and Rutström [2004] present evidence that this switching behavior is generally consistent with the subjects expressing indifference.

[26] The observable characteristics are captured by binary indicators for sex, race (black), a Business major, Sophomore status, Junior status, Senior status, high GPA, low GPA, Graduate student status, expectation of a post-graduate education, college education for the father of the subject, college education for the mother of the subject, and U.S. citizen status. We also included age in years.

[27] Using the predicted CRRA coefficients from the interval regression has a disadvantage in that it throws out much of the individual variation that is not captured by socio-demographics. An alternative to using predicted CRRA is to use subjects' raw responses in terms of the interval in which they switched. If one compares the distribution of risk attitudes based on the raw responses to the distribution of predicted ones used here, it is apparent that the raw

decisions, we generated these predicted CRRA values for each individual assuming a scale of 10 and that this was the first task. Average CRRA is estimated to be 0.68 for this sample.

Figure 7 collects the information needed to evaluate our central hypothesis regarding the CR test. It simply aligns the information from Figure 1 on the differences in the certainty-equivalents conditional on risk attitudes with the information from Figure 6 on the distribution of risk attitudes in our sample. The results are clear by inspection: the distribution of observed risk attitudes puts the vast bulk of the sample in the region in which it is impossible to meaningfully test EUT using the paired choice comparisons. The choices for the A*B* pair (Problem 4), in particular, are virtually meaningless as tests of EUT given the distribution of risk attitudes in the sample, since the two lotteries have virtually the same certainty equivalent in that domain of risk attitudes.

There is no need to draw a comparable figure to compare the elicited distribution of risk attitudes and the certainty-equivalents from Figure 2 for the PR tasks. From Figure 6 we can see that none of the subjects was predicted to be risk-neutral.[28] Thus our first design goal has been met: we have found that one of the CR and PR tests, specifically the PR test, can be used to be meaningfully test EUT *given the range of risk attitudes observed for our subject pool.*

B. Revisiting Common Ratio Tests of EUT

Ignoring indifference induced by risk aversion, our subjects exhibit the CR effect, although we needed relatively large sample sizes to observe it. Out of the sample of 176 that completed one or other of the CR choices, 87 completed the AB choice (Problem 1) and 89 completed the A*B* choice (Problem 4). In the AB choice (Problem 1) option A was chosen by 50.6%, and in the A*B* choice (Problem 4) 62.9% chose option A. Thus we replicate the direction of the violation observed by CSS in the sense that a larger proportion of subjects choose the riskier alternative in the A*B* choice (Problem 4). Using a one-tailed Fisher Exact test, since we know the direction of the

---

distribution is "lumpier" but the qualitative conclusions are unchanged since the location of the raw distribution is essentially the same as the distribution shown in Figure 5.

[28] None of our subjects was estimated to have relative risk aversion greater than 1.11, which is the point of indifference for pair #6 of our PR experiments.

violation from previous literature, we reject the null hypothesis of no effect at the 6.7% significance level. Thus we have an *apparent* EUT violation to evaluate. When we restrict the test to the 152 subjects for whom we have RA responses and CR responses, the *p*-value on the one-sided Fisher Exact test increases from 0.067 to 0.132.

Next we remove all responses to the CR tasks that are given by subjects for whom we cannot reliably reject indifference, and repeat the statistical test. To allow for imprecision in the inferences about risk aversion we take a 90% confidence interval around the predicted CRRA for each subject, based on the estimated model described above. If that confidence interval includes the point of indifference, $r = 0.45$ in the CR tests, we deem the subject to be confounded out. This procedure results in 110 subjects being dropped, so that the statistical test is performed with only 42 (= 152-110) subjects. With such a large drop in observations it becomes immediately obvious that the ability to test EUT using these lottery tasks is very limited. Of these 42 subjects, option A was chosen by 46.4% in the AB choice, and option A was chosen by exactly 50.0% in the A*B* choice. The *p*-value on a one-sided Fisher Exact test in this case is 0.543, so we conclude that there is no evidence that EUT is violated once one removed the confounded subjects from the CR test.[29]

In effect, these results just confirm what the visual display in Figure 7 shows: indifference is a serious confound for CR tests of EUT.

C. Revisiting Preference Reversal Tests of EUT

We recruited 90 subjects, all of whom had participated in prior CR and RA experiments. Roughly one-half of the subjects made the direct binary choices in the order shown in Figure 2, and the other half had the last three direct choices presented to them first, as a check on possible order effects. Apart from adjusting for inflation in the prizes and avoiding any loss framing, our lotteries

[29] A tighter test is obtained if we just drop subjects whose estimated CRRA is within one standard error of 0.45, which amounts to using a 68.3% confidence interval instead of a 90% confidence interval. In this case we only drop 67 subjects, but the inferences are the same. Of the remaining 85 (=152-67) subjects, 54.1% chose option A and 56.8% chose option A*, resulting in a *p*-value of 0.493 for the one-sided Fisher Exact test.

and instructions closely follow Grether and Plott [1979].[30]

The bottom panel of Figure 3 shows the responses in our sample, closely matching those in the top panel that were obtained in the original PR experiments comparing consistency across binary choice and elicited valuations. In effect, this panel also shows the distribution when we drop subjects for whom the 90% confidence interval includes a coefficient for which the subject would be indifferent, since there is only one such subject! Only if we increase the confidence interval used to identify indifference to over 99% do we drop a second subject. So inferences about the validity of EUT from the PR task are not affected by the risk aversion confound, which is the point of the paired CR and PR design. We estimate that there are 2.2 reversals on average, with a standard deviation of 1.38 reversals per subject. Obviously this changes very little if we allow for the one or two subjects that are confounded by potential indifference. Thus we easily reject the EUT hypothesis of zero reversals using the original "reversals metric."

