

NATURALLY OCCURRING PREFERENCES AND EXOGENOUS LABORATORY EXPERIMENTS: A CASE STUDY OF RISK AVERSION

BY GLENN W. HARRISON, JOHN A. LIST, AND CHARLES TOWE¹

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. Because 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. Behavior tended to be moderately risk averse when artificial monetary prizes were used or when there was minimal uncertainty in the natural nonmonetary outcome, but subjects drawn from the same population were much more risk averse when their attitudes were elicited using the natural nonmonetary outcome that had some background risk. These results are consistent with conventional expected utility theory for the effects of background risk on attitudes to risk.

KEYWORDS: Risk aversion, background risk, field experiments, laboratory experiments.

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 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.² In the traditional paradigm, risk aversion is viewed in

¹We are grateful to Michael Price for assistance with running the experiments and to Steffen Andersen, a co-editor, and three referees for helpful comments. Harrison thanks the U.S. National Science Foundation for research support under grant NSF/HSD 0527675. Supporting data and instructions are stored in the Experimental Sciences Digital Archive located at <http://exlab.bus.ucf.edu>.

²The other is the characterization of preferences across arguments of the utility function that are atemporal (e.g., beer and pizza) or time dated (e.g., beer today or beer tomorrow). We focus on just one commodity and one time dating.

terms of diminishing marginal utility of the final prize in some abstract lottery.³ 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 that 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 that type of 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 previously experienced, so that they can use whatever heuristics they have developed for that commodity when making their choices. Finally, it means developing ways to communicate probabilities that correspond to the language with which 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 toward a specific “foreground risk” that are elicited. In many field settings it is not possible to artificially identify attitudes toward 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

³There is some debate as to whether the arguments of the utility function should be terminal wealth, income, or prizes that are final outcomes from some specific choice. We offer an “agnostic empirical argument” in Section 3 that the subjects behave as if lottery prizes; the argument. Irrespective of the position that one takes on this debate, one can view our estimates of risk attitude as informative from two different perspectives. One is from the perspective of expected utility theory in which utility is defined over lottery prizes; and the other is from prospect theory (and related alternatives to standard theory) in which the value function is also defined over lottery prizes. In the latter case one would also be interested in controls for probability weighting and, if applicable, loss aversion; we include such controls subsequently. We recognize that some prospect theory observers reject the view that expected utility can be defined coherently over lottery prizes, and we further accept the view that *if one defines utility over terminal wealth* that there are strong a priori arguments to expect subjects to be risk neutral over the lottery prizes used here (Rabin (2000)). However, our objective is not to resolve this debate, even though we have crisp views on the matter that accord with Cox and Sadiraj (2006) and Rubinstein (2002). Instead, our goal is to inform a wide range of observers (those who use expected utility theory and view utility as defined over lottery prizes, and those who use prospect theory) about the effect on risk attitudes of our treatments.

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.⁴ 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) referred to this type of behavior as *risk vulnerability* and showed 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).⁵

Although 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 of a *small* independent risk. Making use of the concept of constant risk aversion presented by Safra and Segal (1998), Quiggin (2003) showed 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) showed that they are complements in a set of alternative models. Of course, what constitutes a “sufficiently high initial risk” and a “small independent risk” is in the eyes of the beholder.⁶

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,⁷ but that it can provide an unreliable measure in a natural

⁴For example, see Pratt and Zeckhauser (1987), Pratt (1988), Kimball (1993), Gollier, and Pratt (1996), and Eeckhoudt, Gollier, and Schlesinger (1996).

⁵Eeckhoudt, Gollier, and Schlesinger (1996) extended 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.

⁶The theoretical literature provides few insights when background risk is added in the case of risk loving preferences. A slight modification of the Quiggin (2003) analysis of constant risk aversion with risk loving behavior preserves the property of diminishing sensitivity (i.e., preferences will be closer to risk neutrality in the presence of independent background risk). For expected utility preferences, the theory is less clear-cut.

⁷Lusk and Coble (2005) reported evidence consistent with this conclusion, comparing risk preferences elicited for an artificial monetary instrument and comparable preferences for an instrument defined over genetically modified food.

context when the natural counterpart outcome has background risk.⁸ These results are consistent with the available theory from conventional expected utility theory (EUT) for the effects of background risk on attitudes to risk.

Another important inference gained from these results concerns the applicability of laboratory results to the field. The good news for the relevance of findings from previous laboratory studies is that there is little evidence to suggest that risk attitudes vary significantly due to the nature of the commodity. Yet, tempering this optimism is the fact that the observed risk posture depends critically on the level of background risk, a multidimensional phenomenon that has yet to be extensively explored in the lab. This empirical insight calls into question the generalizability of the extant literature purporting to measure risk attitudes in the lab *and* the field if they fail to measure or control for the potential confound of background risk. It is not the case that field estimates of risk attitudes are correct and lab estimates are inherently flawed: both are conditional on assumptions made about background risk and other theoretically relevant variables. However, it is the case that eliciting risk attitudes in a natural field setting requires one to consider the nature and degree of background risk, because it is inappropriate to ignore.

To make this point more succinctly, consider the elicitation of the value that a person places on safety, a critical input in the cost–benefit assessment of environmental policy such as the Clean Air Act (U.S. Environmental Protection Agency (1997)). Conventional procedures to measure such preferences focus on monetary values to avoid mortality risk by asking subjects to value scenarios in which they face different risks of death. The traditional interpretation of responses to such questions ignores the fact that it is hard to imagine a physical risk that could kill you with some probability but that would leave you alive and have no effect whatsoever on your health. Of course, such risks exist, but most of the environmental risks of concern for policy do not fall into such a category. In general, then, responses to the foreground risk question should allow for the fact that the subject likely perceived some background risk. This example represents an important policy issue and highlights the import of the theoretical literature on background risk.

