Risk Attitudes and Evaluation Periods in the Laboratory: A Reconsideration

by

Glenn W. Harrison [†]

August 2005

Abstract. Laboratory tests of the hypothesis of myopic loss aversion do not refute expected utility theory, unless one defines expected utility theory only in terms of needlessly restrictive specifications. The data presented in these tests are consistent with expected utility theory as well as myopic loss aversion. However, these data also demonstrate that the parameters of popular specifications of the myopic loss aversion model vary with changes in the evaluation period, so one cannot assume them to be invariant to changes in the evaluation period. This invariance was a maintained assumption in the use of myopic loss aversion to explain the equity premium puzzle. Finally, these experiments present a challenge to both expected utility theory specifications and alternatives to be explicit about the appropriate evaluation horizon of decision makers. Thus a seminal contribution of the myopic loss aversion literature, and these experimental designs, is to force mainstream economists to pay attention to an issue that they have neglected *within* their own framework.

[†] Department of Economics, College of Business Administration, University of Central Florida, USA. E-mail contact: GHARRISON@BUS.UCF.EDU. I am grateful to Uri Gneezy, Jan Potters, Michael Haigh and John List for generously providing access to their raw data, and to Morten Lau and Elisabet Rutström for helpful comments. Supporting data and statistical code are stored in the *ExLab* Digital Archive at http://exlab.bus.ucf.edu.

Several recent studies propose experimental tests that purport to directly test expected utility theory (EUT) against the alternative hypothesis of "myopic loss aversion" (MLA). Gneezy and Potters [1997] and Haigh and List [2005] use simple experiments in which many potential confounds are removed.[1] Unfortunately, those experiments only test a very special case of EUT against the alternative hypothesis. That special case fails, rather dramatically. But it is easy to come up with utility functions that are consistent with EUT and that can explain the observed data without relying on MLA. For example, any utility function with decreasing relative risk aversion (RRA) and that exhibits risk aversion for low levels of income will suffice at a qualitative level. The empirical outcomes observed at the individual level can then be explained by simply fitting specific parameters to this utility function. We demonstrate this intuitively in section 1, and more formally in section 2.

Section 3 considers some unsettling implications of these experiments for MLA: that the key "loss aversion" parameters of the standard MLA model vary dramatically according to the exogenously imposed evaluation period. Thus the behaviorist explanation is hoisted on the same petard it alleged applied to the EUT explanation.

However, although it is useful and trivial to come up with a standard EUT story that accounts for the data, and even fun to find an anomaly for the behaviorists to ponder, these experiments force one to examine a much deeper question than "can EUT explain the data?" That question is whether utility is best defined over the *outcome* that the subject faces or over the *decision* that the subject is asked to make. Depending on how the *subjects interpret the experimental task*, these frames could differ in this experimental task. Section 4 considers these open issues, arguing that these experiments present a challenge to both expected utility theory specifications and alternatives to be explicit about the appropriate evaluation horizon of decision makers. Thus a seminal

[1] Additional experimental tests include Thaler, Tversky, Kahneman and Schwartz [1997] and Gneezy, Kapteyn and Potters [2003]. These provide results that are qualitatively identical, but harder to evaluate. Thaler et al. [1997] did not provide subjects with precise knowledge of the probabilities involved in the lotteries, but allowed them to infer that over time; hence behavior could have been driven by mistakes in the subjective inference of probabilities rather than MLA. Gneezy et al. [2003] embed the task in an asset market, which may have influenced individual behavior in other ways than predicted by EUT or MLA. These influences are of interest, since markets are the institution in which most stocks and bonds are traded, but from the perspective of wanting the cleanest possible test of competing theories those extra influences are just a confound.

contribution of the myopic loss aversion literature, and these experimental designs, is to force mainstream economists to pay attention to an issue that they have neglected *within* their own framework.

1. A Simple EUT Explanation of the Data

A. The Experimental Task

The experimental task was very simple. Each subject in the baseline treatment made 9 decisions over a fixed stake. In Gneezy and Potters [1997] (GP) this stake was 2 Dutch Guilders, which we will call $2.00 for pedagogic ease. In each round they could choose a fraction of the stake to bet. If they chose to bet nothing then they received $2.00 in that round for certain. If they bet $$x$ then they faced a ⅔ chance of losing $$x$ and a ⅓ chance of winning 2.5x$. These earnings were on top of the initial stake of $2.00. Thus the subject literally ended up with ($2.00 - $$x$) with probability ⅔ and ($2.00 + 2.5x$) with probability ⅓. Since $x could not exceed $2.00, by design, the subject actually faced no losses for the round as a whole. Of course, if one ignores the $2.00 stake the subject did face a loss. In the baseline condition the subject chose a bet in each round, the random outcome was realized, their earnings in that round tabulated, and then the next round decision was made.