Turning to the alternative metric, consistency of direct binary choices, we again observe serious problems for EUT. The bottom panel of Figure 4 shows the responses in our sample. Before directly comparing them to those in the top panel, we must recognize that our subjects only had 5 pairs that they could have expressed the same preference for (there were 6 pairs in the original PR experiments). Nonetheless, the vast majority of subjects make only 3 consistent choices out of 5. The difference is that we know that our subjects were not indifferent, so we can conclude that these are indeed violations of EUT.

[30] Each of the lotteries in the *original* experiments involves the possibility of the subject making a loss relative to an initial endowment. No subject in the original experiments actually lost money from the experiment as a whole, since each subject received an initial endowment that exceeded any possible loss. We removed any references to one of the lottery prizes involving a "loss frame" by merging the initial endowment into the lottery prizes. This is a purely semantic change from Grether and Plott [1979], at least from the perspective of traditional EUT, but one that makes our tests of EUT more pointed. Payoffs for pair #1 are $29 with probability 35/36 and $16 otherwise or $61 with probability 11/36 and $14.50 otherwise. Payoffs for pair #2 are $24 with probability 29/36 and $16 otherwise or $43.50 with probability 7/36 and $17 otherwise. Payoffs for pair #3 are $36 with probability 18/36 and $16 otherwise or $26.50 with probability 34/36 and $13 otherwise. Payoffs for pair #4 are $124.50 with probability 4/36 and $16 otherwise or $29 with probability 32/36 and $17 otherwise. Payoffs for pair #5 are $25 with probability 34/36 and $17 otherwise or $41 with probability 14/36 and $14.50 otherwise. Finally, payoffs for pair #6 are $32 with probability 18/36 and $16.50 otherwise or $24 with probability 33/36 and $13.50 otherwise.

4. Are Our Subjects Representative?

Tests of EUT depend logically and empirically on the distribution of risk attitudes, so it is important to know if our sample is representative. Although our results are internally consistent for our sample, since we conducted the EUT tests on an in-sample basis, claims about generality of the confound for other samples depend on the distribution of risk attitudes in other samples. Figure 8 collects data from Holt and Laury [2002] and Harrison, Lau, Rutström and Sullivan [2005] to compare to our data. HL used student subjects in the United States, and we focus on their 1x and 20x responses since they contain the largest sample sizes (N=180). Harrison, Lau, Rutström and Sullivan [2005] used 253 subjects sampled from the adult population of Denmark, and had prizes that were roughly 150x those in Table 1. Our subjects, of course, were more comparable to those used by HL and we used 10x payoffs.

There are some intriguing differences in the elicited distributions. Ours is clearly more compact, and exhibits less heterogeneity. This is good news and bad news for the tests of EUT we consider. The bad news is that if a test happens to have an indifference point where this tight distribution lies, then virtually none of the subjects will provide informative responses; this was the case for our CR tests. The good news is that our subjects will then provide relatively informative responses for the complementary test of EUT; this was the case for our PR tests.

But more heterogeneity, such as we see in the other samples, means that some subjects will likely provide informative responses for one EUT test and others subjects for the other EUT test. Without knowledge of risk attitudes at the individual level, however, one cannot say which subject is more informative for which test. These data remind us that distributions of risk attitudes that are wide show evidence of heterogeneity of individual preferences, and not necessarily uncertainty about preferences.[31]

[31] There is a tendency in some experimental literature to implicitly assume that all subjects have the same risk preferences. This assumption is consistently rejected by the data when one controls for observable individual characteristics.

5. Conclusions

Previous tests of EUT using lottery choice pairs have not recognized the confounding influence of indifference. Properly identifying indifference requires knowledge of the risk attitudes of each subject. We independently estimate the risk attitudes of a sample of subjects, and then use those estimates to condition our analysis of their responses to two popular tests of EUT. We find that the distribution of risk attitudes places many of these subjects precisely at the point at which indifference makes tests of EUT meaningless. We develop a "complementary slack" experimental design: for a given empirical distribution of risk attitudes, if one test of EUT suffers heavily from the confound then the other should not. With such a design, once the risk attitudes of the subject are known it is indeed possible to undertake tests of EUT that are meaningful in the sense of not being confounded by indifference caused by risk attitudes.

Previous dismissals of EUT that are based on Common Ratio tests are premature, since it is not clear to what extent those tests were confounded by indifference. The most important conclusion we draw is that EUT is hard to test, and that its main weakness as a theory may be the difficulty of undertaking operationally meaningful tests of it. Thus it is not an "ex hypothesis," in the spirited sense of Rabin and Thaler [2001; p. 229], so much as it was *never* an operationally meaningful hypothesis *when tested unconditionally*. EUT can be tested in an operationally meaningful way only when risk attitudes are controlled for, and with sufficient accuracy.

Based on these tests we find significant evidence of inconsistent choices. We therefore confirm the conclusions drawn in previous tests, that EUT is not a good theory for explaining the observations, but here we have ensured that the test is not confounded. This is not an argument in favor of some specific alternative theory to EUT, since any theory that does not explicitly handle indifference would suffer from the same confounding influence and would need to be similarly retested. We leave it to future research to perform appropriate comparative tests of decision theories that avoid the indifference confounds.