A further virtue of extending lab procedures to the field, therefore, is to encourage richer lab designs by forcing the analyst to account for realistic features of the natural environment that have been placed aside. In virtually any market with asymmetric information, whether it is a coins market, an open air market, or a stock exchange, a central issue is the quality of the object being traded. This issue, and attendant uncertainty, arises naturally. In many markets, the grade of the object, or professional certification of the seller, is one of

⁸Lusk and Coble (2007) found that adding abstract background risk to an elicitation procedure using artificial monetary outcomes also generates more risk aversion, although they did not find the effect to be large quantitatively. Our results do not address the effect of background risk on attitudes in a lab context.

the critical variables that determine price. Thus one could scarcely design a test of foreground risk in these markets without attending to the background risk. In our field experiment we exploit the fact that such risks can be exogenously controlled in these settings, and in a manner consistent with the predictions of theory.⁹

A final insight is that we observe considerable individual heterogeneity in risk attitudes, such that one should not readily assume homogeneous 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.

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, and Marschak (1964) procedure (e.g., Harrison (1986) and Kachelmeir and Shehata (1992)). The other is by observing choices that subjects make over lotteries that vary the prizes offered for given probabilities (e.g., Binswanger (1980, 1981)) or varying the probabilities of winning given prizes (e.g., Holt and Laury (2002)). We utilize the latter method, because it has been widely implemented in recent laboratory experiments and involves a relatively transparent task. We limit ourselves to characterization of risks framed in terms of gains.

Holt and Laury (2002) (HL) devised 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 denote A or B. Table I 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 to 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

⁹However, because we do not know the subject probability distribution of background risk in the field, we cannot know if background risk is statistically independent of the foreground risk. We can think of no reason why the two might be correlated, but this illustrates again the type of trade-off one experiences with field experiments. It also points to the complementary nature of field and lab experiments: Lusk and Coble (2007) showed that independent background risk in a lab setting is associated with an increase in foreground risk aversion.

TABLE I
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
$p(\$2)$	Outcome	$p(\$1.60)$	Outcome	$p(\$3.85)$	Outcome	$p(\$0.10)$	Outcome				
0.1	\$2	0.9	\$1.60	0.1	\$3.85	0.9	\$0.10	\$1.64	\$0.48	\$1.17	$-\infty, -0.95$
0.2	\$2	0.8	\$1.60	0.2	\$3.85	0.8	\$0.10	\$1.68	\$0.85	\$0.83	$-\infty, -0.95$
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, ∞

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.¹⁰ 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: it 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.

This procedure has now been widely employed in the laboratory. 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) examined behavior of Danish citizens confronted with an extension of the HL procedure. They found evidence of risk aversion, in general, and considerable heterogeneity associated with observable characteristics of the sample.

2. DESIGN OF THE FIELD EXPERIMENTS

2.1. *General Design*

Our objective is 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 that can easily confound inferences about risk attitudes, we employed

¹⁰As Laury (2005) noted, it has become increasingly common in economics experiments to elicit a series of choices from participants and then pay for only one (selected at random). This approach allows the researcher to observe a large number of individual decisions and to increase payoffs for each decision because only one of them will be used for payment. An important question then becomes, "Do subjects behave as if each of these choices involves the stated payoffs or do subjects scale-down payoffs to account for the random selection that is made?" Laury (2005) reported one laboratory experiment that tests this directly in an environment that shares similarities with our experiment. In an environment where payoff scale effects have been demonstrated to matter, three treatments are conducted: pay for 1 of 10 choices under low payoffs, pay for all 10 choices under low payoffs, and pay for 1 of 10 choices under $10\times$ the low payoff level. Increasing payoff scale has a significant effect on choices compared with the low payoff treatments where all 10 decisions are paid or where one decision is paid. However, there is no significant difference in choices between paying for 1 or all 10 decisions at the low payoff level. Similar evidence was reported by Starmer, and Sugden (1991) and Cubitt, Starmer, and Sugden (1998). Moreover, we should note that this feature is common across all of our treatments. Thus one would have to argue that there is an interaction effect with the treatments for it to be a confound for our inferences.

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 this design, the grade is a scalar, typically from 1 to 70 as subsequently discussed further. In the *ungraded coins* treatment, we employ exactly the same graded commodity, but without certification of the grade. Thus, we add “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 because 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, because 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 multiple price list with which 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.

2.2. *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, and between 1878 and 1904 in the New Orleans mint.¹² All were struck in 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 in Figure 1.

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 graded as a 1. In 1986 the Sheldon scale was formalized by the Professional Coin Grading Ser-

¹¹The 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 the design of William Barber for the dollar. Morgan was passed over for the job of Chief Engraver 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.

¹²Some 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.

¹³These 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; GV, government issue price. In addition, designations exist for color (RD, red; RB, red-brown; 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; DC, deep cameo).



FIGURE 1.—A Morgan silver dollar.

vice (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.

¹⁴Web page <http://www.pcg.com> contains information on the grading scales employed, historical background to the grading schemes, and prices on many rare coins including the 1897 Morgans. Coins graded by PCGS can be verified at <http://www.pcg.com/verify.shtml>.

