Naturally Occurring Preferences and Exogenous Laboratory Experiments:

A Case Study of Risk Aversion

by

Glenn W. Harrison, John A. List and Charles Towe[†]

December 2004

Abstract. Does individual behavior in a laboratory setting provide a reliable indicator of behavior in a naturally occurring setting? We consider this general methodological question in the context of eliciting risk attitudes. The controls that are typically employed in laboratory settings, such as the use of abstract lotteries, could lead subjects to employ behavioral rules that differ from the ones they employ in the field. Since it is field behavior that we are interested in understanding, those controls might be a confound in themselves if they result in differences in behavior. We find that the use of artificial monetary prizes provides a reliable measure of risk attitudes when the natural counterpart outcome has minimal uncertainty, but that it can provide an unreliable measure when the natural counterpart outcome has background risk. These results are consistent with conventional expected utility theory for the effects of background risk on attitudes to risk. Behavior tended to be risk loving when artificial monetary prizes were used or when there was minimal uncertainty in the natural non-monetary outcome. But subjects drawn from the same population were risk averse when their attitudes were elicited using the natural non-monetary outcome that had some background risk. Theory predicts this effect of background risk, but not the change from risk-loving to risk-aversion.

[†] Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison) and Department of Agricultural & Resource Economics, University of Maryland, USA (List and Towe). Harrison is also affiliated with the Institute of Simulation & Training, University of Central Florida; List is also affiliated with the Department of Economics, University of Maryland, and the National Bureau of Economic Research. E-mail: gharrison@bus.ucf.edu, jlist@arec.umd.edu, and ctowe@arec.umd.edu. We are grateful to Michael Price for assistance with the running of these experiments. Supporting data and instructions are stored in project "Risk Aversion in Naturally Occurring Markets" at the Experimental Sciences Digital Archive located at http://exlab.bus.ucf.edu.

One of the main attractions of experimental methods is the control that it provides over factors that could influence behavior. The ability to control the environment allows the researcher to study the effects of treatments in isolation, and hence makes it easier to draw inferences as to what is influencing behavior. In most cases we are interested in making inferences about existing or possible field behavior. We hypothesize that there is a danger that the imposition of an exogenous laboratory control might make it harder, in some settings, to make reliable inferences about field behavior. The reason is that the experimenter might not understand something about the factor being controlled, and might impose it in a way that is inconsistent with the way it arises naturally in the field, and that affects behavior.

We take as a case study the elicitation of measures of risk aversion, which is arguably one of the most primitive characteristics of the standard specification of utility functions.[1] In the traditional paradigm, risk aversion is viewed in terms of diminishing marginal utility of the final prize in some abstract lottery. The concept of a lottery here is just a metaphor for a real lottery, although in practice the metaphor has been used as the primary vehicle for laboratory elicitation of risk attitudes. In general there is some commodity x and various levels i of x, x_i, that depend on some state of nature which occurs with a probability p_i that is known to the individual whose preferences are being elicited. Thus the lottery is defined by $\{x_i\,;p_i\}$. Traditional measures of risk aversion are then defined in terms of the curvature of the utility function with respect to x.

We consider the evaluation of risk attitudes in the field. This entails more than just "leaving the classroom" and recruiting outside of a university setting, as emphasized by Harrison and List [2004]. In terms of sample composition, it means finding subjects who deal with uncertainty in varying degrees, and trying to measure the extent of their field experience with uncertainty. Moreover, it means developing stimuli that more closely match those that the subjects have experience with, so that they can use whatever heuristics they have for that commodity when making their choices. Finally, it means developing ways of communicating probabilities that correspond with

[1] The other is the characterization of preferences across arguments of the utility function that are a-temporal (e.g., beer and pizza) or time-dated (e.g., beer today or beer tomorrow). We focus on just one commodity and time-dating, and do not therefore consider such issues.

the language that the subjects are familiar. Thus, field experimentation in this case (and in general) involves several simultaneous changes from the lab setting with respect to subject recruitment and the development of stimuli that match the field setting.

A second theme is the importance of "background risk" for the attitudes towards a specific "foreground risk" that are elicited. In many field settings it is not possible to artificially identify attitudes towards one risk source without worrying about how the subjects view that risk as being correlated with other risks. For example, mortality risks from alternative occupations tend to be highly correlated with morbidity risks. It is implausible to ask subjects their attitude toward one risk without some coherent explanation as to why a higher or lower level of that risk would not be associated with a higher or lower risk of the other.

In the field experiment, which uses numismatists at a coin show as experimental subjects, we focus on a well-studied special case of this issue of multiple risks: the response of decision-makers to the addition of independent background risk.[2] Assuming expected-utility preferences and risk averse agents, the addition of independent risk reduces welfare. More importantly for our purposes, the literature has yielded a set of preferences that guarantee that the addition of an unfair background risk to wealth reduces the certainty equivalent of any other independent risk. That is, the addition of background risk of this type makes risk averse individuals behave in a more risk averse way with respect to any other independent risk. Gollier and Pratt [1996] refer to this type of behavior as "risk vulnerability," and show that all weakly Decreasing Absolute Risk Averse utility functions are risk vulnerable; this class includes many popular characterizations of risk attitudes, such as Constant Absolute Risk Aversion (CARA) and Constant Relative Risk Aversion (CRRA).[3]

While the notion of risk vulnerability is intuitively appealing, it is at odds with the spirit of psychological work that provides evidence of diminishing sensitivity. This is the notion that if the agent is already at a point of sufficiently high risk, she will not pay particular attention to the addition

[2] For example, see Pratt and Zeckhauser [1987], Pratt [1988], Kimball [1993], Gollier and Pratt [1996] and Eeckhoudt et al. [1996].

[3] Eeckhoudt et al. [1996] extend these results by providing the necessary and sufficient conditions on the characterization of risk aversion to ensure that any increase in background risk induces more risk aversion.

of a *small* independent risk. Making use of the concept of constant risk aversion due to Safra and Segal [1998], Quiggin [2003] shows that the premium for a given risk is always reduced by the presence of independent background risk. Thus, in this setting, aversion to one risk is reduced by the addition of an independent background risk. Rather than arguing that independent risks are substitutes, as in the standard expected utility framework, Quiggin [2003] shows that they are complements in a set of alternative models. Of course, what constitutes a "sufficiently high initial risk" and a "small independent risk" are in the eyes of the beholder.

Our main conclusion is that the use of artificial monetary prizes provides a reliable measure of risk attitudes when the natural counterpart outcome has minimal uncertainty,[4] but that it can provide an unreliable measure when the natural counterpart outcome has background risk.[5] These results are consistent with the available theory from conventional expected utility theory (EUT) for the effects of background risk on attitudes to risk.

In addition, the qualitative measure of risk attitudes differs in the three cases we examined. Behavior tended to be risk *loving* when artificial monetary prizes were used or when there was minimal uncertainty in the natural non-monetary outcome. But subjects drawn from the same population were risk *averse* when their attitudes were elicited using the natural non-monetary outcome that had some background risk. Theory predicts this effect of background risk, but not the change from risk-loving to risk-aversion.

We observe considerable individual heterogeneity in risk attitudes, such that one should not readily assume homogenous risk preferences for the population. This heterogeneity is also correlated with observable characteristics of the individual. These results are consistent with those from laboratory experiments in college settings and in the field using artificial monetary prizes.

Finally, we find that our procedural controls have a measurable effect on elicited risk attitudes. Our estimates can be affected by skewing the frame of choices offered. We also show that

---

[4] Lusk and Coble [2004a] show that risk attitudes elicited over abstract monetary outcomes are useful predictors of behavior over risky non-monetary outcomes, which is consistent with this finding.

[5] Lusk and Coble [2004b] find that adding abstract background risk to an elicitation procedure using artificial monetary outcomes also generates more risk aversion, although they do not find the effect to be large quantitatively.

there is an "experimenter effect" on responses collected by one particular experimenter, and that there is an effect from eliciting responses on the third day of the show. Thus, one might have to worry more about procedural controls when eliciting risk attitudes in the field than in the lab, which is consistent with the generic difficulties of ensuring the same controls in the field as in the lab.