Figure 1: Risk Attitudes and Common Ratio Tests of EUT

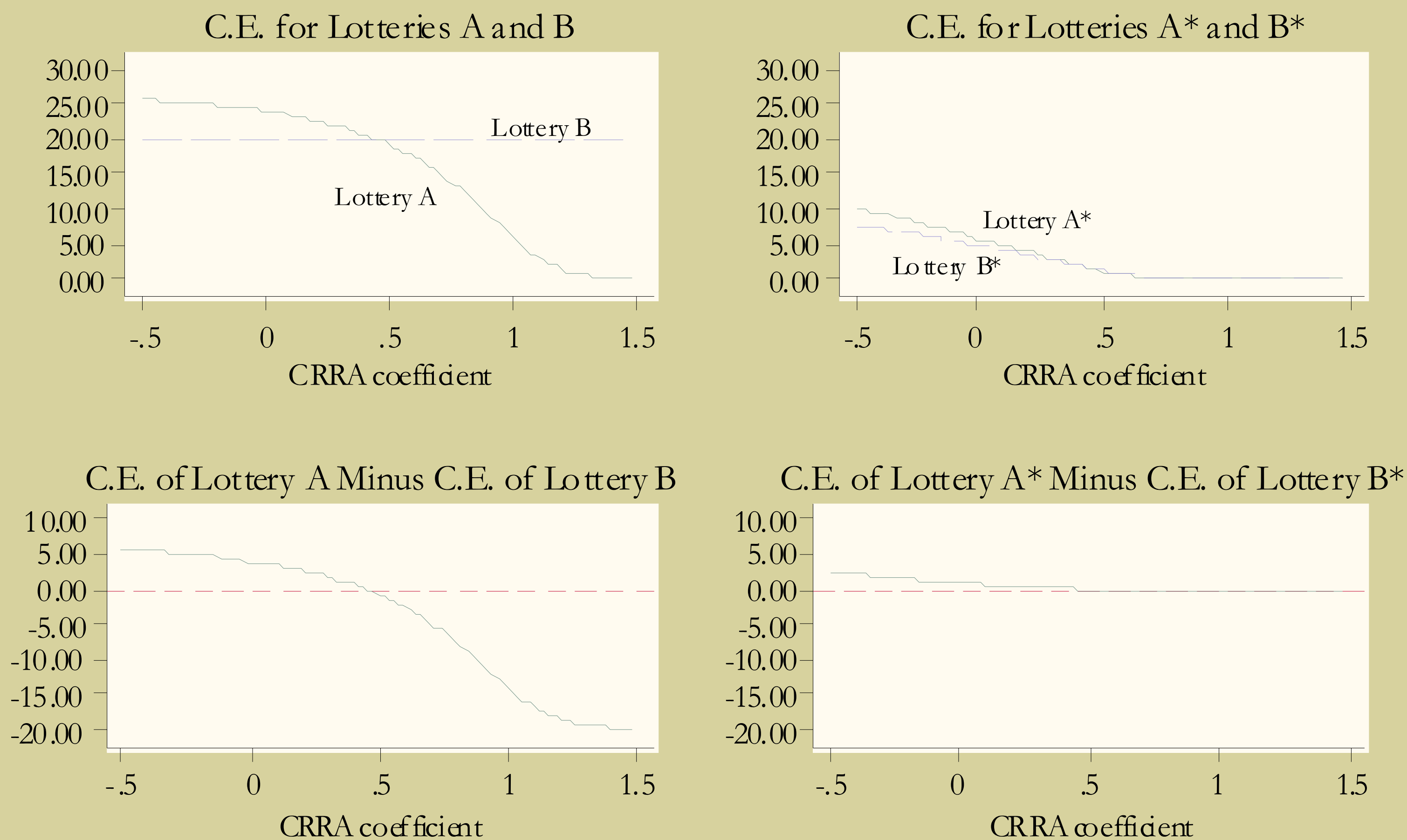
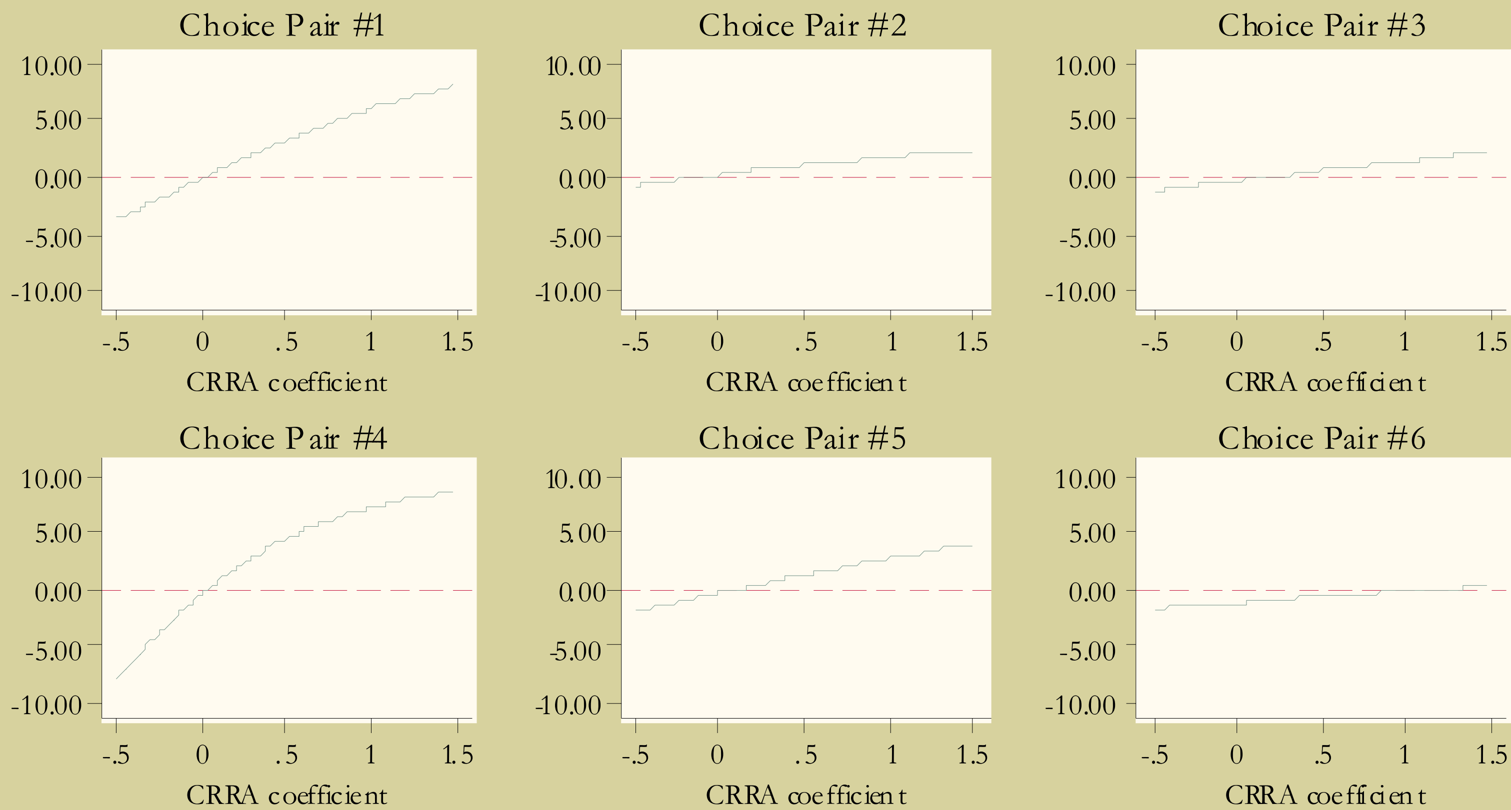