The specific coins we employ are MS/PR-63, MS/PR-65, MS/PR-65PL, and 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 prooflike surface.
- 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 in Figure 2 illustrates a PCGS-graded Morgan dollar. The circled number is a sonically sealed certification number, which can be entered on the PCGS web site to



FIGURE 2.—Graded Morgan silver dollar.

¹⁵Retail values were obtained from the PCGS web site, <http://www.pcg.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.

verify the date, mint mark, 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, The Central Florida Coin Show, held April 16–18, 2004, at the Orlando Expo Center. We offered subjects \$5 to participate 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 survey questions to better characterize the individual.¹⁶

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 that reflected our 3×3 design. We randomized the money, graded coins, and ungraded coins treatments every hour, over the three days of the show.¹⁷

Figure 3 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. One can use CRRA values to characterize the switch point in each row, following Holt and Laury (2002, Table 3, p. 1649). 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 midpoint 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 3.¹⁸ 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

¹⁶In addition, given our design, we must admit to responding to some background risk. Although experienced at sports card and related shows, this was the first coin show that in which 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 sports card markets. On the other hand, casual evidence from show activity indicates that coin shows are currently booming, much as sports card markets did in the 1980s and early 1990s.

¹⁷Given 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. We find no differences in observed risk posture over the three days.

¹⁸The 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.

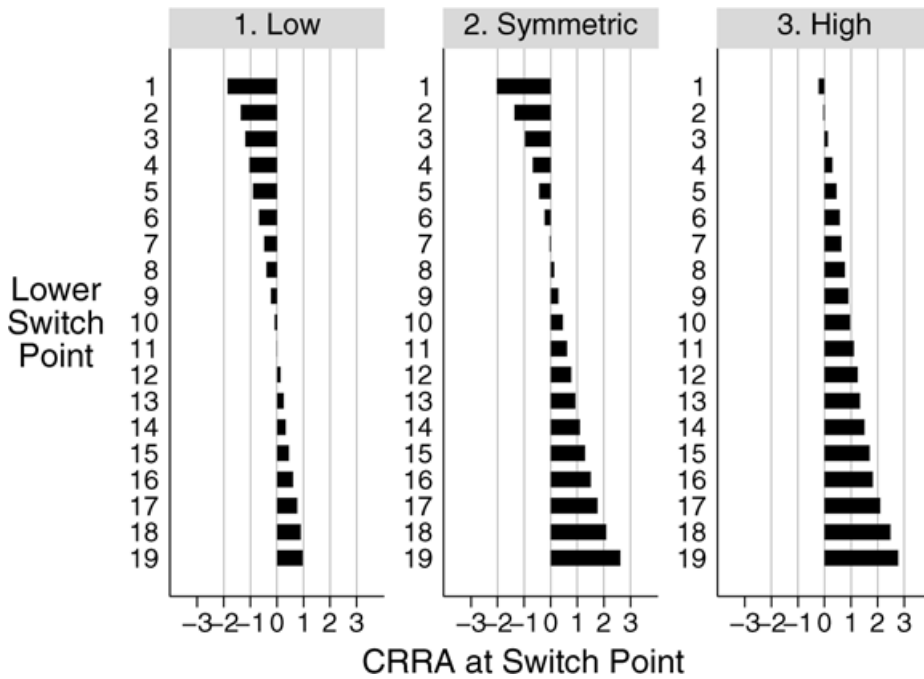


FIGURE 3.—Effect of anchoring treatment.

may reveal a behavioral pull in that direction. For this to occur in a statistically detectable manner, one needs to have several rows of separation.¹⁹

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

¹⁹The design of Harrison, Lau, Rutström, and Sullivan (2005) may have failed to detect a large effect because to their initial decision sheet had only 10 rows and, hence, only a separation of 2 rows in their comparable frames.

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.²⁰ You need not be present to win, we will mail the prize to you.

Now, let's 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 between 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 20 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 \$350.²¹

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 the following:

	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

²⁰We implemented this random draw by adding a prize every time the number of subjects exceeded a multiple of 20. In fact, we had 6 prizes, so each of the 113 subjects actually had a 0.053 chance of winning.

²¹The reference in these instructions to Decision 20 was actually typed as Decision 10 in the instructions provided. This 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 because the bottom row was of course row 20. No subject appeared concerned with this typo.

All responses were collected by “paper and pencil,” and an identification for each subject was 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 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 for the sex, age, educational level, income, marital status, size of household, and smoking status of the subject.

Responses from 113 subjects were collected: 42 in the money treatment, 38 in the graded coins treatment, and 33 in the ungraded coins treatment. Subjects were successfully randomized to the anchoring frame.²²

3. RESULTS

Over all 2,260 binary choices, 48.5% were for the safe lottery. In the money treatment, we observed 43.6% of the choices for the safe lottery, whereas in the graded coins treatment, we observed 51.8% safe choices and in the ungraded coins treatment, we observed 51.1% safe choices. These choices are not independent, of course, because each subject made 20 binary choices. Furthermore, we have an anchoring frame that varies from subject to subject. However, these raw responses do *tentatively* suggest that there was an increase in risk aversion when we compare the coins treatments to the money treatment, and that there was virtually no difference between risk attitudes elicited in the graded coins and ungraded coins treatments.

To go beyond this casual evidence, we require some economic structure and statistical structure. We describe each informally and formally, and then discuss estimation results. We discover that the unconditional inferences from the raw results are misleading with respect to the graded coins treatment: the graded coins and money treatments elicited comparable risk attitudes, and the ungraded coins treatment elicited responses consistent with greater risk aversion than either of the other two.