In section 1 we review the standard lab procedures used to elicit risk aversion. In section 2 we propose a field counterpart. In section 3 we discuss the empirical results. Section 4 concludes.

1. Estimating Risk Aversion in the Laboratory

There are several popular ways in which one can measure risk aversion in the laboratory. One is by eliciting the certainty-equivalent of a given lottery using open-ended valuation procedures such as a Vickrey auction or the Becker-DeGroot-Marschak procedure (e.g., Harrison [1986] and Kachelmeir and Shehata [1992]). The other is by observing choices that subjects make over lotteries that vary the probabilities of winning different prizes (e.g., Binswanger [1980][1981]).[6] We propose utilizing a structured variant of the latter approach, due to Holt and Laury [2002], since it has been heavily used in recent laboratory experiments and involves a relatively transparent task.[7] We limit ourselves to the characterization of risks framed in terms of gains.[8]

---

[6] The earliest use of this type of design in the context of elicitation of risk attitudes is, we believe, Miller, Meyer and Lanzetta [1969]. Their design confronted each subject with 5 alternatives that constitute an MPL, although the alternatives were presented individually over 100 trials.

[7] Related methods were used earlier by Murnighan, Roth and Schoumaker [1987][1988], although they only used the results to sort subjects into one group that was less risk averse than the other. Beck [1994] utilized it to identify risk aversion in subjects, prior to them making group decisions about the dispersion of everyone else's potential income. This allowed an assessment of the extent to which subjects in the second stage chose more egalitarian outcomes because they were individually averse to risk or because they cared about the distribution of income.

[8] A major issue in the evaluation of risk preferences has been the extent to which risk attitudes change with the "frame" offered to subjects. The term "loss frame" refers to a task in which the choices are framed as losses from some reference point. All of the non-hypothetical experiments that we know of provide a positive reference point such that the subject cannot suffer a net loss (e.g., Battalio, Kagel and Jiranyakul [1990] and Kagel, MacDonald and Battalio [1990]). An experimental session in which the subject faces choices in which some outcomes involved net losses for the session should be referred to as involving choices in the *loss domain*. These are important semantics when evaluating the experimental evidence accurately, since none of the non-hypothetical evidence pertains to the loss domain but is often interpreted casually as if it does. In any event, conducting experiments with a loss frame is possible ethically, and procedurally simple, as long as one restricts the overall task to the gain domain.

A. The Basic Procedure

Holt and Laury [2002] (HL) devise a simple experimental measure for risk aversion using a multiple price list (MPL) design. Each subject is presented with a choice between two lotteries, which we denote 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-last column as \$1.64, although the EV columns were not presented to subjects. Similarly, lottery B in the first row has 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 chooses A or B in each row, and one row is later selected at random for payout for that subject. The logic behind this test 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 penultimate row. Arguably, the last row is simply a test that the subject understood the instructions, and has no relevance for risk aversion. 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.

Data from this exercise may be analyzed using a variety of statistical models. Each subject made 10 responses. The responses can be reduced to a scalar if one looks at the *lowest* row in Table 1 at which the subject "switched" from lottery A to lottery B.[9] This reduces the response to a scalar for each subject and task, but a scalar that takes on integer values between 0 and 10. Alternatively, one could study the effects of experimental conditions in terms of a CRRA characterization, employing an interval regression model. The dependent variable is the CRRA interval that subjects implicitly choose when they switch from lottery A to lottery B. For each row of panel A in Table 1, one can calculate the implied bounds on the CRRA coefficient, and these are in fact reported by HL

[9] Some subjects switched several times, but the minimum switch point is always well-defined. In the laboratory setting of most previous experiments it turns out not to make much difference how one handles these "multiple switch" subjects, but our analysis considers the effect of accounting for them in the field as explained below.

[2002; Table 3]. These intervals are shown in the final column of Table 1. Thus, for example, a subject who made 5 safe choices and then switched to the risky alternatives would have revealed a CRRA interval between 0.14 and 0.41, and a subject who made 7 safe choices would have revealed a CRRA interval between 0.68 and 0.97, and so on.

This procedure has now been widely employed in the laboratory (e.g., Harrison, Johnson, McInnes and Rutström [2003a][2003b] and Laury and Holt [2002]). It has also been employed in some *artefactual field experiments*, to use the terminology of Harrison and List [2004] to indicate experiments that use laboratory procedures and commodities with field subjects in an artificial setting. For example, Harrison, Lau, Rutström and Sullivan [2005] and Harrison, Lau and Rutström [2004] examine behavior of Danish citizens confronted with an extension of the HL procedure. They find evidence of risk aversion in general, and considerable heterogeneity associated with observable characteristics of the sample.

B. Characterizing Risk Attitudes

The most popular functional form for the characterization of risk attitudes has been the CRRA specification. This function can be defined by the utility function $U(x) = (x^{1-r})/(1-r)$, 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 $r = 1$, $U(x) = \ln(x)$.

There are several ways in which these characterizations can be applied. The first is to use the assumption of a locally CRRA utility function to define RRA values for the "switch points" in the HL design. This leads to a minimally parametric measure of risk aversion for a given subject, consisting of an interval of CRRA values defining the switch point. The second approach is to estimate a statistical model that uses the CRRA specification, and allows the parameters of each specification to be conditioned on observable individual characteristics of the individual or the task.[10]

[10] An interim approach is to estimate the CRRA or EP specifications and assume that all subjects have the same risk preferences (e.g., Holt and Laury [2002]). This approach typically requires that one augment the statistical model with some assumption that allows individuals to deviate randomly from the estimated model, such as a stochastic choice

The CRRA characterization is nested in the Expo-Power (EP) utility function proposed by Saha [1993]. Following Holt and Laury [2002], the EP function is defined as

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

where α and r are parameters to be estimated. Relative risk aversion (RRA) is then $r + \alpha(1-r)y^{1-r}$. So RRA varies with income if $\alpha \neq 0$. This function nests CRRA (as $\alpha \rightarrow 0$) and CARA (as $r \rightarrow 0$). We use this flexible functional form later to test if the CRRA characterization is acceptable for our data.

2. Design of the Field Experiments

A. General Design

Our objective was to evaluate risk aversion using a more natural representation of the elicitation task than previous laboratory experiments, and to compare the effect of that representation on elicited risk attitudes. To avoid order effects which can easily confound inferences about risk attitudes, we employed a between-subjects design. Our primary treatments reflect differences in the nature of the prize being considered. Our secondary treatments reflect differences in the way in which the HL design was framed to subjects, to detect any "anchoring" biases in responses.

The primary treatments vary the commodity used as the prize. In the *Money* treatments we follow the procedures of the laboratory experiments and simply use monetary rewards as the prizes. This treatment parallels the abstract lottery representation of the laboratory, and serves as a control. In the *Graded Coins* treatment we employ graded coins as the prize, where higher grades would be expected to generate higher utility for individuals in this sample (just as more money is expected to generate higher utility in the abstract setting). To ensure control over the utility ranking of each prize, we employ different grades of the *same* coin in each lottery. Conveniently for us, the grade is a scalar, typically from 1 to 70 as discussed further below. In the *Ungraded Coins* treatment we employ exactly the same graded commodity, but without the certification of the grade. Thus, we add

function linking observed choices to the estimates in the latent model. Apart from general tests of functional form for a sample, we see little virtue in such a stark representation of individual preferences when one has information on the individual characteristics of the subjects.

"background risk" to the valuation of the lottery by removing the grading information from graded commodities with known grades. This adds additional, natural uncertainty to the lottery outcomes since the subject must additionally consider the possible grades that the prize might have. Our design ensures that the ungraded prize is actually a mean-zero deviation of the grade of the graded prize, since we were extremely careful when removing the certification to avoid damage to the commodity. Grading is costly, albeit with a known and stable fee, so we also added a certificate for "a free grading" to each ungraded prize. These procedures provide a controlled way to identify the effect of background risk in a natural manner.