In the alternative treatment the subject made 3 decisions instead of 9. The first decision was a single amount to bet in each of rounds 1 through 3, the second decision was a single amount to bet in each of rounds 4 through 6, and the third decision was a single amount to bet in each of rounds 7 through 9. Thus the subject made *one decision or choice* for each of the outcomes in rounds 1, 2 and 3. To state it equivalently, since this is critical to follow, one decision was simply applied three times: it is not the case that the subject made three separate decisions at round 1 that were applied in rounds 1, 2 and 3 respectively. The subject could not say "bet x%, y% and z% in rounds 1, 2 and 3," but could only instead say "bet x%," meaning that x% would be bet for the subject in each of round 1, 2 and 3. In all other respects the experimental task was the same: the only thing that varied was the horizon over which the choices were made. This is referred to as the Low Frequency treatment (L), and the baseline is referred to as the High Frequency treatment (H).

B. The Observed Outcomes

The raw data in the two sets of experiments are presented in Figures 1 and 2, which show the distribution of percentage bets. The general qualitative outcome is for subjects to bet more in the L treatment than in the H treatment. Gneezy and Potters [1997; Table I, p.639] report that 50.5% was bet in their treatment H and 67.4% in their treatment L over all 9 rounds. They conducted their experiments with 83 Dutch students, split roughly evenly across the two treatments in a between-subjects design. Haigh and List [2005] (HL) report virtually the same outcomes: for their sample of 64 American college students, the fractions were 50.9% and 62.5%, respectively, and for their sample of 54 current and former traders from the Chicago Board of Trade the fractions were 45% and 75%, respectively.[2] Using unconditional non-parametric tests or panel Tobit models, these differences are statistically significant at standard levels.[3] Thus it appears that samples of subjects drawn from the same population behave as if more risk averse in treatment H compared to treatment L, and that the average subject is risk averse. The latter inference follows from the fact that a risk-neutral subject, according to EUT, would bet 100% of the stake.

Figures 1 and 2 also alert us to one stochastic feature of these data that will play a role later: that there is a substantial spike at the 100% bet level. From an EUT perspective, this corresponds to subjects that are risk neutral or risk loving.

If we just consider "interior bets" then the same qualitative results obtain. In GP the Low

---

[2] The stakes in the experiments of Gneezy and Potters [1997] were actually 2 Dutch guilders, which converted at the time of the experiment to roughly $1.20. Haigh and List [2005] used a stake of $1.00 for their students, to be comparable to the earlier stake. They quadrupled the stakes to $4.00 for the traders, on the grounds that it would be more salient for them. Of course, this change in monetary stake size adds a potential confound to the comparability of results across students and traders, but one that has no obvious resolution without an elaborate investigation into the purchasing power of a dollar to students and traders.

[3] Gneezy and Potters generously provided their individual data, and we used the same statistical model as Haigh and List [2005; Table II, specification 2, p. 530] on their data. Haigh and List also generously provided their individual data, and we replicated their statistical conclusions.

frequency treatment generates an average 42.1% bet compared to an average 33.9% bet in the High frequency treatment. In HL the students (traders) bet an averages of 37.7% (25.3%) and 51.4% (59.3%) in each treatment.

C. An Important Aside: EUT and the Arguments of Utility Functions

MLA justifiably captured the attention of mainstream economists when Benartzi and Thaler [1995] (BT) used it to offer an explanation of the "equity premium puzzle." However, somewhere on the way to BT, a funny thing happened to the definition of a utility function within EUT. All of a sudden, EUT was to be *exclusively* defined in terms of utility as a function of terminal wealth. Thus BT (p.74) could write that in prospect theory "...utility is defined over gains and losses relative to some neutral reference point, such as the status quo, as opposed to wealth as in expected utility theory." There is more to the BT explanation of the equity premium than just this interpretation of the arguments of the utility function, as noted in section 4, but this issue needs to be clarified independently since it is extraordinarily pervasive.[4]

Whether or not one models utility as a function of terminal wealth (EUTw) or income (EUTi) depends on the setting. Both specifications have been popular. The EUTw specification was widely employed in the seminal papers defining risk aversion and the application of those concepts to portfolio choice. It has also been employed exclusively in the finance literature, such as the debates over the equity premium reviewed by Kocherlakota [1996]. On the other hand, the EUTi specification has been widely employed by auction theorists and experimental economists testing EUT. The central point is that either is valid, since the axioms of EUT are silent about the objects of the utility function.

One is tempted to think that this point is well-known since Markowitz [1952] and Samuelson

[4] Remarkably, it is pervasive even among economists that would view themselves as working solely within standard EUT. For example, in a major textbook devoted to *The Economics of Risk and Time*, Gollier [2001; p.9] writes, "In this book, we will consider the outcome [i.e., the argument of the utility function] to be a monetary wealth." There are no references to the literature cited below, in which alternative interpretations are allowed.

[1952; ¶13, p.676], but that may just be a hindsight bias. Cox and Sadiraj [2005] and Rubinstein [2002] make these points quite clearly, in the context of controversies over the validity of EUT generated by Rabin [2000] and Rabin and Thaler [2001].

Cox and Sadiraj [2005] go further to propose a generalization of EUTw and EUTi that allows initial wealth to be an argument of the utility function along with income (as long as initial wealth is not simply added to income, which would be EUTw). They also note that "loss aversion," the alternative favored by Rabin [2000] and Rabin and Thaler [2001] as a descriptive model of low-stakes risk aversion, is perfectly consistent with EUTi.