Figure 2:
Risk Attitudes and Preference Reversal Choice Pairs

Difference in Certainty Equivalents Favoring P-Bet in Each Pair



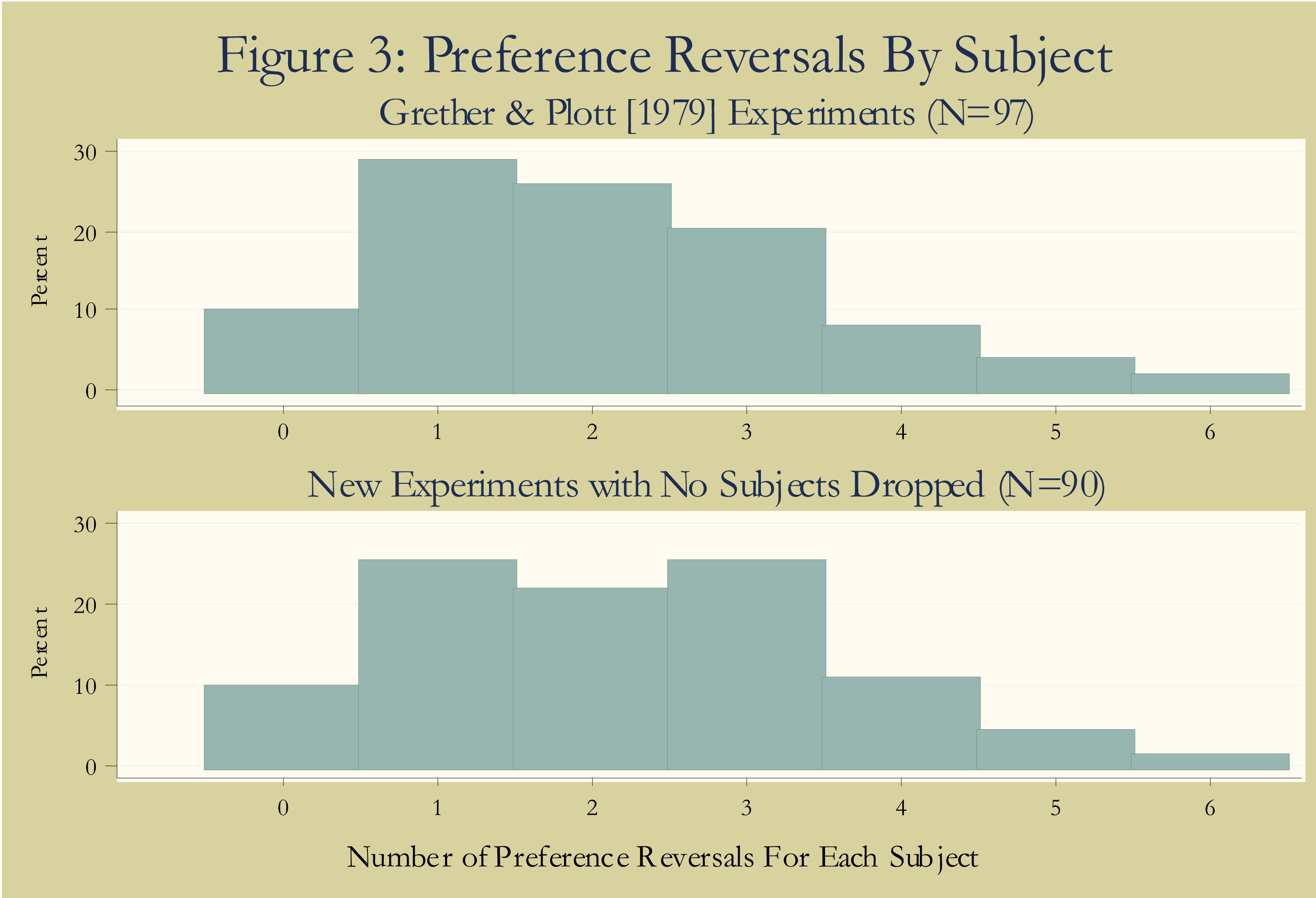


Figure 3: Preference Reversals By Subject
Grether & Plott [1979] Experiments (N=97)
Percent
30
20
10
0
0
1
2
3
4
5
6
New Experiments with No Subjects Dropped (N=90)
Percent
30
20
10
0
0
1
2
3
4
5
6
Number of Preference Reversals For Each Subject