Our secondary treatment explicitly varies the skewness frame, to detect if subjects anchor in the middle of the MPL that they are presented. We employ three variants. In the *Symmetric Frame* we allow the probabilities of the two lotteries to increment by a constant 5 percentage points from row to row. Thus a risk-neutral subject would switch from option A to option B roughly in the middle of the table. In the *Lower* frame we have smaller increments in the probability for risk-loving choices than we do for risk-averse choices, such that a risk-neutral subject should switch lower in the table than in the symmetric frame, at least in the absence of any anchoring effect. Conversely for the *Higher Frame*, the risk-neutral subject should switch higher up the table than in the symmetric frame if there are no anchoring effects. We name these frames to reflect the bias in elicited risk attitudes that they are intended to induce if there is anchoring behavior.

B. Specific Design

The specific rare coin that we employ is a U.S. Silver Dollar known as the Morgan Dollar.[11] These were primarily minted between 1878 and 1921 in the Philadelphia and San Francisco mints,

[11] The expression "Morgan Dollar" is a common term used for the Liberty Head Silver Dollar struck from 1878 until 1904, and again briefly in 1921. George Morgan was the assistant engraver at the time the coin was first struck, but his design was selected over William Barber's for the dollar. Morgan was passed over for the Chief Engraver's job when William Barber died in 1879 (the job actually went to Charles Barber, William's son). However, Morgan became the next Chief Engraver in 1918, a position he held until his death in 1925.

and between 1878 and 1904 in the New Orleans mint.[12] All were struck in fine 0.900 fine silver, and were extremely popular in the wild West, adding to their appeal to collectors. To ensure control, we only use 1897 coins minted from San Francisco, known in the trade as "1879-S Morgans," pictured above.

Coins are graded using a relatively common 1-to-70 classification scheme.[13] This scheme originated with the numismatist William Sheldon in 1948. His original objective, interestingly, was to develop a "ratio scale" of value, such that a coin graded as a 70 would be worth 70 times a coin

[12] Some very rare Morgans were minted in Denver in 1921, and in Carson City between 1878 and 1893. The mint is readily identified by a mark just above and between the "D" and "O" in the "DOLLAR" imprinted on one side. For the Philadelphia mint there is no mint mark, and for the San Francisco mint there is an "S" mark.

[13] These are: PO-1 Identifiable date and type; FR-2 Mostly worn, though some detail is visible; AG-3 Worn rims but most lettering is readable though worn; G-4 Slightly worn rims, flat detail, peripheral lettering nearly full; G-6 Rims complete with flat detail, peripheral lettering full; VG-8 Design worn with slight detail; VG-10 Design worn with slight detail, slightly clearer ; F-12 Some deeply recessed areas with detail, all lettering sharp; F-15 Slightly more detail in the recessed areas, all lettering sharp; VF-20 Some definition of detail, all lettering full and sharp; VF-25 Slightly more definition in the detail and lettering; VF-30 Almost complete detail with flat areas; VF-35 Detail is complete but worn with high points flat; EF-40 Detail is complete with most high points slightly flat; EF-45 Detail is complete with some high points flat; AU-50 Full detail with friction over most of the surface, slight flatness on high points; AU-53 Full detail with friction over 1/2 or more of surface, very slight flatness on high points; AU-55 Full detail with friction on less than 1/2 surface, mainly on high points; AU-58 Full detail with only slight friction on the high points; MS/PR-60 No wear. May have many heavy marks/hairlines, strike may not be full; MS/PR-61 No wear. Multiple heavy marks/hairlines, strike may not be full; MS/PR-62 No wear. Slightly less marks/hairlines, strike may not be full; MS/PR-63 Moderate number/size marks/hairlines, strike may not be full; MS/PR-64 Few marks/hairlines or a couple of severe ones, strike should be average or above; MS/PR-65 Minor marks/hairlines though none in focal areas, above average strike; MS/PR-66 Few minor marks/hairlines not in focal areas, good strike; MS/PR-67 Virtually as struck with minor imperfections, very well struck ; MS/PR-68 Virtually as struck with slight imperfections, slightest weakness of strike allowed; MS/PR-69 Virtually as struck with minuscule imperfections, near full strike necessary; MS/PR-70 As struck, with full strike; and GV Government issue price. In addition, designations exist for color (RD Red; RB Red-Brown; and BN Brown), "strike" (FS Full Steps; FB Full Bands; FH Full Head; FBL Full Bell Lines; CA Cameo; and BM Branch Mint Proof) and surface (DM Deep Mirror Prooflike; PL Prooflike; and DC Deep Cameo).

graded as a 1. In 1986 the Sheldon scale was formalized by the *Professional Coin Grading Service* (PCGS), a private company of coin collectors wanting to establish a thick market in rare coins.[14] They employed the "cardinal scale" proposed by Sheldon, and dropped the presumption that it also reflected relative value. Since then the PCGS has established a dominant position in the coin grading market, and has graded almost 9 million coins worth roughly $11.3 billion.

The specific coins we employ are MS/PR-63, MS/PR-65, MS/PR-65PL, MS/PR-66, with retail values of $40, $125, $200 and $350 at the time of the experiment.[15] These grades are defined as follows:

- MS/PR-63 Moderate number/size marks/hairlines, strike may not be full
- MS/PR-65 Minor marks/hairlines though none in focal areas, above average strike;
- MS/PR-65PL Minor marks/hairlines though none in focal areas, above average strike, with a proof-like surface; and
- MS/PR-66 Few minor marks/hairlines not in focal areas, good strike.

These grades were provided by PCGS, and the coins were sealed in a protective cover. This cover can be "cracked," but it is effectively tamper-proof if one wants to leave the official certification in place. The image shown below illustrates a PCGS-graded Morgan Dollar. The circled number is a sonically sealed certification number, which can be entered on the PCGS web site[16] to verify the date, mintmark, denomination, variety, and grade. In our experience, virtually any serious coin collector in the United States would be familiar with this grading system.

Our procedures were to set up a stall at a major coin show in 2004.[17] We offered subjects $5 for participating in an experiment. In most of our previous field experiments we have not used such a participation fee, but we wanted to ensure a relatively large sample size and to ask relatively more

---

[14] Web page http://www.pcgs.com contains information on the grading scales employed, historical background to the grading schemes, and prices on many rare coins including the 1897 Morgans.

[15] Retail values were obtained from the PCGS website, http://www.pcgs.com, just prior to the show. Excluding shipping and insurance, which would not be relevant at a show, we purchased 12 coins worth $2,166 on the open market just prior to the show. The retail values for these coins were $2,105, so our average costs were 2.9% above retail. On average, we paid 11%, 3% and 21% over retail for the MS63, MS65 and MS65PL coins, and 8% under retail for the MS66 coins.

[16] Specifically, at http://www.pcgs.com/verify.chtml.

[17] The Central Florida Coin Show, April 16-18, at the Orlando Expo Center.

survey questions to better characterize the individual.[18]

After subjects agreed to participate, and completed the consent form, we presented them with the instructions for the risk aversion task, and one of the nine payoff tables reflecting our 3×3 design. We randomized the Money, Graded Coins, and Ungraded Coins treatments every hour, over the three days of the show.[19]

Figure 1 shows the effect of the anchoring treatment in which we varied the probabilities associated with each row. The symmetric frame has 20 decision rows in which the probability of the high prize outcome in each lottery option increases by a fixed 0.05 in each row. The associated CRRA values for each switch point range from a risk-loving -2.98 to a risk-averse 2.50, with risk-neutrality corresponding to a switch from option A to option B around row 7, which is the proximate mid-point of the decision table actually presented to subjects.