Rubinstein [2002] draws the important connection between adopting an EUTi assumption and the question of temporal consistency of preferences, since the income that one receives in today's experiment must be "integrated" in some consistent way with the income received in the past (viz., wealth prior to the experiment). This suggests links back to the older literature on the "asset integration hypothesis," reviewed by Quizon, Binswanger and Machina [1984]. In other words, just because one adopts an EUTi characterization and thereby avoid the problems posed by Rabin [2000], one is not free to make any arbitrary assumptions about behavior over time. The laboratory evidence on this matter has it's own controversies: see Frederick, Loewenstein and O'Donoghue [2002] and Harrison and Lau [2005].

It is worth pausing to trace the intellectual history of this confusion in behavioral economics. Kahneman and Tversky [1979; p.264] explicitly state the "asset integration hypothesis" that is perhaps implicit in the view that EUT is defined solely over terminal wealth. They note (p.276) that Markowitz [1952] "... was the first to propose that utility be defined on gains and losses rather than on final asset positions, an assumption which has been implicitly adopted in most experimental measurements of utility." One prominent experimental study that they cite is by Mosteller and Nogee [1951]. In their theoretical statement of the assumptions underlying the construction of utility curves *defined explicitly over gains*, Mosteller and Nogee [1951; p.373] clearly note the role played by Friedman, Savage and Samuelson in earlier drafts. Furthermore, they note (fn.7, p.372) that "Plans for this experiment grew directly out of discussions with Friedman and Savage at the time they were writing their paper," referring of course to Friedman and Savage [1948]. One might have assumed that Friedman or Savage, not known for their willingness to fudge such matters or hide scholarly disagreements, might have raised such a fundamental issue at the time – if, in fact, it had been an issue at all. Thus the use of EUTi over 50 years ago does not seem to have an unusual or limited pedigree, notwithstanding the *ad hominem* nature of the case.

In any event, we proceed here as if it is valid to use EUT defined over changes in wealth, recognizing that this is not the exclusive definition of EUT.

D. An EUT-Consistent Explanation

Returning to the experiments of Gneezy and Potters [1997] and Haigh and List [2005], the key point is to view subjects as having a utility function that is defined over prize income that reflects the stakes that *choices* are being made over. The High frequency subjects can be viewed as making a series of 9 choices over stakes defined, for each choice, by a vector y which takes on a range of integer values between \$0 and \$7. The subject could get \$0 if they bet 100% of the stake and lost it; or they could get as much as \$7 if they bet 100% of the stake and won 2.5 × \$2.00.

The Low frequency subjects, on the other hand, made 3 choices over stakes defined by the possible combinations of gains and losses over 3 random draws. Thus they could end up with three losses, 2 losses and 1 gain, 1 loss and 2 gains, or 3 gains. The probabilities for each outcome, irrespective of order, are 0.30, 0.44, 0.22 and 0.04, respectively. The monetary outcome in each case depends on the fraction of the stake that the subject chose to bet.

Table 1 spells out the arithmetic for different bets. For simplicity we evaluate the possible choices in increments of 10 cents, but of course the choices could be in pennies.[5] The second column shows the bet as a percent of the stake of \$2.00. Columns 3 though 7 show the components

[5] In fact, subjects tended to pick in round percentages. In the Low frequency treatment 76% of the choices were for 0%, 25%, 50% or 100% bets, and in the High frequency treatment 81% of the choices were for the 25%, 50% or 100% bets.

of the lottery facing the subject in the High frequency treatment for each possible bet, and columns 8 through 16 show the same components for the subject in Low frequency treatment. Consider, for example, a bet of 10 cents, which is 5% of the stake. If the subject is in the High treatment and loses, they earn 190 (=200-10) cents in that period; this occurs with probability ⅔. If the subject is in the High treatment and wins, they earn 225 (= 200 + 10×2.5 = 200 + 25) cents; this occurs with probability ⅓. In the corresponding entry for the subject in the Low treatment, the value of prizes is calculated similarly, but for three random draws. Thus in the LLL outcome, the subject earns 570 (= 200-10 + 200-10 + 200-10) cents.

From Table 1 we see instantly that a risk neutral (RN) subject that obeyed EUT would bet 100% of the pie in both treatments and thereby maximize expected value. It can also be inferred that a moderately risk averse subject would bet some fraction of the pie in each treatment, less than 100%, and that a risk loving subject would always bet 100% of the pie.

The outcomes of the lotteries being evaluated by subjects in the High and Low treatments differ significantly. Consider the 50% bet, in the middle of Table 1. For subjects in the High treatment the two final outcomes from each choice are 100 and 450, occurring with the probabilities shown there. For subjects in the Low Treatment there are four final outcomes from each choice: 300, 650, 1000 and 1350. Thus the monetary rewards from the same percentage choice differ significantly. So, to explain why subjects in the Low treatment are more risk averse than subjects in the Low treatment, it suffices at a qualitative level to find some utility function that has moderate amounts of risk aversion for "low" income levels and smaller amounts of risk aversion for "higher" income levels.