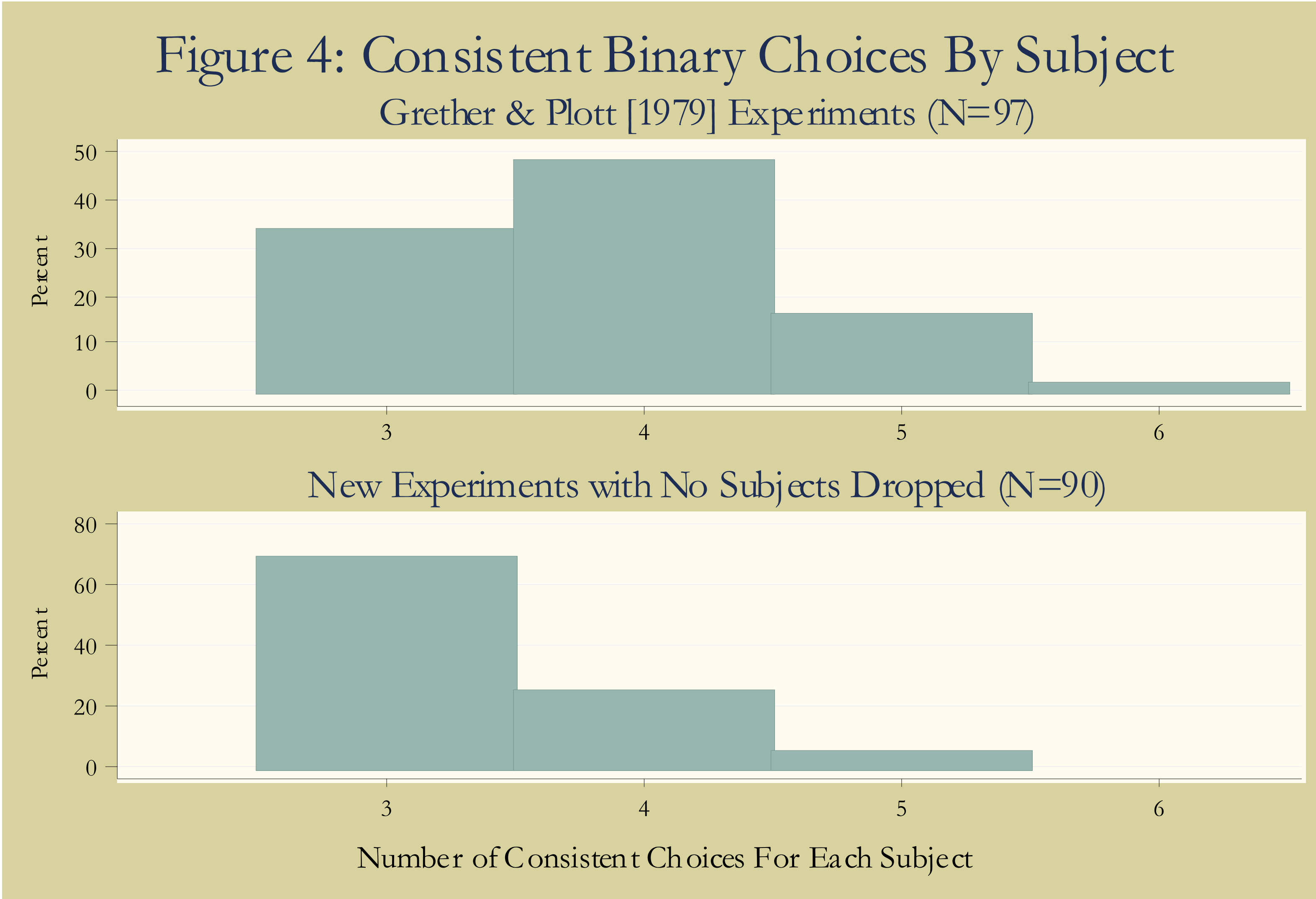


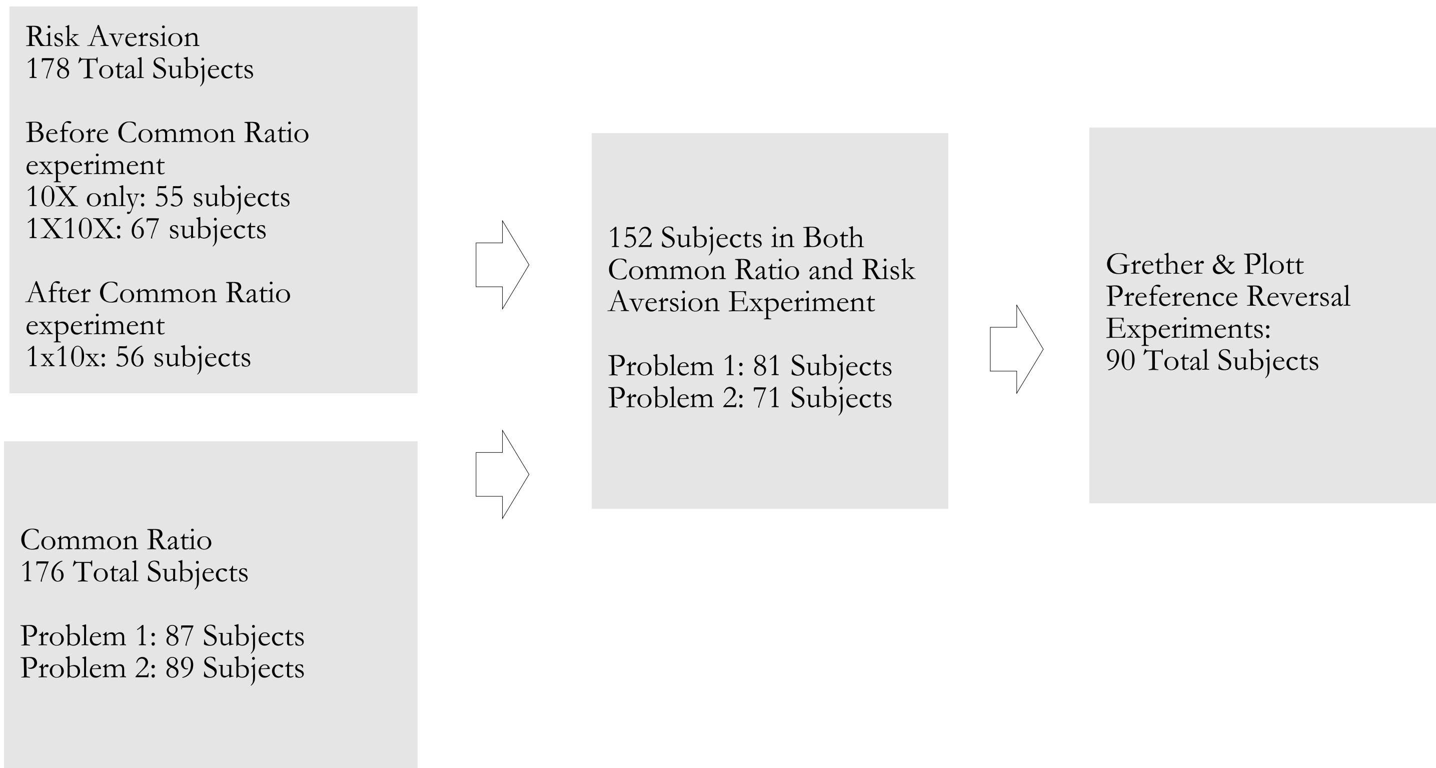
Figure 4: Consistent Binary Choices By Subject
Grether & Plott [1979] Experiments (N=97)
Percent
0
10
20
30
40
50
3
4
5
6
New Experiments with No Subjects Dropped (N=90)
Percent
0
20
40
60
80
3
4
5
6
Number of Consistent Choices For Each Subject