The "low frame" changes the probabilities so that the risk-neutral switch point occurs at a later, or lower, point in the decision table: at rows 10 or 11. Conversely, the "high frame" changes the probabilities so that the risk-neutral switch point occurs at an earlier, or higher, point in the decision table: at rows 2 or 3.[20] The objective is to generate some separation in the observed row choices under the hypothesis that subjects are simply picking a row that is close to the middle of the table, rather than according to the implied risk aversion of that row. Even if subjects do not pick exactly in the middle, these frames may reveal a behavioral pull in that direction. For this to occur in

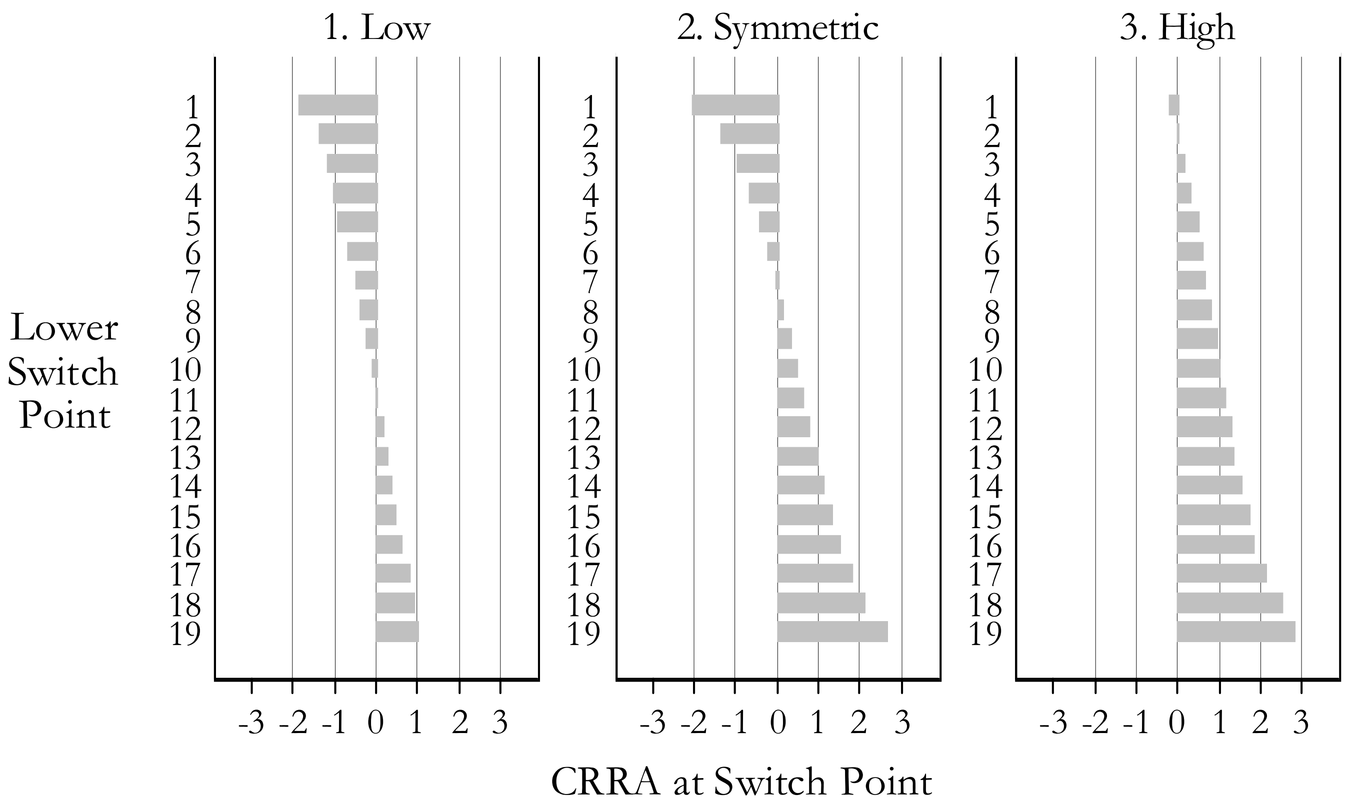
---

[18] In addition, given our design, we must admit to responding to some background risk. Although experienced at sportscard and related shows, this was the first coin show that we had participated. The older demographics we expected in our sample led us to be cautious that subjects would be as willing to participate in such an activity as we have experienced in sportscard markets. On the other hand, casual evidence from show activity indicates that coin shows are currently booming, much as sportscard markets did in the 1980's and early 1990's.

[19] Given the vagaries of hourly attendance, and a desire to have roughly balanced samples in each treatment cell, we did not elicit responses for the Ungraded Coins treatment on the third day of the show.

[20] The probabilities for the low frame are 0.06, 0.1, 0.12, 0.14, 0.16, 0.2, 0.24, 0.26, 0.3, 0.34, 0.36, 0.4, 0.44, 0.46, 0.5, 0.55, 0.6, 0.64, 0.66 and 0.7. The probabilities for the high frame are 0.3, 0.35, 0.4, 0.45, 0.5, 0.54, 0.56, 0.6, 0.64, 0.66, 0.7, 0.74, 0.76, 0.8, 0.84, 0.86, 0.9, 0.94 and 0.96.

Figure 1: Effect of Anchoring Treatment



a statistically detectable manner one needs to have several rows of separation. The design of Harrison, Lau, Rutström and Sullivan [2005] may have failed to detect a large effect due to their initial decision sheet only having 10 rows, and hence only having a separation of 2 rows in their comparable frames.

The instructions given to subjects were virtually identical across treatments. In the baseline case, the Money treatment with a symmetric frame, they were as follows:

> On the next page there is a decision sheet and each decision is a paired choice between OPTION A and OPTION B. You will make twenty choices and record these in the final column. Before you start making your twenty choices, let me explain how these choices will affect your potential earnings.
>
> Even though you will make twenty decisions, only one of these will end up affecting your earnings, but you will not know in advance which decision will be used. After completing the twenty choices you will roll a 20-sided die the result corresponds to the row from the decision sheet you will play. Obviously, each decision has an equal chance of being used in the end.
>
> Once the row has been selected we will use a 100-sided die to play the option A or B corresponding to your decision on the selected row. The final result will be

written on the game sheet and an index card which you will add to the raffle box. A raffle will be conducted on Sunday at 2pm to determine the participants that will receive payment. You will have a 1/20 chance of winning the prize on the index card which is the result from the upcoming game. You need not be present to win, we will mail the prize to you.

Now, lets discuss the decision sheet. Please look at Decision 1 at the top of the decision sheet:

OPTION A pays $200 if the die shows a number between 1 and 5. It pays $125 if the die shows between 6 and 100.

OPTION B yields $350 if the die shows a number between 1 and 5. It pays $40 if the die shows 6 and 100.

The other decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between $200 and $350. {The reference here to Decision 10 was an error, from an earlier version of the decision sheet. The experimenters corrected this orally during the experiment, noting that it was a typo since the bottom row was of course row 20.}

So now please look at the empty boxes on the right side of the record sheet. You will have to write a decision, A or B in each of these boxes, and then a roll of the dice will determine which choice will count. We will look at the decision that you made for the choice that counts, and circle it, before rolling a die again to determine your potential earnings. Then you will write your potential earnings in the blank at the bottom of the page.

Are there any questions?

In the Coins treatment, references to $200, $125, $350 and $40 were replaced by COIN A, COIN B, COIN C and COIN D, respectively. In that case the subjects were shown the coins. The first three lines of the decision sheet for the Money treatment appeared as follows:

	OPTION A	OPTION B	DECISION
1	5/100 of $200, 95/100 of $125	5/100 of $350, 95/100 of $40	A B
2	10/100 of $200, 90/100 of $125	10/100 of $350, 90/100 of $40	A B
3	15/100 of $200, 85/100 of $125	15/100 of $350, 85/100 of $40	A B

The comparable lines for the Coins treatment were:

	OPTION A	OPTION B	DECISION
1	5/100 of COIN A, 95/100 of COIN B	5/100 of COIN C, 95/100 of COIN D	A B
2	10/100 of COIN A, 90/100 of COIN B	10/100 of COIN C, 90/100 of COIN D	A B
3	15/100 of COIN A, 85/100 of COIN B	15/100 of COIN C, 85/100 of COIN D	A B

All responses were collected by "paper and pen," and an ID for each subject placed in a visible, sealed glass jar until the drawing near the end of the show.