Although less obvious than the RN prediction, any subject exhibiting Constant Relative Risk Aversion (CRRA) would choose *the same bet fraction in each row*. The more risk averse they were, the smaller would be the bet, but it would be the same bet in each of the High and Low treatments. This result is important since *every* statement of "the EUT null hypothesis" in the MLA literature that we can find uses RN or CRRA specifications for the utility function.[6] Thus it is easy to see why evidence of a difference between the bet fractions in the High and Low treatments is viewed as a rejection of EUT.

Of course, this does not test EUT at all. It only tests a very special case of EUT, where the specific functional form seems to have been chosen to perform poorly.[7] It is easy to propose more flexible utility function than CRRA. There are many such functions, but one of the most popular in recent work that is fully consistent with EUT has been the Expo-Power (EP) utility function proposed by Saha [1993]. Following Holt and Laury [2002], the EP function is defined as

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

where α and r are parameters to be assumed or estimated. 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 $r \rightarrow 0$). At a qualitative level, if $r>0$ and $\alpha<0$ one can immediately rationalize the qualitative data in these experiments: RRA $= r + \alpha(1-r)y^{1-r} \rightarrow$ r as $y \rightarrow 0$, and then one has declining RRA with higher prize incomes since $\alpha<0$.

2. Statistical Analysis

The qualitative insight that one can explain these data with a simple EUT specification can be formalized by estimating the parameters of a model that account for the data.[8] Such an exercise also helps explain some differences between the traders and students in Haigh and List [2005].

As noted earlier, Figures 1 and 2 alert us to the fact that the behavioral process generating data at the 100% bet level may be different than the process generating data at the "interior" solutions. From a statistical perspective, this is just a recognition that a model that tries to explain the interior modes of these data, and why they vary between the High and Low treatments, might

[6] For example, Kahneman and Lovallo [1993; p. 20], Benartzi and Thaler [1995; p.79], Gneezy and Potters [1997; p.632], Thaler, Tversky, Kahneman and Schwartz [1997; p.650], Gneezy, Kapteyn and Potters [2003; p. 822] and Haigh and List [2005; p.525].

[7] In other words, there are settings in which a CRRA or even RN utility function might be appropriate for some theoretical, econometric or policy exercise. But this experiment is not obviously one of those settings.

[8] Supporting data and statistical code are stored in the *ExLab* Digital Archive at http://exlab.bus.ucf.edu. Uri Gneezy, Jan Potters, Michael Haigh and John List generously provided access to their raw data.

have a difficult time also accounting for the spike at 100%. One approach is just to ignore that spike, and see what estimates obtain. Another approach is to construct a model and likelihood function that accounts for these two processes.[9] We apply both approaches, although favoring the latter *a priori*.

The dependent variable is naturally characterized as the fraction of the stake bet, denoted π. Therefore the likelihood function is constructed using the specification developed by Papke and Wooldridge [1996] for fractional dependant variables. Specifically, the log-likelihood of observation i is defined as $l_i(\xi) = \pi_i \times \log[G(x_i, \xi)] + (1-\pi_i) \times \log[1-G(x_i, \xi)]$ for parameter vector ξ, a vector of explanatory variables x, and some convenient cumulative distribution function $G(\cdot)$. We use the cumulative Gamma distribution function $G(z) = \Gamma(a,z)$, where a is a parameter that can be estimated.[10] The index z_i is the expected utility of the bet chosen, conditional on some parameter estimates of ξ and some characteristics x_i for observation i.

The index z is constructed using information on the lottery for the actual bet, reflecting a more detailed version of the arithmetic underlying Table 1. Thus, for a particular fractional bet, the parameters of the task imply that the subject was facing a particular lottery. So one element of the x vector is whether or not the subject was in the High or Low treatment. Another element is the stake. And another element is the set of parameters of the experimental task defining the lottery outcomes (e.g., the probabilities of a loss or a gain, and the numbers defining how the bet is scaled to define the loss or the gain). Using this information, and candidate estimates of r and α for the EP utility function, the likelihood constructs the expected utility of the observed choice, and the maximum likelihood estimates find the parameters of the EP utility function that best explain the observed choices.

This approach can be applied directly to the data in Figures 1 and 2, recognizing that one

---

[9] Yet another approach would be to modify the experimental design and allow subjects to leverage their bets beyond 100% of their stake. There are some logistical problems running such experiments in a laboratory setting, although of course stock exchanges and futures markets allow such trades.

[10] The *a* parameter may be viewed as a counterpart in this specification of the noise parameter used by Holt and Laury [2002].

model must explain the multiple modes of these distributions. Alternatively, one can posit a natural two-step decision process, where the subject first decides if they are going to bet everything or not, and then if they decide not to, decides how much to bet (including 0%). This might correspond to one way that a risk averse or risk loving subject might process such tasks: first figure out what a RN decision-maker would do, since that is computationally easier, and then shade one's choice in the direction dictated by risk preferences. Since the matrix in Table 1 was not presented to subjects in such an explicit form, this would be *one* sensible heuristic to use.

Irrespective of the interpretation, this proposed decision process implies a statistical "hurdle" model. First the subject makes a binary choice to contribute 100% or less. Then the subject decides what fraction to contribute, conditional on contributing less than 100%. The first stage can be modeled using a standard probit specification, although it is the second stage that is really of greatest interest.