²²In fact we had 19, 12, and 11 subjects from the money treatment in each of low-skew, symmetric, and high-skew anchoring frames, respectively; 10, 10, and 18 in the graded coins treatment; and 12, 11, and 10 in the ungraded coins treatment. A Fisher exact test of the null hypothesis of independence of treatment and frame has a p -value of 0.29. Nonetheless, we control for the frames in our statistical analysis because we do have more of the low-skew subjects in the money treatment, and more of the high-skew subjects in the graded coins treatment, and we want to avoid those sampling differences as a possible confound.

3.1. *Informal Specification Assumptions*

The economic structure we need is provided by specifying a parametric CRRA utility function and some assumptions about the relationship between utility and choices. One natural assumption for this relationship is EUT: that the subjects took the stated probabilities of the foreground risk task as the decision weights to calculate expected utility in the standard manner. Another natural assumption is prospect theory (PT): that the subjects used some probability weighting function to transform the stated probabilities of the foreground risk task into decision weights that are then used to calculate expected utility (sometimes called prospective utility). To simplify, we assume the single-parameter probability weighting function used by Tversky and Kahneman (1992), recognizing that there are more flexible functional forms available.²³

The statistical structure we need is provided by allowing for a “stochastic error” in choices, correcting for the fact that sampling errors may be clustered for the same subject and specifying a likelihood function that relates latent specifications of preferences to observed choices. The stochastic error specification we adopt allows the subject to make an error when they are comparing the expected utility of the two lotteries and follows the “Fechner” specification employed by Hey and Orme (1994) in their analysis of comparable binary choices.²⁴ We assume that the responses from the same subject might be clustered and we correct for unobserved individual effects using methods that are standard in the statistical survey literature.²⁵ We allow the parameters of the statistical model to be linear functions of observable characteristics of the individual and task, because our design is between subjects in terms of the treatments of interest.

²³We adopt the “separable prospect theory” of Kahneman and Tversky (1979), rather than the later cumulative prospect theory. Given the structure of our decision problem, with only two prizes in each lottery and no losses, the differences are likely to be minor.

²⁴See Ballinger and Wilcox (1997) for a methodological review of this development, which builds on an earlier tradition started by Becker, DeGroot, and Marschak (1963). There are alternative error specifications, depending on the stage of the latent decision process that one assumes the error to occur. Hey (1995), Carbone (1998), and Loomes and Sugden (1998) reviewed the alternatives.

²⁵Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. For example, Williams (2000, p. 645) noted that it could arise from dental studies that “collect data on each tooth surface for each of several teeth from a set of patients” or “repeated measurements or recurrent events observed on the same person.” The procedures that allow for clustering allow heteroscedasticity between and within clusters, as well as autocorrelation within clusters. They are closely related to the “generalized estimating equations” approach to panel estimation in epidemiology (see Liang and Zeger (1986)), and generalize the “robust standard errors” approach popular in econometrics (see Rogers (1986)). Wooldridge (2003) reviewed some issues in the use of clustering for panel effects, in particular noting that significant inferential problems may arise with small numbers of panels.

3.2. Formal Specification Assumptions

The most popular functional form for the characterization of risk attitudes has been the CRRA specification, assuming a standard EUT decision process. We first specify the latent structural model implied by that specification. We then consider two extensions to allow for (i) prospect theory and probability weighting in the latent process, and (ii) the possibility that subjects might be making decisions with something other than the lottery prizes as arguments of their utility function.

The base case under EUT

Utility is defined by the standard function

$$(1) \quad U(m) = m^{1-r}/(1-r),$$

where r is the CRRA coefficient and m is the monetary value of the lottery prize. With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(m) = \ln(m)$. We estimate the CRRA parameter using maximum likelihood applied to the binary choices made by each subject. Probabilities for each outcome k_n , $p(k_n)$, are those induced by the experimenter, so expected utility is then simply the probability weighted utility of each outcome in each lottery. Given that there were two outcomes in each lottery, the EU for lottery i is

$$(2) \quad EU_i = \sum_n [p(k_n) \times U(k_n)]$$

for $n = 1, 2$. The EU for each lottery pair is calculated for a candidate estimate of r , and the difference

$$(3) \quad \nabla EU = EU_A - EU_B$$

is calculated, where EU_A is the left lottery in the display and EU_B is the right lottery.

The subject is assumed to make some structural error in the comparison of the EU of the two lotteries. Let this error ε be normally distributed with zero mean and standard deviation σ . Then, if $\Phi(\cdot)$ is the standard normal cumulative distribution function, it follows²⁶ that the likelihood is $\Phi((EU_A - EU_B)/\sigma)$ if A is chosen, and $1 - \Phi((EU_A - EU_B)/\sigma)$ if B is chosen. Thus σ is a structural

²⁶If the subject does not make mistakes, then lottery A is chosen if $EU_A - EU_B > 0$; otherwise B is chosen. If the subject makes measurement errors, denoted by ε , then the decision is made on the basis of the value of $EU_A - EU_B + \varepsilon$. That is, A is chosen if $EU_A - EU_B + \varepsilon > 0$; otherwise B is chosen. If ε is random, then the probability that A is chosen is $P(EU_A - EU_B + \varepsilon > 0) = P(\varepsilon > -(EU_A - EU_B))$. Because ε is assumed to be normally distributed with mean 0 and standard deviation σ , it follows that $Z = \varepsilon/\sigma$ is normally distributed with mean 0 and stan-

“noise parameter” used to allow some errors from the perspective of the deterministic EUT model. As $\sigma \rightarrow 0$ this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries, but as σ gets larger the choice essentially becomes random.