Table 1: Design of the Holt and Laury Risk Aversion Experiments

Standard Payoff Matrix

Lottery A				Lottery B						
p($2)		p($1.60)		p($3.85)		p($0.10)		EV^A	EV^B	Difference
0.1	$2	0.9	$1.60	0.1	$3.85	0.9	$0.10	$1.64	$0.48	$1.17
0.2	$2	0.8	$1.60	0.2	$3.85	0.8	$0.10	$1.68	$0.85	$0.83
0.3	$2	0.7	$1.60	0.3	$3.85	0.7	$0.10	$1.72	$1.23	$0.49
0.4	$2	0.6	$1.60	0.4	$3.85	0.6	$0.10	$1.76	$1.60	$0.16
0.5	$2	0.5	$1.60	0.5	$3.85	0.5	$0.10	$1.80	$1.98	-$0.17
0.6	$2	0.4	$1.60	0.6	$3.85	0.4	$0.10	$1.84	$2.35	-$0.51
0.7	$2	0.3	$1.60	0.7	$3.85	0.3	$0.10	$1.88	$2.73	-$0.84
0.8	$2	0.2	$1.60	0.8	$3.85	0.2	$0.10	$1.92	$3.10	-$1.18
0.9	$2	0.1	$1.60	0.9	$3.85	0.1	$0.10	$1.96	$3.48	-$1.52
1	$2	0	$1.60	1	$3.85	0	$0.10	$2.00	$3.85	-$1.85