A key feature of these estimates is that they pool the data from High and Low treatments. The objective is to ascertain if one EUT-consistent model can explain the shift in the distributions between these treatments in Figures 1 and 2. Since each subject provided multiple observations there are corrections for the possible correlation of errors associated with a given subject.[11]

Table 2 reports the results of maximum likelihood estimation of these models. Panels A and B provide estimates for the individual responses from Gneezy and Potters [1997]. The estimates using all bets show some initial risk aversion at zero income levels ($r = 0.21$) and then some slight evidence of declining RRA as income rises ($\alpha = -0.019$). However, the evidence of declining RRA is not statistically significant, although the 95% confidence interval is skewed towards negative values.

[11] The use of clustering to allow for "panel effects" from unobserved individual effects is common in the statistical survey literature. Clustering 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] notes 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 for allowing for clustering allow heteroskedasticity 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 [1993]). Wooldridge [2003] reviews some issues in the use of clustering for panel effects, in particular noting that significant inferential problems may arise with small numbers of panels.

Much more telling evidence comes from the estimates in panel B, for the interior bets. Here we find striking evidence of the qualitative explanation presented earlier: initial risk aversion at zero income levels (r = 1.12) and sharply declining RRA as income rises (α = -0.57).

Panels C through F report estimates for the treatments of Haigh and List [2005], estimated separately for traders and students since that was their main treatment. With the exception of the estimates in panel E, for *all* bets by UMD students, these results again confirm the qualitative explanation proposed above. The results for traders are strikingly consistent with the observed pattern of choices. Therefore one must simply reject the conclusion of Haigh and List [2005; p.531] that their "findings suggest that expected utility theory may not model professional traders' behavior well, and this finding lends credence to behavioral economics and finance models, which are beginning to relax inherent assumptions used in standard financial economics." Whether MLA models the behavior of traders better than EUT is a separate matter, but EUT easily explains the data. In fact, these data are more consistent with the priors that motivated the Haigh and List [2005] study, illustrated by List [2003][2004], that students would be more likely to exhibit anomalies than field traders. Focusing on the *a priori* preferred estimates of panels D and F, our results show that students and traders simply have different preferences, each of which is consistent with EUT and these data. Of course, *de gustibus non est disputandum*, even for behaviorists.

3. Coals To Newcastle: An Anomaly for the Behaviorists

The reason that MLA is interesting is that BT use it to provide an intuitive explanation for the equity premium puzzle. Their empirical approach is to assume a particular numerical specification of MLA, and then solve for the "evaluation horizon"[12] of returns to stocks and equities that makes their expected utility[13] equivalent. They find that this horizon is roughly 12 months,

[12] BT (p.80) are clear that this *evaluation* horizon is not the same thing as a *planning* horizon: "A young investor, for example, might be saving for retirement 30 years off in the future, but nevertheless experience the utility associated with the gains and losses of his investment every quarter when he opens a letter from his mutual fund. In this case his [planning] horizon is 30 years but his evaluation period [evaluation horizon] is 3 months."

[13] They prefer the expression "prospective utility," but there is no confusion as long we are clear about which utility functions and probabilities are being used to calculate expected utility.

which strikes one as *a priori* plausible if one had to pick a single representative evaluation horizon for all investors.[14] Thus they assume a particular empirical version of MLA and further assume that these coefficients do not change as they counter-factually calculate the effects of alternative evaluation horizons:

> According to our theory, the equity premium is produced by a combination of loss aversion and frequent evaluation. Loss aversion plays the role of risk aversion in standard models, and can be considered a fact of life (or, perhaps, a fact of preferences). In contrast, the frequency of evaluations is a policy choice that presumably could be altered, at least in principle. Furthermore, as the charts (...) show, stocks become more attractive as the evaluation period increases.

So the parameters of the MLA specification are assumed invariant to evaluation horizon, as an essential premiss of the empirical methodology.

Thus the motivation for the experiments of GP and HL. As GP note, BT "... do not present direct (experimental) evidence for the presence of MLA. The evidence presented in [BT] is only circumstantial. [...] We have experimental subjects making a sequence of risky choices. To analyze the presence of MLA, we do not try to estimate the period over which subjects evaluate financial outcomes, but rather we try to manipulate this evaluation period." Hence the data from GP can be used to recover the MLA preferences that are consistent with the observed behavior, and the empirical premiss of BT evaluated.

Since behavioral economists are so enamored of anomalies, it may be useful to point out one or two in the MLA literature being considered here. The first anomaly is that the data from the experiments of GP demonstrate that *the MLA specification itself depends on the evaluation horizon*, which of course was varied by experimental design in their data. Hence one cannot assume that those parameters stay fixed as one calibrates the equity premium by varying the evaluation horizon. The second anomaly is that these data also *imply risk attitudes defined over the utility function that are qualitatively*

[14] Mankiw and Zeldes [1991] make the important observation that only 12% of Americans hold stocks worth more than $10,000, using a 1984 survey, so one really has to explain *their* indifference between holding bonds and stocks. Presumably, the remaining "corner-solution" individuals face some transactions costs to undertaking such investments. It would be an easy and important extension of the approach of BT to allow for such heterogeneity in the composition of stockholders and others.

the opposite of those customarily assumed.[15]

The MLA parameterization adopted by BT (p.79) is taken directly from Tversky and Kahneman [1992], both in terms of the functional forms and parameter values. They assume a power utility function defined separately over gains and losses: $U(x) = x^{\alpha}$ if $x \geq 0$, and $U(x) = -\lambda(-x)^{\beta}$ for $x < 0$. So α and β are the risk aversion parameters, and λ is the coefficient of loss aversion.[16] Tversky and Kahneman [1992; p.59] provide estimates that have been universally employed in applied work by behaviorists: $\alpha = \beta = 0.88$ and $\lambda=2.25$.