The likelihood of the risk aversion responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of r and σ given the preceding statistical specification and the observed choices. The log-likelihood is

$$(4) \quad \ln L^{\text{EUT}}(r, \sigma; y, X) = \sum_i [(\ln(\Phi(\nabla \text{EU})/\sigma)|y_i = 1) + (\ln(1 - \Phi(\nabla \text{EU})/\sigma)|y_i = 0)],$$

where $y_i = 1$ (0) denotes the choice of lottery A (B) in risk aversion task i and X is a vector of individual characteristics.

We allow each parameter to be a linear function of the characteristics of the subject or the task. This is the X vector referenced in the foregoing text. Our most important treatment variables are whether the task involved a coin (COIN) or whether it involved an ungraded coin and, hence, added background risk (BRISK), so the implicit control is the task using money. The estimates of *each* parameter in the aforementioned likelihood function entail estimation of the coefficients of a linear function of these characteristics. Thus, if we included these two characteristics, the estimate of r , \hat{r} , would actually be

$$(5) \quad \hat{r} = \hat{r}_0 + (\hat{r}_{\text{COIN}} \times \text{COIN}) + (\hat{r}_{\text{BRISK}} \times \text{BRISK}),$$

where \hat{r}_0 is the estimate of the constant, normalized on the money treatment. We also include other task and subject characteristics in our estimation that are subsequently explained. If we collapse this specification by dropping all characteristics, then we would simply be estimating the constant terms for each of r and σ .

Prospect theory

For the “value function” of PT we use the same CRRA functional form specified for EUT:

$$(6) \quad U(x) = (x^{1-\alpha})/(1 - \alpha).$$

dard deviation 1: Z has a unit normal distribution. Hence the probability that A is chosen is $P(\varepsilon > -(\text{EU}_A - \text{EU}_B)) = P(\varepsilon/\sigma > -(\text{EU}_A - \text{EU}_B)/\sigma)$. It follows that the probability that A is chosen is $1 - \Phi(-(\text{EU}_A - \text{EU}_B)/\sigma) = \Phi((\text{EU}_A - \text{EU}_B)/\sigma)$, because the distribution is symmetrical about 0. Hence the probability that B is chosen is given by $\Phi(-(\text{EU}_A - \text{EU}_B)/\sigma) = 1 - \Phi((\text{EU}_A - \text{EU}_B)/\sigma)$. We implicitly assume that the probability that the subject is indifferent is zero.

We do not have any losses in the lotteries considered here, so we drop the part of the utility function in PT that is defined over losses.²⁷ We use the probability weighting function proposed by Tversky and Kahneman (1992), which implies decision weights

$$(7) \quad w(p) = p^\gamma / [p^\gamma + (1 - p)^\gamma]^{1/\gamma}$$

for $0 < p < 1$. The parameter γ determines the curvature of the function.²⁸ Assuming that PT is the true model, prospective utility (PU) is defined in much the same manner as when EUT is assumed to be the true model. The PT utility function (6) is used instead of the EUT utility function (1) and $w(p)$ is used instead of p , but the steps are otherwise identical. The difference in prospective utilities is defined similarly as

$$(8) \quad \nabla\text{PU} = \text{PU}_A - \text{PU}_B$$

and the same Fechner noise specification is assumed to apply, with error term ϕ instead of σ . Thus the likelihood, conditional on the PT model being true, depends on the estimates of α , γ , and ϕ given the foregoing specification and the observed choices. The log-likelihood is

$$(9) \quad \ln L^{\text{PT}}(\alpha, \gamma, \phi; y, X) = \sum_i [(\ln(\Phi(\nabla\text{PU}/\phi))|y_i = 1) + (\ln(1 - \Phi(\nabla\text{PU}/\phi))|y_i = 0)].$$

The parameters α , γ , and ϕ can again be estimated, in principle, as linear functions of the vector X .

²⁷To be accurate, we did not frame any of the prizes as negative numbers, which would make them losses relative to a reference point of zero. If the subject arrived at the experiment expecting some average level of earnings per task, prizes that implied a lower earning might be *subjectively* framed as a loss. We assume that this is not the case and that all subjects implicitly used zero as their reference point. In fact, when we imposed nonnegative reference points between \$0 and \$4 in \$0.10 increments, and extended the statistical model to estimate the degree of loss aversion relative to any positive reference point, the likelihood of the model is maximized at \$0. Thus we believe it appropriate to assume a reference point of zero *in this field setting*. Andersen, Harrison, and Rutström (2006) examined this issue in greater depth using traditional laboratory experiments. Unlike our field experiments, subjects in that setting have some experience with earning substantial positive rewards for much longer tasks. They elicited subjects' beliefs about expected earnings for the *session*, which averaged roughly \$30. However, when they estimated an endogenous "homegrown reference point" with the full structure of a prospect theory model that also allows for loss aversion, the estimate was only \$1.11 per lottery choice, with a 95% confidence interval between \$0.43 and \$1.80.

²⁸The normal assumption, backed by a substantial amount of evidence reviewed by Gonzalez and Wu (1999), is that $0 < \gamma < 1$. This gives the weighting function an "inverse S shape," characterized by a concave section that signifies the overweighting of small probabilities up to a crossover point where $w(p) = p$, beyond which there is then a convex section that signifies underweighting. If $\gamma > 1$ the function takes the less conventional "S shape," with convexity for smaller probabilities and concavity for larger probabilities.