After these decisions were made, we asked each subject to complete a short survey asking them for information about themselves. We asked several questions to characterize the depth of their experience in the coin market. Specifically, we asked how long they had been active in the coin and paper money market; whether they were a coin and paper money professional dealer; how many coin and paper money shows they attend in a typical year; how many of their coins had been professionally graded in a given year; whether they dealt only in graded coins, only in ungraded coins, or in both; and whether they were affiliated with a grading company. We also asked the subject to identify their sex, age, educational level, income, marital status, size of household, and smoking status.

Responses from 113 subjects were collected: 42 in the Money treatment, 38 in the Graded Coins treatment, and 33 in the Ungraded Coins treatment.

3. Results

Our main working hypothesis is that there are differences in risk attitudes elicited using the artificial Money frame as compared to the natural Coins frames. To examine this hypothesis we rely on a popular and simple CRRA characterization of the individual utility function. We then explore the sensitivity of our inferences to that assumed functional form, by considering more flexible utility specifications and more flexible specifications of the stochastic choice process.

A. Results Assuming Constant Relative Risk Aversion

We employ an interval regression model in which the dependent variable is the CRRA interval at which the subject switches. One attractive feature of this specification is that it allows us to treat subjects that "switch back" from option B to option A, after having chosen B at least once in a prior row, as having "fat preferences," or as effectively expressing indifference for the intermediate options. Such responses should correctly result in less precise estimates of the CRRA for the sample than subjects who only switch once from option A to option B. Further, the interval regression model allows for subjects whose responses only allow us to define an interval from minus infinity (extreme risk-loving) to some finite CRRA, or to define an interval from some finite CRRA to plus infinity (extreme risk-aversion).

We encountered an unusual amount of noise in the responses received by our subjects, at least compared to the behavior observed in comparable tasks in the lab and the field. One source of this noise was observed *a priori* during the experiment: many of the female respondents were spouses of male dealers, and did not appear to be as keen participants in the market as their spouse. We therefore exclude 28 females in the primary statistical analysis, but report the effects of including them. Another source of noise is that 14 of our 85 male subjects gave responses that could not be interpreted as including a finite CRRA interval, and which we dropped.[21] There is no evidence that the propensity to respond in this manner is related to any of our treatments.[22]

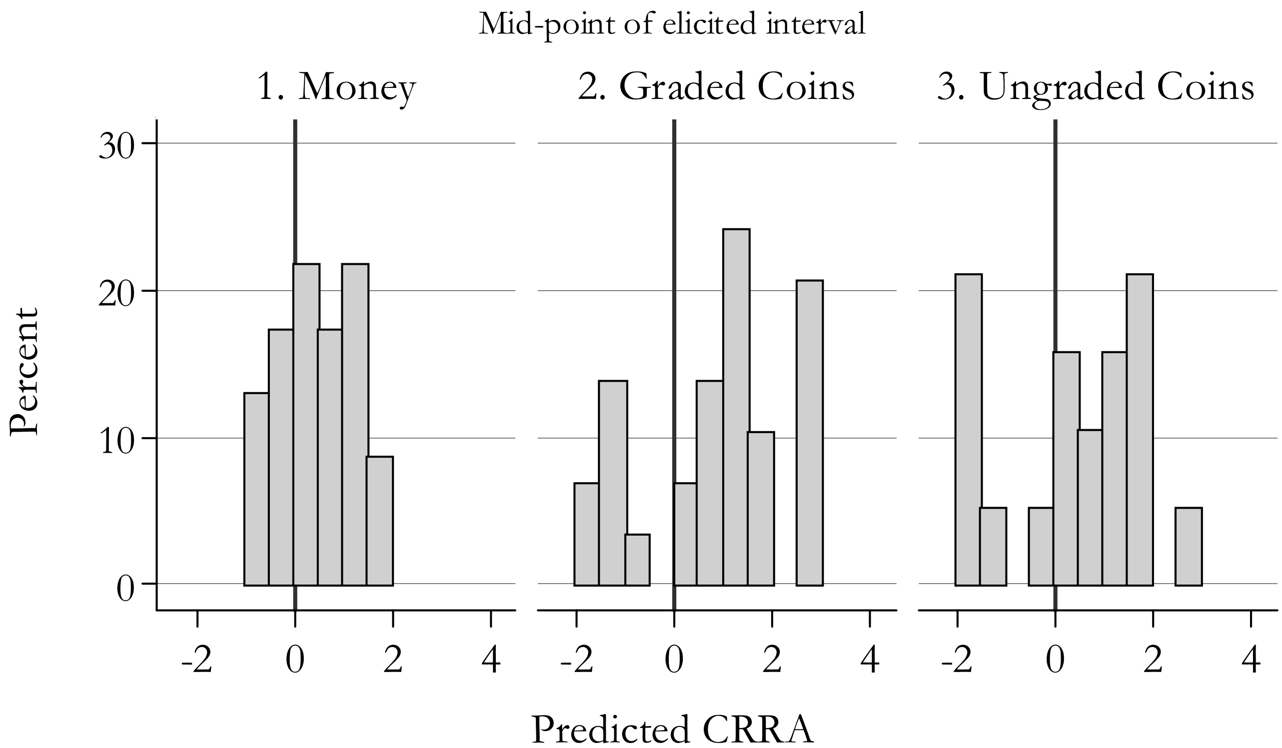
Table 2 provides descriptive statistics in the sample used in the primary estimation analysis, along with definitions of all variables. Figure 2 displays the raw data, in terms of the mid-points of the elicited interval in each of the main treatments.[23] These data generally exhibit greater variability

[21] These responses are dropped automatically by the interval regression model we employ. One could formulate likelihood functions that take their responses into account, using either (a) "mixture models" that allow for several distinct population processes to generate the observed data, or (b) "stochastic choice" models that allow for error terms to account for errant choice behavior. However, many of these subjects gave readily identifiable signs of being confused or unmotivated: for example, one respondent chose the pattern "BA" for each consecutive row, and another chose the pattern "BBAA" throughout.

[22] We estimated a probit model on a dummy variable denoting valid and invalid responses, and found no significant effects from any of our treatments.

[23] Those subjects whose choices imply an open interval, with the lower or upper bound CRRA being $-\infty$ or $+\infty$, were assigned mid-points equal to the finite value in their interval. The raw data can also be misleading with respect to those subjects who switched back and forth several times, since it fails to convey the relative imprecision of the elicited risk attitude. The subsequent statistical analysis correctly treats all of these subjects as having relatively imprecise elicited CRRA intervals.

Figure 2: Elicited Risk Attitudes By Context



than observed in comparable laboratory exercises over abstract Money prizes. Furthermore, there appears to be a significant increase in variability as one moves to the Graded and Ungraded Coins treatments. Because we understood elicitation tasks on a between-subjects basis, and our samples were not designed to be large enough for randomization to ensure comparable samples, we examine the effects of the treatments using statistical procedures which condition on observable characteristics of the individual, the experimental treatment, and our procedures.

Table 3 reports the main estimation results. We easily reject the homogeneity null hypothesis that all coefficients are jointly equal to zero. The treatment effects are relatively clear, with the Symmetric Money treatment as the norm. The Graded Coins treatment is associated with an average CRRA that is 0.10 higher, with a standard error of 0.43 and a *p*-value of 0.81. The 95% confidence interval for this effect is between -0.75 and +0.96. However, the Ungraded Coins treatment is associated with a large increase in the average CRRA of 2.43, with a *p*-value of 0.002. Thus we find a significant effect of the Ungraded Coins treatment, and no statistically significant effect from the

Graded Coins treatment. *These findings indicate that the use of artificial monetary prizes provides a reliable measure of risk attitudes when the natural counterpart outcome has minimal uncertainty, but that it can provide an unreliable measure when the natural counterpart outcome has background risk.* These results are consistent with conventional EUT for the effects of background risk.