Using the data from GP one can estimate the parameters of this MLA model. For simplicity we assume no probability weighting, although that could be included. BT (p.83) and GP stress that it is the loss aversion parameter λ that drives the main prediction of MLA, rather than probability weighting or even risk aversion in the utility function. The likelihood function is again constructed using the specification developed by Papke and Wooldridge [1996] for fractional dependant variables. Since there are no data on personal characteristics in the GP data, the x vector refers solely to whether or not the decision was made in the Low frequency setting or the High frequency setting. Thus $\xi = (\alpha, \beta, \lambda)$, and each of those fundamental parameters is estimated as a linear function of binary dummies for the Low and High frequencies.[17]

Table 3 reports the maximum likelihood estimates obtained. The "good news" for MLA is that they provide strong evidence that the loss aversion parameter is greater than 1. The "bad news" for MLA is that they provide equally striking evidence that all of the parameters of the MLA

[15] These specifications have not been estimated for the data in HL, although one could do so.

[16] Köbberling and Wakker [2005; p. 128] argue against using this CRRA specification of the utility function, or at least constraining $\alpha = \beta$, so that a general index of loss aversion that they propose is well-defined. They suggest using a Constant Absolute Risk Aversion utility function instead. The popularity of the CRRA specification justifies studying it, and it should be an easy matter to replicate the analysis using alternative specifications. However, their concerns with the CRRA specification derive from their assumption (p.121) that the utility function should have well-defined directional derivatives at a zero income level, which is a sufficient condition for their index to be scale-invariant. It is not obvious why this assumption is as "plausible" as they claim.

[17] The constant term in this linear function is suppressed, since it would be perfectly correlated with the sum of these two binary variables. To be explicit, denote these dummy variables for the treatments as L and H, respectively. Then we actually estimate α_L, α_L, β_L, β_H, λ_L and λ_H, where $\alpha = \alpha_L \times L + \alpha_{H\times}H$, $\beta = \beta_L \times L + \beta_{H\times}H$ and $\lambda = \lambda_L \times L + \lambda_{H\times}H$. Thus the logic of the likelihood function is as follows: candidate values of these six parameters are proposed, the linear function evaluated so that we know candidate value of α, β and λ for each of the Low and High frequency treatments, the expected utility of the actual choice is evaluated using the Tversky and Kahneman [1992] specification, and then the log-likelihood function specified above is evaluated.

specification vary with the evaluation horizon. The "awkward news" for MLA is that they provide inconsistent evidence about risk attitudes in relation to the received empirical wisdom.

The estimates for α indicate *risk-loving behavior over gains.*[18] There does not appear to be much difference in risk attitudes over gains, and indeed one cannot reject the null hypothesis that they are equal with a Wald test (*p*-value = 0.391). The estimates for β indicate a severe case of *risk aversion over losses*. Moreover, subjects appear to be *more* risk averse in the Low frequency setting than in the High frequency setting: a Wald test of the null hypothesis of equality has a *p*-value of 0.074. Finally, the estimates for λ are *consistent with loss aversion*, since they are both each significantly greater than 1 (*p*-values < 0.0001). However, these subjects appear to be significantly *more* loss averse in the High frequency setting than in the Low frequency setting (*p*-value = 0.0005).

4. Concluding Remarks

The experiments in Gneezy and Potters [1997] and Haigh and List [2005] do not test expected utility theory against myopic loss aversion. It is easy to estimate models using EUT-consistent utility functions that can fit the qualitative properties of their treatments.

However, these estimates provide no test of whether EUT is a better descriptor of these data than MLA. No doubt MLA can explain these data, and it is a separate matter to devise appropriate statistical ways to reconcile the two theories. One approach, advocated by Harrison and Rutström [2005] and restating a theme of Hey and Orme [1994], is to recognize that *some* individuals might make choices that are better described by EUT and that *some other* individuals might make decisions that are better described by MLA. That is, to discard the "representative agent" approach altogether. Such an approach, we conjecture, would leave room for both EUT and MLA to explain these data.