Figure 5: Experimental Design



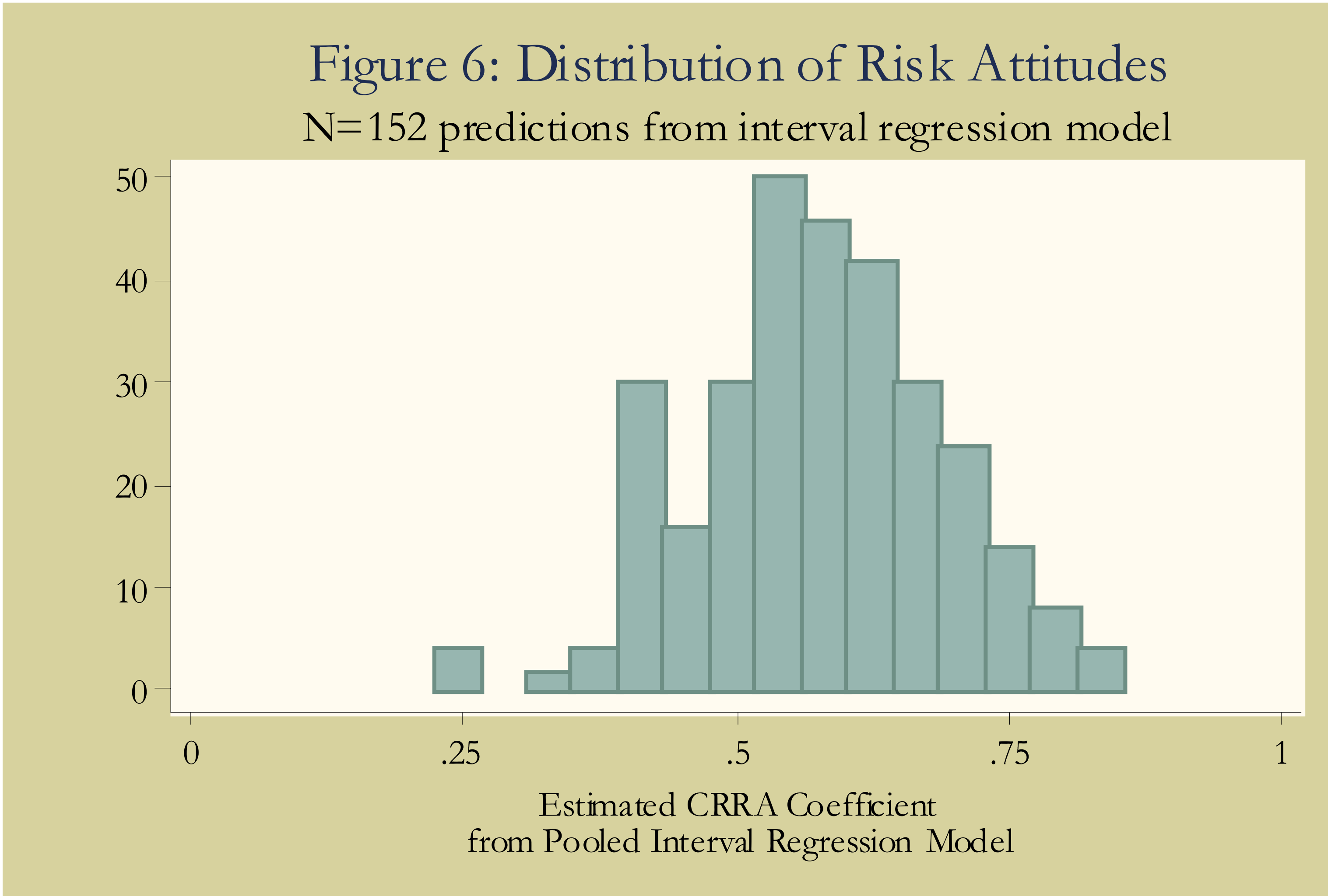


Figure 6: Distribution of Risk Attitudes

N=152 predictions from interval regression model

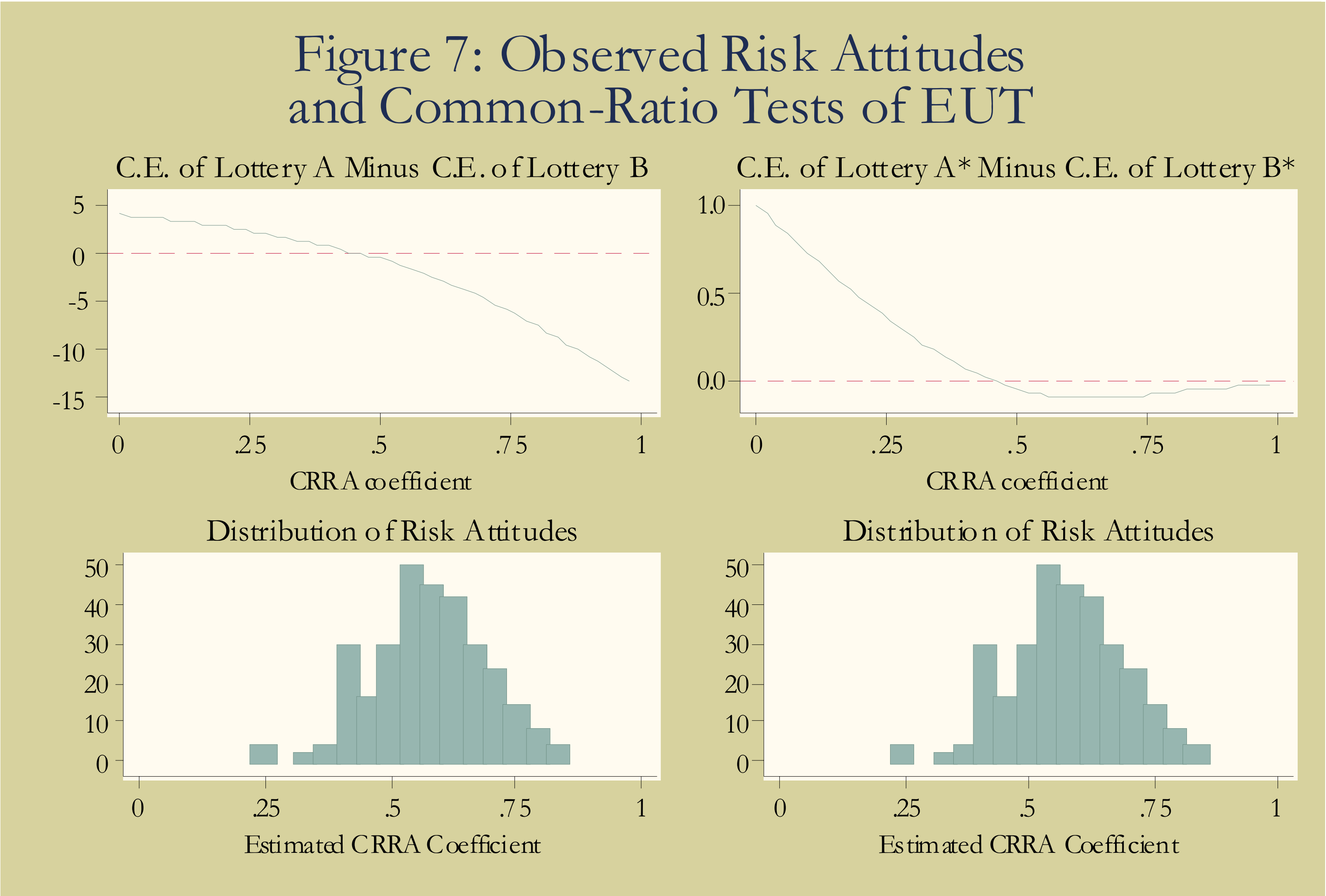


Figure 7: Observed Risk Attitudes
and Common-Ratio Tests of EUT
C.E. of Lottery A Minus C.E. of Lottery B
5
0
-5
-10
-15
0
.25
.5
.75
1
CRRA coefficient
C.E. of Lottery A* Minus C.E. of Lottery B*
1.0
0.5
0.0
0
.25
.5
.75
1
CRRA coefficient
Distribution of Risk Attitudes
50
40
30
20
10
0
0
.25
.5
.75
1
Estimated CRRA Coefficient
Distribution of Risk Attitudes
50
40
30
20
10
0
0
.25
.5
.75
1
Estimated CRRA Coefficient