The skewness frames are each associated with different measures of risk aversion in the direction expected. The skewLO treatment is associated with a reduction in measured risk aversion of 0.26, which is the predicted direction if subjects are anchoring in the middle of the decision table, but the effect is only significant at the 68% level. The skewHI treatment is associated with a significant increase in risk aversion of 2.74, and this effect is statistically significant at any conventional level. *Thus, we conclude that the positive skewness treatment did have a significant influence on elicited risk aversion, and that future exercises eliciting risk attitudes should control for such effects.*

We also observe that our procedural controls have a measurable effect on elicited risk aversion. We have an "experimenter effect" on responses collected by one particular experimenter, as well as an effect from eliciting responses on the third day of the show.[24] This effect may be correlated with the effect from eliciting responses on the third day, since the experimenter in question collected twice as many responses on that day as the other experimenter and the two effects largely offset each other. *Thus, one might have to worry more about procedural controls when eliciting risk attitudes in the field than in the lab*, which is consistent with the generic difficulties of ensuring the same controls in the field as in the lab.

Many individual subject characteristics appear to be significantly associated with the elicited risk attitude. Older subjects have lower risk aversion. Even though the effect may seem small, at a reduction in average CRRA of only 0.092 per year of age, our subjects have many years under their belt since the average is just over 50. Single subjects were more risk loving than others, as were those with larger households. Those with some college education had a higher aversion to risk, but those who had completed college did not. The poor and the rich are relatively more risk averse than

[24] Day 1 was a Friday afternoon, Day 2 was a full Saturday, and Day 3 was a slight-truncated Sunday. We often find in comparable shows that the Friday afternoon tends to be when one finds relatively more dealers and experienced traders.

others. These results are consistent with those from laboratory experiments in college settings and in the field: *there is considerable individual heterogeneity in risk attitudes, such that one should not readily assume homogenous risk preferences for the population.*

Turning to the trader characteristics of the subject as a coin collector or trader, we observe no significant effect from being a dealer. On the other hand, there are strong effects from other characteristics that are correlated with being a dealer: those with more experience in the coin market, some affiliation with a grading company, and having owned a Morgan Silver Dollar, exhibited more risk aversion. Conversely, attending more shows was associated with a lower risk aversion, although the effect is not large for those attending just 2 or 3 shows per year.

We also correct for the effects of the main treatments on residual variance, using a multiplicative heteroskedasticity specification. The Coins treatments is associated with a significant increase in unexplained variance, primarily due to the Ungraded Coins treatment (although the negative effect of the Graded Coins treatment is only significant at the 18.6% level). Both of the skewness treatments reduced unexplained variance, perhaps by restricting the domain of choices in probability-space to be considered.[25]

We generate predictions from this estimated model to examine the net effect of the treatments in elicited risk aversion. In this case all skewness effects were set to zero, to ensure comparability. We also assume that all subjects had the average characteristics of the pooled sample, so that we can control for sample differences in the three treatments. For example, if one treatment had slightly more single subjects as a fraction of the sample compared to the other treatment, and this characteristic is significantly associated with differences in elicited risk aversion, then there could be an effect of sample composition that is driving the differences in Figure 2. With this procedure we predict that the CRRA is -1.11 in the Money treatment, -1.06 in the Graded Coins treatment, and +1.33 in the Ungraded Coins treatment. *Thus we tend to see risk-loving behavior in the Money and Graded Coins treatments, and risk aversion in the Ungraded Coins treatment.* Although these estimates are not precise

[25] All three treatments used 20 rows in the decision table, but the skewLO frame offered probabilities between 0.06 and 0.70, the skewHI frame offered probabilities between 0.30 and 1, and the Symmetric frame offered probabilities between 0 and 1.

in some absolute sense, the statistical significance of the difference in the Ungraded Coins treatment is highlighted in Table 3.[26]

Including women in the analysis changes the results significantly. In terms of the primary treatments, neither Graded Coins nor Ungraded Coins has a significant effect on elicited risk aversion. In one case the *p*-value of the effect is 0.52, and in the other it is 0.87, indicating that women react to the coins treatments very differently than men. The skewness treatments have roughly the same effect on elicited risk aversion. And the experimenter effect becomes statistically insignificant. Average CRRA for the Money and Ungraded Coins treatments remain negative (-0.70 and -1.20, respectively), indicating a general risk-loving propensity, but the average CRRA for the Ungraded Coins treatment is now negative as well (-0.62).

B. More Flexible Specifications

There are two issues that arise with use of a CRRA specification of the utility function. First, and perhaps obviously, is it appropriate to assume that RRA is constant over the domain of prizes considered here? Second, the statistical evidence of significant residual heteroskedasticity leads one to wonder if a more flexible specification of the "stochastic choice" process might lead to different results. We consider each of these issues here.

Using the EP functional form introduced earlier, we undertook tests that show that CRRA is an appropriate characterization for this sample. We estimated an EP utility function in which the α parameter was allowed to be a linear function of the experimental treatments (whether the outcome was a graded or ungraded coin, which skewness treatment was employed, and which experimenter collected the responses). Our estimates were obtained via maximum likelihood, accounting for the correlation of responses provided by each subject. We then directly explore the CRRA specification by testing if the estimates of α are jointly equal to 0. The resulting χ^2 test statistic has 6 degrees of

[26] The methodological point of using comparable samples to predict treatment effects may be clearer if one considers what would happen if we use the estimated model to predict average CRRA values in each distinct sample. In this case we would estimate the average CRRA as 0.16 in the Money treatment, 0.24 in the Graded Coins treatment, and 1.2 in the Ungraded Coins treatment.

freedom, and is calculated to be 2.23 with a *p*-value of 0.90. Hence *we cannot reject the null hypothesis that CRRA is appropriate for this population and these tasks.*

However, this specification remains restrictive in another sense, also noted by Holt and Laury [2002]: it assumes deterministic choice behavior. In other words, it assumes that subjects behave as if they have some latent index function indicating what the expected utility of each lottery is, that they compare the values of that function for each of the two lotteries they are being asked to make a binary choice over, and that they exactly execute their choice even if the difference in expected utility is "small" in some sense. The probit function linking these choices to the latent index function then "tacks on" an error at the final stage of this decision process.

A methodologically important literature has evolved in which errors are allowed to enter structurally at different stages in this process.[27] For example, a "random preference error story" would allow for errors at the first stage of the above process: as the subject evaluates the expected utility of a lottery there is some error.[28] Or a "lottery comparison error story" would allow for errors at the second stage of the process: as the subject compared the expected utility of the two lotteries to determine which one was larger, some error is made.[29] Or a "tremble error story" would allow for errors at the end of the process: after the subject deterministically evaluated the expected utility of each lottery, and then deterministically compared the two to decide which is higher, a small mistake

[27] See Ballinger and Wilcox [1997] for a methodological review of this development. A parallel literature has evolved in which similar errors are allowed to enter the specification of equilibria in non-cooperative games. McKelvey and Palfrey [1995] and Goeree and Holt [1999] develop specifications that have been widely used to study experimental games. The idea of error-prone decisions has an even longer tradition in game theory, of course, with notions such as "trembling hand equilibria" and "ϵ-equilibria" being widely used.

[28] Becker, DeGroot and Marschak [1963] first proposed this specification, which has been popularized by Loomes and Sugden [1995]. In most of the specifications of random preference the same error is attached to the evaluation of each lottery. This avoids the possibility of stochastically dominated lotteries being chosen, on the grounds that this violates *a priori* beliefs about behavior or evidence that subjects do not make these types of violations. However, if one also allows an error to occur at any of the subsequent stages in the decision process, this implication does not follow. In that case one might argue that the logic of the random preference approach suggests that there should be two distinct error terms for each lottery, perhaps reflecting the complexity of evaluating the lottery (e.g., as noted by Wilcox [1993], lotteries with two prizes, or with simple probabilities such as 0.5, are easier to cognitively evaluate than other lotteries).