The arguments of utility

There is a debate about the arguments of the utility function. Our view is that one is free to define utility over virtually anything, such as final states of nature, lottery prizes, income, or wealth at some point in time. This view is consistent with the analyses specified previously in which the argument is assumed to be the lottery prize offered to subjects. However, it is also possible to resolve this issue empirically by extending the earlier analysis to estimate the argument of the utility function.²⁹ Specifically, modify (1) to be

$$(1') \quad U(\omega + m) = (\omega + m)^{1-r} / (1 - r)$$

and include ω as a parameter to be estimated in the likelihood function (4). If ω is close to zero, then there is empirical support for the view that subjects behave as if maximizing a utility function defined over lottery prizes.

3.3. *Estimates*

Table II contains maximum likelihood estimates of the EUT specification applied to our data. We include several characteristics in binary form to account for some of the heterogeneity of responses in risk attitudes. Each is self-explanatory from the description in Table II. One likely source of 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 spouses. We therefore allow for this as a control in the estimation of the noise parameter σ . In addition, we also control for potentially confounding effects of the skewness treatments and differences in individual characteristics across treatment samples.

The most important result from Table II is that there is a significant effect from the ungraded coins treatment. The constant term shows the estimated CRRA under the money treatment: an estimate of 0.951 indicates that individuals are moderately risk averse. There is no statistically significant difference with graded coins: the CRRA coefficient is 0.160 lower than the treatment with money, but the effect is statistically insignificant, with a p -value of 0.794. However, the effect of ungraded coins and background risk is to increase elicited risk aversion dramatically. The coefficient estimate of 3.974 in Table II is clearly statistically significant. To our surprise, given our priors from related markets, there appears to be no interaction between this striking effect of background risk and whether the subject was a dealer.

The economic significance of these results can be illustrated by considering a simple lottery in which outcomes of \$10, \$20, or \$30 occur with equal probability. Hence the expected value of the lottery is \$20. If we assume that utility is defined solely in terms of the lottery prizes, CRRA values of 0.88, 0.76, and

²⁹Heinemann (2007) and Andersen, Harrison, and Rutström (2006) developed this approach.

TABLE II
MAXIMUM LIKELIHOOD ESTIMATION RESULTS FOR EUT SPECIFICATION

Parameter	Variable	Estimate	Std. Error	<i>p</i> -Value	95% Confidence Interval	
<i>r</i>	Constant	0.951	0.444	0.032	0.080	1.822
	Coins as final outcomes	-0.160	0.610	0.794	-1.356	1.036
	Ungraded coins as final outcomes	3.974	0.744	0.000	2.517	5.431
	Frame to skew RA lower	0.756	0.400	0.059	-0.027	1.539
	Frame to skew RA higher	0.142	0.299	0.634	-0.443	0.728
	Female	-1.259	0.712	0.077	-2.654	0.136
	College education or higher	0.044	0.209	0.832	-0.366	0.455
	Single and never married	0.838	0.379	0.027	0.094	1.582
	Ever owned Morgan silver dollars	0.032	0.573	0.956	-1.092	1.156
	Coin dealer	-0.984	0.597	0.099	-2.153	0.185
	Dealer × coins as final outcome	0.394	0.473	0.405	-0.533	1.320
	Affiliated with a grading company	-0.124	0.284	0.661	-0.680	0.432
<i>σ</i>	Constant	0.408	0.734	0.578	-1.030	1.847
	Female	12.534	38.278	0.743	-62.490	87.558

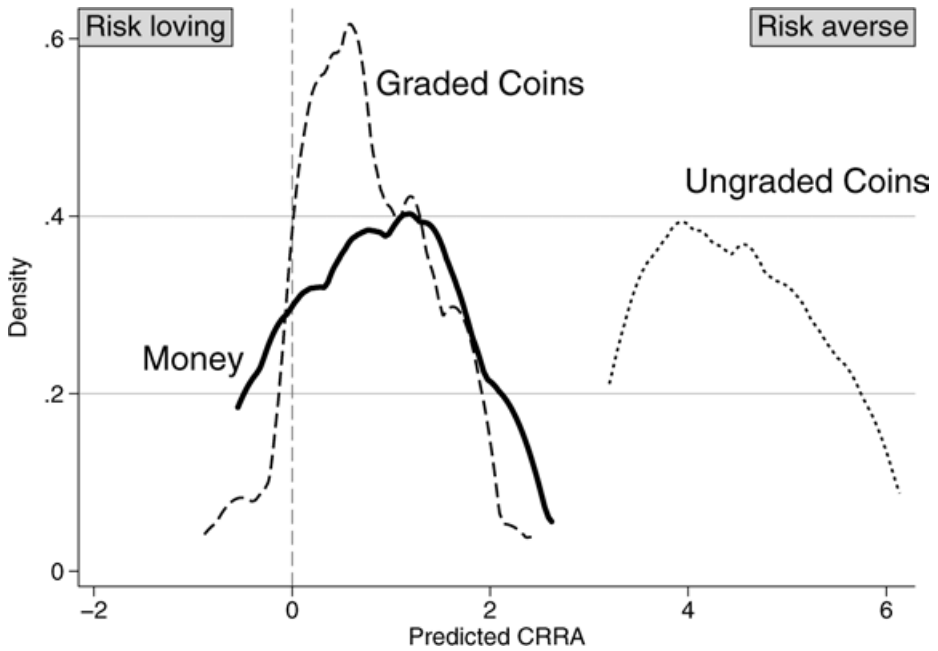


FIGURE 4.—Distribution of risk attitudes by treatment: the kernel density of predicted CRRA an using Epanichnikov kernel.