Moreover, as demonstrated by Harrison and Rutström [2005], the parameter estimates for each theory are often much more consistent with *a priori* expectations when one allows for the

[18] The Arrow-Pratt coefficient of relative risk aversion is $1-\alpha$, so $\alpha=1$ implies risk neutrality, $\alpha<1$ implies risk aversion, and $\alpha>1$ implies risk-loving behavior. These benchmarks are worth noting, to avoid confusion, given the popularity of specifications from Holt and Laury [2002] that estimate $1-\alpha$ directly (the risk-neutral value is 0 in that case, positive estimates indicate risk aversion, and negative estimates indicate risk-loving).

sample to be generated by two population processes instead of forcing the data to fit into one assumed population process. This may help reconcile the differences between estimates for these models presented above and estimates found in the previous literature.[19]

One might anticipate a defense that MLA provides a more parsimonious explanation of the data, and is therefore superior. This view is methodologically odd. First, it does not follow just in terms of "bean counting" the "things that have to be assumed." The EUT explanation would seem to require a two-parameter utility function, such as EP. But MLA requires a one-parameter[20] utility function, a loss aversion parameter, and some assumption that one know what is a loss and what is a gain, which presumes estimation of a reference point. Arbitrarily setting that reference point to 0 does not suffice, since this 0 must be defined relative to the task: in this case, it is implicit that the reference point for MLA is the stake, but that is still an assumption. Second, parsimony defined in terms of the cardinality of the parameters needed to specify a function misses many of the most important features of adopting one specification over another: local flexibility and global regularity being two of the most important.[21] If some function is parameter-parsimonious but locally inflexible (e.g., RN and CRRA), there is no obvious trade-off to another function that is slightly less parameter-parsimonious but locally flexible enough (e.g., EP).[22] Such parameter-counting debates are simply meaningless.

A deeper issue underlying the assessment of behavior from these experiments is asset integration within the laboratory session. What incomes are arguments of the utility functions of the

[19] The MLA differences have been noted. But it is only fair to point out that estimates of declining RRA are not consistent with the evidence from a wide range of lab experiments reported in Holt and Laury [2002][2005] and Harrison, Johnson, McInnes and Rutström [2005]. They find evidence of slightly *increasing* RRA over the prize levels considered here, although there is evidence of CRRA in the field data collected by Harrison, Lau and Rutström [2005]. On the other hand, that evidence of increasing RRA assumes that every agent is characterized best by EUT, and may itself be confounded by that assumption if some agents are better characterized as non-EUT decision-makers. In other words, one should not rush to *dismiss* a model because it generates estimates that differ from some other received estimates, providing there are ways to try to reconcile the two sets of estimates.

[20] The original version in Tversky and Kahneman [1992; p.57] assumed a two-parameter value function, with different power exponents for gains and losses. Most empirical work found those to be quite close, and the two are often assumed to be the same value.

[21] See Perroni and Rutherford [1995] for a careful statement of these concepts, which are of great practical importance in policy simulations in which shocks are "always" non-local.

[22] The use of parameter-parsimony as a metric of the *gravitas* of a model in experimental and behavioral economics has become quite common, and has marred otherwise important work in the area of learning models.

subjects? In the earlier analysis it was assumed to be simply the prizes over which they were making *choices* whenever they got to make a choice. But what about asset integration of income earned during the sequence of rounds? Gneezy and Potters [1997; p.636] note that this could affect risk attitudes in a more general specification, but assert that the effect is likely to be small given the small stakes. This may be true, but is just an assertion and deserves more complete study using the framework proposed by Cox and Sadiraj [2005]. These data provide an opportunity to study this question, since subjects received information on their intra-session income flows at different rates. Hence one could, in principle, test what linear function of accumulating wealth was relevant for their choices.

Another issue that is latent in this discussion is the one fundamental insight of Benartzi and Thaler [1995]: what is the evaluation horizon of decision-makers? The fact that the tests of MLA are simply not tests of MLA should not turn one away from examining this issue. The point is that it is an open issue for someone using an MLA specification just as it is an issue for someone using an EUTi specification. In this sense, the seminal contribution of the MLA literature and the experimental design of Gneezy and Potters [1997], is to force mainstream economists to pay attention to an issue they have neglected *within* their own framework.

Figure 1: Distribution of Percentage Bets
in Gneezy and Potters [1997] Experiments