[29] Becker, DeGroot and Marschak [1963] also were the first to propose this specification, which has been developed by Hey and Orme [1994] and Hey [1995]. If the process linking the latent function to the likelihood of the data is of the probit form, this implies an additive error term to the expected utility difference, and is often called a "Fechner error term." Although they did not use a probit (or logit) specification, Holt and Laury [2002] employed an error term which acts to change the relative weight of the expected utilities of the two lotteries in a similar manner vis-a-vis the structural stage of the assumed decision process.

is made when actually implementing that choice.[30] Although the evidence for one or the other error story is mixed, it appears to support either of the first two over the last when one assumes conventional EUT models of the underlying preference functional.[31]

We implement an error model in which subjects make some error at the lottery comparison stage of the assumed decision process. We extend the literature by allowing this error term to be a linear function of experimental treatments, just as in our earlier specification of multiplicative heteroskedasticity. We find errors in the Money treatment are positive (+0.64) and statistically significant (*p*-value of 0.088), but that errors in the Graded and Ungraded treatments are positive but much smaller and statistically insignificant (estimated values of +0.10 and +0.25, respectively, with *p*-values of 0.82 and 0.40). None of the treatments had any statistically significant effect on errors. Thus we conclude that our subjects appear to be *much better able to make consistent decisions about uncertain prospects when the uncertainty derives from a natural and familiar process than when it derives from being asked to evaluate prospects couched in terms of an abstract Money frame.*

In addition, allowing for the nature of the stimuli to affect the error has implications for whether one would conclude that these subjects had a CRRA-consistent utility specification or not. When we include the full set of treatments as determinants of the error, as above, we continue to be unable to reject the hypothesis that α is zero, so we find support for CRRA (the χ^2 test for this hypothesis has a value of 7.21 and 6 degrees of freedom, hence a *p*-value of 0.30). But if we restrict the error to be a scalar across the different stimuli and treatments, we convincingly reject CRRA (χ^2 value of 39.1, for a *p*-value less than 0.001). Hence if one simply included a structural error specification without conditioning on the differences in stimuli in our treatments, one would erroneously reject a popular functional form, CRRA. Just as the structural error specification is starting to open up the behavioral black box of errors that subjects are prone to make, our explicit recognition of the effects of artificial and natural stimuli appears to be informative with respect to

[30] Harless and Camerer [1994] developed this approach in the context of testing expecting utility theory with experiments.

[31] See Carbonne [1997]. The evidence becomes mixed when one simultaneously evaluates alternative error stories *and* alternative underlying preference functionals: see Carbonne and Hey [1999] and Loomes, Moffatt and Sugden [2002]. Our analysis is restricted to conventional EUT specifications.

the characterization of functional form for utility under uncertainty.

4. Conclusions

Does individual behavior in a laboratory setting provide a reliable indicator of behavior in a naturally occurring setting? We consider this general methodological question in the context of eliciting risk attitudes. We find that the use of artificial monetary prizes provides a reliable measure of risk attitudes when the natural counterpart outcome has minimal uncertainty, but that it can provide an unreliable measure when the natural counterpart outcome has background risk. These results are consistent with conventional EUT for the effects of background risk on attitudes to risk. Behavior tended to be risk loving when artificial monetary prizes were used or when there was minimal uncertainty in the natural non-monetary outcome. But subjects drawn from the same population were risk averse when their attitudes were elicited using the natural non-monetary outcome that had some background risk. Theory predicts this effect of background risk, but not the change from risk-loving to risk-aversion.

Our findings therefore contribute to the broader evaluation of the effect of "context" on elicited preferences. Risk attitudes depend on outcomes, on subjective probabilities of outcomes, and on the framing of the choices. Our field experiments allow us to study the effect of a naturally occurring context in one of these dimensions, the outcome. This limitation was deliberate, to allow clearer inferences within design and budgetary limits that would be free of possible confounds. And our choice of context was also deliberate: we see limited relevance in studies showing just that context matters if that context is patently artificial. The challenge is to identify those aspects of context that matter for the task at hand.[32]

The practical implications for policy are twofold. First, estimates of risk aversion elicited using monetary outcomes may be reliable for non-monetary outcomes that do not involve a considerable amount of background risk. However, they should not be applied directly to non-

[32] For example, Carlson and Clark [1992] examine what types of context matter for language comprehension, in an effort to avoid making references to context "a kind of conceptual garbage can" (p.61). Similarly, Wilcox [1993] examines certain aspects of the complexity of abstract lotteries from a cognitive perspective in lab experiments.

monetary outcomes with significant background risk. At the very least they should be viewed as lower-bound estimates, consistent with the predictions of standard theory about the effects of background risk on risk attitudes.[33] Second, one should not assume in welfare evaluation under uncertainty that all individuals have the same attitudes to risk. Consistent with laboratory evidence, we identify statistically significant evidence that observable individual heterogeneity is correlated with risk attitudes. Thus one should apply appropriate measures of risk attitudes when evaluating welfare impacts for different individuals or households.

[33] We certainly do not propose that the *quantitative* difference between our Money and Graded treatments be used to calibrate between, say, Money and health outcomes. Such calibration factors should be generated on a case-by-case basis for appropriate populations.

Table 1: Payoff Matrix in the Holt and Laury Risk Aversion Experiments

Lottery A				Lottery B				EV^A	EV^B	Difference	Open CRRA Interval if Subject Switches to Lottery B
p($2)		p($1.60)		p($3.85)		p($0.10)					
0.1	$2	0.9	$1.60	0.1	$3.85	0.9	$0.10	$1.64	$0.48	$1.17	-∞, -0.95
0.2	$2	0.8	$1.60	0.2	$3.85	0.8	$0.10	$1.68	$0.85	$0.83	-∞, -0.95
0.3	$2	0.7	$1.60	0.3	$3.85	0.7	$0.10	$1.72	$1.23	$0.49	-0.95, -0.49
0.4	$2	0.6	$1.60	0.4	$3.85	0.6	$0.10	$1.76	$1.60	$0.16	-0.49, -0.15
0.5	$2	0.5	$1.60	0.5	$3.85	0.5	$0.10	$1.80	$1.98	-$0.17	-0.15, 0.14
0.6	$2	0.4	$1.60	0.6	$3.85	0.4	$0.10	$1.84	$2.35	-$0.51	0.14, 0.41
0.7	$2	0.3	$1.60	0.7	$3.85	0.3	$0.10	$1.88	$2.73	-$0.84	0.41, 0.68
0.8	$2	0.2	$1.60	0.8	$3.85	0.2	$0.10	$1.92	$3.10	-$1.18	0.68, 0.97
0.9	$2	0.1	$1.60	0.9	$3.85	0.1	$0.10	$1.96	$3.48	-$1.52	0.97, 1.37
1	$2	0	$1.60	1	$3.85	0	$0.10	$2.00	$3.85	-$1.85	1.37, ∞

Table 2: Descriptive Statistics

N = 85; males only.

Sample collected at The Central Florida Coin Show, April 16-18, 2004.

Variable	Description	Mean	Standard Deviation	Minimum	Maximum
	Treatments				
money	Money treatment	0.34	0.48	0	1
graded	Graded treatment	0.37	0.48	0	1
ungraded	Ungraded treatment	0.28	0.45	0	1
skewLO	Frame to skew RA lower	0.35	0.48	0	1
skewHI	Frame to skew RA higher	0.35	0.48	0	1
	Procedural Controls				
experimenter	Charles Towe effect	0.70	0.46	0	1
day1	Day 1 of show	0.28	0.46	0	1
day3	Day 3 of show	0.19	0.39	0	1
	Individual Demographic Characteristics				
age	Age	51.46	13.25	18	84
single	Single and never married	0.15	0.36	0	1
nhhd	Size of household	2.52	1.40	0	7
Scollege	Some College education or post-high school	0.12	0.32	0	1
college	College education or higher	0.58	0.50	0	1
poor	Income under $30k per year	0.12	0.32	0	1
rich	Income over $75k per year	0.34	0.48	0	1
	Individual Trader Characteristics				
yrexp	Years of experience in coins	21.77	17.46	0	75
shows	Number of shows attended in a year	7.86	10.81	0	50
grcomp	Affiliated with a grading company	0.09	0.29	0	1
ms	Ever owned Morgan Silver dollars	0.89	0.31	0	1
dealer	Coin dealer	0.34	0.48	0	1
grade	Has coins graded	0.65	0.48	0	1