4.78 imply that the certainty equivalent would be \$18.40, \$18.62, and \$13.08, or 92%, 93%, and 65% of the expected value.

Another way to study the effects of different types of coins is to use the maximum likelihood estimates to predict CRRA for each subject and to show the density of predicted CRRA by treatment. Figure 4 shows these estimated densities, confirming the clear finding from Table II of a marked increase in risk aversion when we allow for background risk. With this procedure, we predict that the CRRA is 0.88 in the money treatment, 0.76 in the graded coins treatment, and 4.78 in the ungraded coins treatment.

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 other characteristics have some effects on elicited risk attitudes. Women exhibit less risk aversion in this setting, and those who have not enjoyed the comfort of marriage exhibit more risk aversion (we choose not to speculate on causality of this effect). Coin dealers are less risk averse in such tasks, as

one might expect from their self-selection into a part-time or full-time occupation that thrives on careful evaluation of risky prospects such as these. The skewness frames have a mixed, and unexpected effect. The frame to skew risk aversion (RA) higher has no appreciable effect, but the frame to skew RA lower is associated with higher risk attitudes. We conjecture that the frame may have actually triggered a “frame reaction,” in the sense that the prospects may have appeared to be skewed to the subjects, leading them to overreact to avoid responding to the frame.³⁰

The *prospect theory specification mitigates the quantitative effects of the treatments, but does not change the main qualitative conclusions*. For parsimony, we only summarize the results here. The probability weighting parameter γ is estimated to be 0.83, which is significantly less than 1 and consistent with previous estimates from laboratory experiments. The effect of the graded coins treatment on risk aversion is again statistically insignificant, after controlling for the same characteristics used in Table II. The CRRA parameter α increases by 0.65 with graded coins, but has p -value of 0.86. On the other hand, α increases by 0.59 with ungraded coins, and has a standard error of only 0.18 and a p -value of 0.001. The 95% confidence interval on this effect from background risk is between +0.24 and +0.94, so the treatment effect magnitude is much smaller than under EUT, but we find qualitatively similar insights.

Finally, there is *striking evidence that subjects have lottery prizes as arguments of their utility function under EUT*. The parameter ω is estimated to be only \$0.037 on average, and to have a 95% confidence interval across subjects and treatments between \$0.023 and \$0.052. This variation includes sampling error as well as heterogeneity across subjects and tasks, and is very small. Again, this is an “agnostic empirical argument” for this specification of the utility function, quite distinct from the theoretical issues discussed by Rabin (2000), Cox and Sadiraj (2006), and Rubinstein (2002). The estimates of the effects of our primary treatments on risk attitudes are virtually identical to the results in Table II, where we assumed $\omega = 0$ on an a priori basis.

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

³⁰In these models we also explored empirically whether agents responded differently to the treatments. We found little heterogeneous treatments effects across naïve and sophisticated agents in our sample, so we suppress these results.

for the effects of background risk on attitudes to risk. Behavior tended to be slightly risk averse when artificial monetary prizes were used or when there was minimal uncertainty in the natural nonmonetary outcome. However, subjects drawn from the same population were quite risk averse when their attitudes were elicited using the natural nonmonetary outcome that had some background risk.

The practical implications for policy are twofold. First, estimates of risk aversion elicited using monetary outcomes may be reliable for nonmonetary outcomes that do not involve a considerable amount of background risk. However, they should not be applied directly to nonmonetary 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.³¹ 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.

These results highlight the suggestion that field experiments represent a useful bridge between the lab and the field, providing a useful middle ground between the tight controls of the laboratory and the vagaries of completely uncontrolled field data. Such a bridge permits a discussion of the generalizability and overall applicability of experimental results from the lab. Indeed, in transferring the insights gained in the laboratory with student subjects to the field, a necessary first step is to explore how market professionals behave in strategically similar situations. In this spirit, our treatments begin by focusing on the representativeness of the sampled population to lend insights into heterogeneity across subject pools. Much like student data, we find substantial heterogeneity in our data.

A related issue concerns the representativeness of the environment. For example, before we can begin to make reasonable arguments that behavior observed in the lab is a good indicator of behavior in the field, we must explore whether the other dimensions of the laboratory environment might cause differences in behavior, including the abstract task, the stakes, the good, and the institution. In this regard, we find that the nature of the environment is critical in considering the general applicability of results from the lab. We find these insights a necessary first step in the discovery process, and we hope that future efforts will explore more fully other potentially important dimensions of the controlled laboratory experiment.

³¹We 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.

Dept. of Economics, College of Business Administration, University of Central Florida, Orlando, FL 32816, U.S.A.; gharrison@research.bus.ucf.edu,

Dept. of Economics, University of Chicago, 59th Street, Chicago, IL 60422, U.S.A.; jlist@uchicago.edu,

and

Dept. of Agricultural and Resource Economics, University of Maryland, College Park, MD 20742, U.S.A.; ctowe@arec.umd.edu.

Manuscript received January, 2005; final revision received July, 2006.