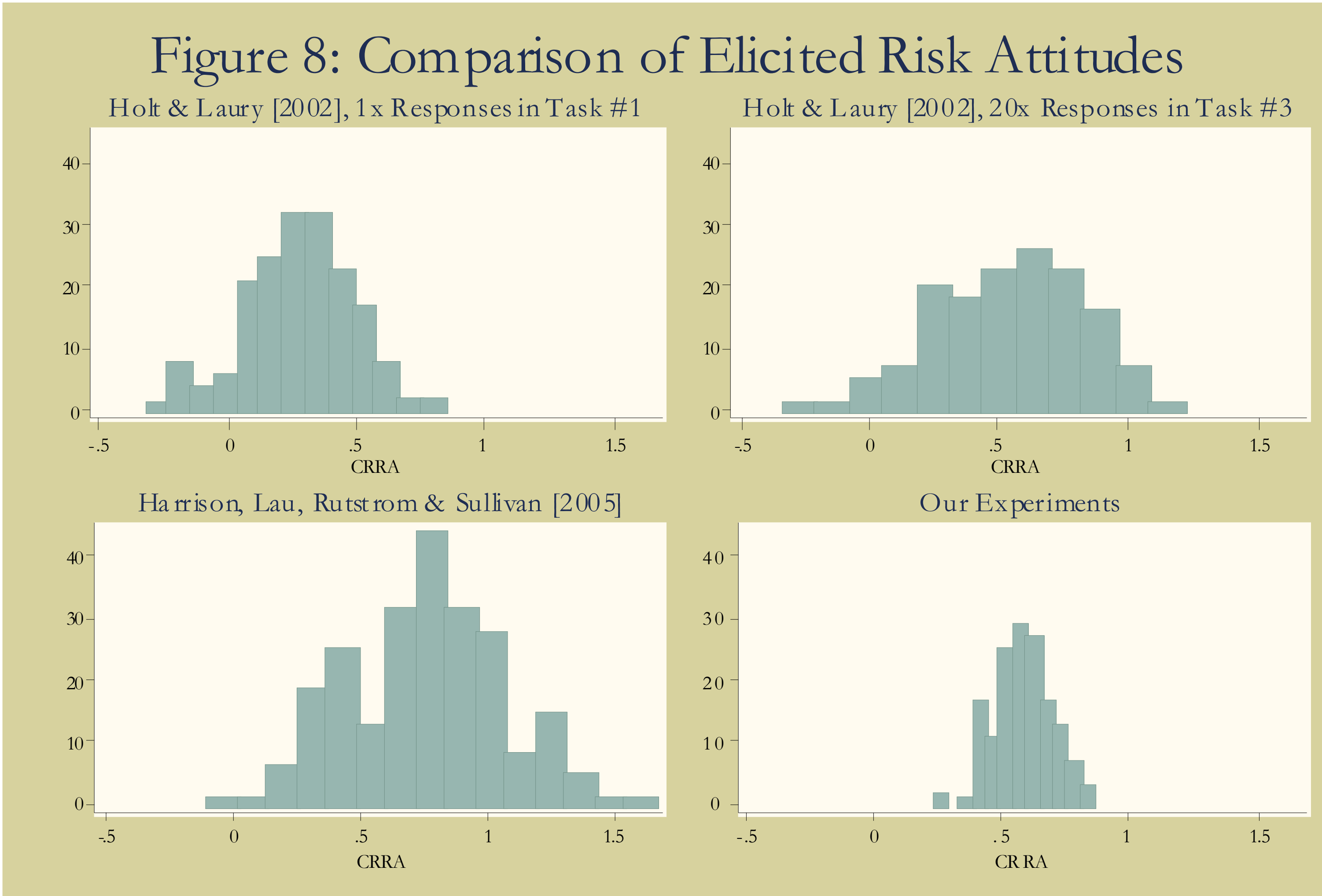


Figure 8: Comparison of Elicited Risk Attitudes

References

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Elicitation Using Multiple Price List Formats," *Working Paper 04-8*, Department of Economics, College of Business Administration, University of Central Florida, 2004.

Ballinger, T. Parker, and Wilcox, Nathaniel T., "Decisions, Error and Heterogeneity," *Economic Journal*, 107, July 1997, 1090-1105.

Charness, Gary, and Rabin, Matthew, "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), August 2002, 817-869.

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

Cox, James C., and Sadiraj, Vjollca, "Implications of Small- and Large-Stakes Risk Aversion for Decision Theory," *Unpublished Manuscript*, Department of Economics, University of Arizona, May 2004.

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

Grether, David M., and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69(4), September 1979, 623-648.

Harrison, Glenn W., "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 82, December 1992, 1426-1443.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Temporal Stability of Estimates of Risk Aversion," *Applied Financial Economics Letters*, 1, 2005a, 31-35.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005b, forthcoming.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Controls for Risk Aversion in the Laboratory," *Working Paper 05-5*, Department of Economics, College of Business Administration, University of Central Florida, 2005c.

Harrison, Glenn W.; Lau, Morten Igel; Rutström, E. Elisabet, and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).

Hey, John D., "Experimental Investigations of Errors in Decision Making Under Risk," *European Economic Review*, 39, 1995, 633-640.

Holt, Charles A., "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, 1986, 508-515.

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

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, forthcoming.

Loomes, Graham; Moffatt, Peter, and Sugden, Robert, "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24, 2002, 103-130.

Loomes, Graham, and Sugden, Robert, "Incorporating a Stochastic Element Into Decision Theories," *European Economic Review*, 39, 1995, 641-648.

Rabin, Matthew, "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68, 2000, 1281-1292.

Rabin, Matthew and Richard Thaler, "Anomalies: Risk Aversion," *Journal of Economic Perspectives*, 15, Winter 2001, 219-232.

Rubinstein, Ariel, "Comments on the Risk and Time Preferences in Economics," *Unpublished Manuscript*, Department of Economics, Princeton University, 2002.