Table 3: Interval Regression Model of Risk Attitudes

Variable	Description	Estimate	Standard Error	z	*p*-value	95% confidence interval lower bound	95% confidence interval upper bound
	Mean Effects						
Constant		2.65	0.84	3.17	0.00	1.01	4.28
	Treatments						
graded	Graded treatment	0.10	0.43	0.24	0.81	-0.75	0.96
ungraded	Ungraded treatment	2.44	0.79	3.09	0.00	0.89	3.98
skewLO	Frame to skew RA lower	-0.26	0.63	-0.41	0.68	-1.50	0.98
skewHI	Frame to skew RA higher	2.75	0.74	3.72	0.00	1.30	4.19
	Procedural Controls						
experimenter	Charles Towe effect	-2.43	0.43	-5.67	0.00	-3.27	-1.59
day1	Day 1 of show	-0.32	0.57	-0.56	0.57	-1.45	0.80
day3	Day 3 of show	2.07	0.36	5.69	0.00	1.36	2.79
	Individual Demographic Characteristics						
age	Age	-0.09	0.01	-6.94	0.00	-0.12	-0.07
single	Single and never married	-2.10	0.32	-6.65	0.00	-2.72	-1.48
nhhd	Size of household	-0.71	0.24	-2.96	0.00	-1.19	-0.24
Scollege	Some College education or post-high school	1.43	0.68	2.12	0.03	0.11	2.76
college	College education or higher	-0.11	0.21	-0.52	0.61	-0.53	0.31
poor	Income under $30k per year	1.00	0.52	1.91	0.06	-0.03	2.02
rich	Income over $75k per year	2.49	0.54	4.57	0.00	1.42	3.56
	Individual Trader Characteristics						
yrexp	Years of experience in coins	0.02	0.00	4.91	0.00	0.01	0.03
shows	Number of shows attended in a year	-0.09	0.01	-7.23	0.00	-0.11	-0.06
grcomp	Affiliated with a grading company	2.04	0.65	3.12	0.00	0.76	3.33
ms	Ever owned Morgan Silver dollars	3.86	0.83	4.67	0.00	2.24	5.48
dealer	Coin dealer	-0.28	0.37	-0.78	0.44	-1.00	0.43
grade	Has coins graded	0.11	0.30	0.38	0.70	-0.47	0.69
	Residual Variance Effects						
graded	Graded treatment	2.32	0.65	3.56	0.00	1.05	3.60
ungraded	Ungraded treatment	4.12	1.58	2.62	0.01	1.03	7.21
skewLO	Frame to skew RA lower	-4.17	1.62	-2.58	0.01	-7.33	-1.00
skewHI	Frame to skew RA higher	-2.11	0.78	-2.70	0.01	-3.63	-0.58

Notes: Sample size is 70. A Wald test that all coefficients are jointly equal to zero has a χ^2 value of 24163 and 20 degrees of freedom, hence a *p*-value < 0.001.

References

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

Battalio, Raymond C.; Kagel, John C., and Jiranyakul, K., "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty*, 3, 1990, 25-50.

Beck, John H., "An Experimental Test of Preferences for the Distribution of Income and Individual Risk Aversion," *Eastern Economic Journal*, 20(2), Spring 1994, 131-145.

Becker, Gordon; DeGroot, Morris, and Marschak, Jacob, "Stochastic Models of Choice Behavior," *Behavioral Science*, 8, 1964, 41-55.

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

Binswanger, Hans P., "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal*, 91, December 1981, 867-890.

Carbone, Enrica, "Investigation of Preference Theory Using Experimental Data," *Economics Letters*, 57, 1998, 305-312.

Carbone, Enrica, and Hey, John, "Estimation of Expected Utility and Non-Expected Utility Preference Functionals Using Complete Ranking Data," in B. Munier and M. Machina (eds.), *Models and Experiments on Risk and Rationality* (Dordrecht: Kluwer, 1994).

Carbone, Enrica, and Hey, John, "Which Error Story Is Best?," *Discussion Paper 33*, Department of Economics, University of York, 1999.

Carlson, Thomas B., and Clark, Herbert H., "Context for Comprehension," in H.E. Clark, *Arenas of Language Use* (Chicago: University of Chicago Press, 1992).

Eeckhoudt, Louis; Gollier, Christian, and Schlesinger, Harris, "Changes in Background Risk and Risk-Taking Behavior," *Econometrica*, 64(3), May 1996, 683-689.

Goeree, Jacob K. and Holt, Charles A., "Stochastic Game Theory: For Playing Games, Not Just For Doing Theory," *Proceedings of the National Academy of Sciences*, 96, 1999, 10564-10567.

Gollier, Christian, and Pratt, John W., "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica*, 64(5), September 1996, 1109-1123.

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

Harrison, Glenn W., "An Experimental Test for Risk Aversion," *Economics Letters*, 21(1), 1986, 7-11.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Temporal Stability of Estimates of Risk Aversion," *Working Paper 3-17*, Department of Economics, College of Business Administration, University of Central Florida, 2003a; forthcoming, *Applied Financial Economics Letters*.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *Working Paper 3-19*, Department of Economics, College of Business Administration, University of Central Florida, 2003b; forthcoming, *American Economic Review*.

Harrison, Glenn W.; Lau, Morten I; and Rutström, E. Elisabet, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Working Paper 04-07*, Department of Economics, College of Business Administration, University of Central Florida, 2004.

Harrison, Glenn W.; Lau, Morten I; 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).

Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1013-1059.

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

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

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

Kachelmeier, Steven J., and Shehata, Mohamed, "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82(5), December 1992, 1120-1141.

Kagel, John H.; MacDonald, Don N., and Battalio, Raymond C., "Tests of 'Fanning Out' of Indifference Curves: Results from Animal and Human Experiments," *American Economic Review*, 80(4), September 1990, 912-921.

Kimball, Miles S., "Standard Risk Aversion," *Econometrica*, 61(3), May 1993, 589-611.

Laury, Susan K., and Holt, Charles A., "Further Reflections on Prospect Theory," *Working Paper*, Department of Economics, Georgia State University, 2002; available at http://www.people.virginia.edu/~cah2k/reflect.pdf

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.

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

Lusk, Jayson L., and Coble, Keith H., "Risk Perceptions, Risk Preference, and Acceptance of Risky Food," *Working Paper*, Department of Agricultural Economics, Purdue University, 2004a; forthcoming, *American Journal of Agricultural Economics*.

Lusk, Jayson L., and Coble, Keith H., "Risk Aversion in the Presence of Background Risk: Evidence from the Lab," *Working Paper*, Department of Agricultural Economics, Purdue University, 2004b.

McKelvey, Richard D., and Palfrey, Thomas R., "Quantal Response Equilibria for Normal Form

Games," *Games and Economic Behavior*, 10, 1995, 6-38.

Miller, Louis; Meyer, David Edward, and Lanzetta, John T., "Choice Among Equal Expected Value Alternatives: Sequential Effects of Winning Probability Level on Risk Preferences," *Journal of Experimental Psychology*, 79(3), 1969, 419-423.

Murnighan, J. Keith; Roth, Alvin E., and Shoumaker, Francoise, "Risk Aversion and Bargaining: Some Preliminary Results," *European Economic Review*, 31, 1987, 265-271.

Murnighan, J. Keith; Roth, Alvin E., and Shoumaker, Francoise, "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty*, 1(1), March 1988, 101-124.

Pratt, John W., "Aversion to One Risk in the Presence of Others," *Journal of Risk and Uncertainty*, 1(4), December 1988, 395-413.

Pratt, John W., and Zeckhauser, Richard J., "Proper Risk Aversion," *Econometrica*, 55(1), January 1987, 143-154.

Quiggin, John, "Background Risk in Generalized Expected Utility Theory," *Economic Theory*, 22(3), October 2003, 607-611.

Safra, Zvi, and Segal, Uzi, "Constant Risk Aversion," *Journal of Economic Theory*, 83(1), November 1998, 19-42.

Wilcox, Nathaniel T., "Lottery choice: Incentives, complexity, and decision time," *Economic Journal*, 103, 1993, 1397-1417.