REFERENCES

- ANDERSEN, S., G. W. HARRISON, AND E. E. RUTSTRÖM (2006): "Dynamic Choice Behavior: Asset Integration and Natural Reference Points," Working Paper 06-07, Department of Economics, College of Business Administration, University of Central Florida. [451,452]
- BALLINGER, T. P., AND N. T. WILCOX (1997): "Decisions, Error and Heterogeneity," *Economic Journal*, 107, 1090–1105. [448]
- BECKER, G., M. H. DEGROOT, AND J. MARSCHAK (1963): "Stochastic Models of Choice Behavior," *Behavioral Science*, 8, 41–55. [448]
- (1964): "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 226–232. [437]
- BINSWANGER, H. P. (1980): "Attitudes Toward Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, 395–407. [437]
- (1981): "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal*, 91, 867–890. [437]
- CARBONE, E. (1998): "Investigation of Stochastic Preference Theory Using Experimental Data," *Economics Letters*, 57, 305–312. [448]
- COX, J. C., AND V. SADIRAJ (2006): "Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory," *Games & Economic Behavior*, 56, 45–60. [434,455]
- CUBITT, R., C. STARMER, AND R. SUGDEN (1998): "On the Validity of the Random Lottery Incentive System," *Experimental Economics*, 1, 115–131. [439]
- EECKHOUDT, L., C. GOLLIER, AND H. SCHLESINGER (1996): "Changes in Background Risk and Risk-Taking Behavior," *Econometrica*, 64, 683–689. [435]
- GOLLIER, C., AND J. W. PRATT (1996): "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica*, 64, 1109–1123. [435]
- GONZALEZ, R., AND G. WU (1999): "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 129–166. [451]
- HARRISON, G. W. (1986): "An Experimental Test for Risk Aversion," *Economics Letters*, 21, 7–11. [437]
- HARRISON, G. W., M. I. LAU, E. E. RUTSTRÖM, AND M. B. SULLIVAN (2005): "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in *Field Experiments in Economics, Research in Experimental Economics*, Vol. 10, ed. by J. Carpenter, G. W. Harrison, and J. A. List. Greenwich CT: JAI Press, 125–218. [439,445]
- HARRISON, G. W., AND J. A. LIST (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1013–1059. [434,439]
- HEINEMANN, F. (2007): "Measuring Risk Aversion and the Wealth Effect," in *Risk Aversion in Experiments, Research in Experimental Economics*, Vol. 12, ed. by J. C. Cox and G. W. Harrison. Greenwich, CT: JAI Press, forthcoming. [452]
- HEY, J. D. (1995): "Experimental Investigations of Errors in Decision Making under Risk," *European Economic Review*, 39, 633–640. [448]

- HEY, J. D., AND C. ORME (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326. [448]
- HOLT, C. A., AND S. K. LAURY (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655. [437,444]
- KACHELMEIER, S. J., AND M. SHEHATA (1992): "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82, 1120–1141. [437]
- KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291. [448]
- KIMBALL, M. S. (1993): "Standard Risk Aversion," *Econometrica*, 61, 589–611. [435]
- LAURY, S. K. (2005): "Pay One or Pay All: Random Selection of One Choice for Payment," Research Paper Series 06-13, Andrew Young School of Policy Studies, <http://ssrn.com/abstract=894271>. [439]
- LIANG, K.-Y., AND S. L. ZEGER (1986): "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22. [448]
- LOOMES, G., AND R. SUGDEN (1998): "Testing Alternative Stochastic Specifications for Risky Choice," *Economica*, 65, 581–598. [448]
- LUSK, J. L., AND K. H. COBLE (2005): "Risk Perceptions, Risk Preference, and Acceptance of Risky Food," *American Journal of Agricultural Economics*, 87, 393–404. [435]
- (2007): "Risk Aversion in the Presence of Background Risk: Evidence from the Lab," in *Risk Aversion in Experiments, Research in Experimental Economics*, Vol. 12, ed. by J. C. Cox and G. W. Harrison. Greenwich, CT: JAI Press, forthcoming. [436,437]
- PRATT, J. W. (1988): "Aversion to One Risk in the Presence of Others," *Journal of Risk and Uncertainty*, 1, 395–413. [435]
- PRATT, J. W., AND R. J. ZECKHAUSER (1987): "Proper Risk Aversion," *Econometrica*, 55, 143–154. [435]
- QUIGGIN, J. (2003): "Background Risk in Generalized Expected Utility Theory," *Economic Theory*, 22, 607–611. [435]
- RABIN, M. (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292. [434,455]
- ROGERS, W. H. (1986): "Regression Standard Errors in Clustered Samples," *Stata Technical Bulletin*, 13, 19–23. [448]
- RUBINSTEIN, A. (2002): "Comments on the Risk and Time Preferences in Economics," Working Paper 06-02, Foerder Institute for Economic Research, Eitan Berglas School of Economics, Tel Aviv University, <http://www.econ.tau.ac.il/research/abstract.asp?id=62002>. [434,455]
- SAFRA, Z., AND U. SEGAL (1998): "Constant Risk Aversion," *Journal of Economic Theory*, 83, 19–42. [435]
- STARMER, C., AND R. SUGDEN (1991): "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review*, 81, 971–978. [439]
- TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323. [448,451]
- U.S. ENVIRONMENTAL PROTECTION AGENCY (1997): "The Benefits and Costs of the Clean Air Act: 1970 to 1990," Office of Air and Radiation, US EPA, Washington, DC. [436]
- WILLIAMS, R. L. (2000): "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics*, 56, 645–646. [448]
- WOOLDRIDGE, J. (2003): "Cluster-Sample Methods in Applied Econometrics," *American Economic Review (Papers & Proceedings)*, 93, 133–138. [448]