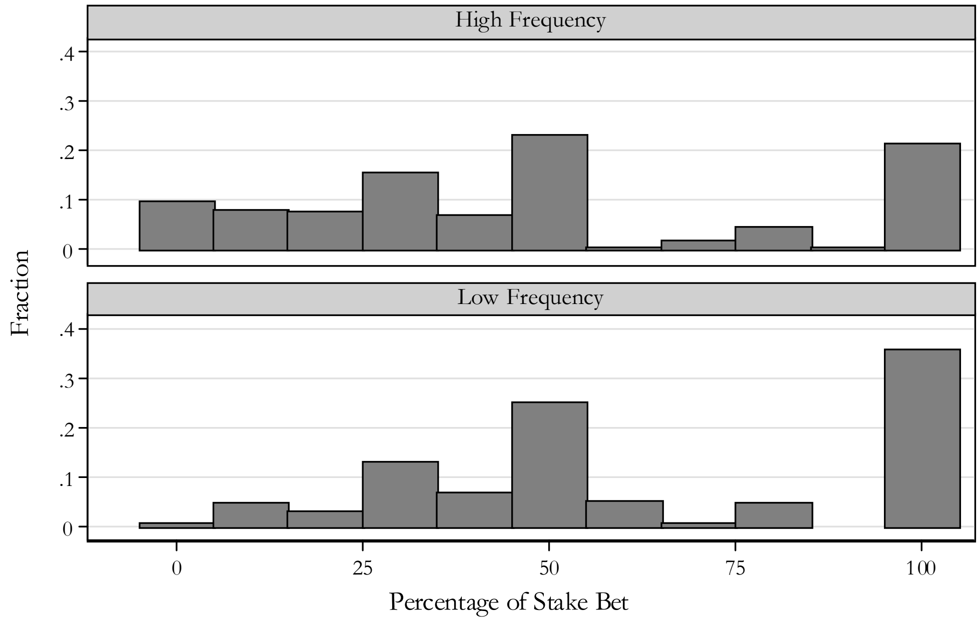

Figure 2: Distribution of Percentage Bets
in Haigh and List [2005] Experiments

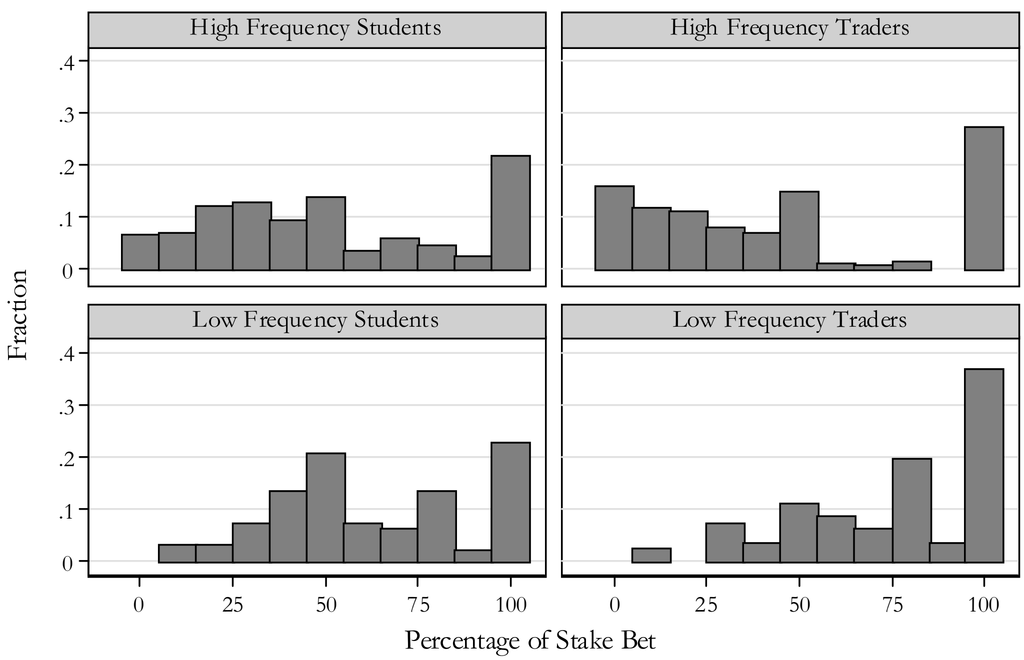

Table 1: Illustrative Calculations Assuming Risk Neutrality

Bold rows show EUT-consistent choices

Possible Choices		*High Frequency Treatment*					*Low Frequency Treatment*								
Bet in cents	Bet as %	L	p(L)	G	p(G)	EV	LLL	p(LLL)	LLG	p(LLG)	LGG	p(LGG)	GGG	p(GGG)	EV
0	0%	200	0.67	200	0.33	200.0	600	0.30	600	0.44	600	0.22	600	0.04	600
10	5%	190	0.67	225	0.33	201.7	570	0.30	605	0.44	640	0.22	675	0.04	605
20	10%	180	0.67	250	0.33	203.3	540	0.30	610	0.44	680	0.22	750	0.04	610
30	15%	170	0.67	275	0.33	205.0	510	0.30	615	0.44	720	0.22	825	0.04	615
40	20%	160	0.67	300	0.33	206.7	480	0.30	620	0.44	760	0.22	900	0.04	620
50	25%	150	0.67	325	0.33	208.3	450	0.30	625	0.44	800	0.22	975	0.04	625
60	30%	140	0.67	350	0.33	210.0	420	0.30	630	0.44	840	0.22	1050	0.04	630
70	35%	130	0.67	375	0.33	211.7	390	0.30	635	0.44	880	0.22	1125	0.04	635
80	40%	120	0.67	400	0.33	213.3	360	0.30	640	0.44	920	0.22	1200	0.04	640
90	45%	110	0.67	425	0.33	215.0	330	0.30	645	0.44	960	0.22	1275	0.04	645
100	50%	100	0.67	450	0.33	216.7	300	0.30	650	0.44	1000	0.22	1350	0.04	650
110	55%	90	0.67	475	0.33	218.3	270	0.30	655	0.44	1040	0.22	1425	0.04	655
120	60%	80	0.67	500	0.33	220.0	240	0.30	660	0.44	1080	0.22	1500	0.04	660
130	65%	70	0.67	525	0.33	221.7	210	0.30	665	0.44	1120	0.22	1575	0.04	665
140	70%	60	0.67	550	0.33	223.3	180	0.30	670	0.44	1160	0.22	1650	0.04	670
150	75%	50	0.67	575	0.33	225.0	150	0.30	675	0.44	1200	0.22	1725	0.04	675
160	80%	40	0.67	600	0.33	226.7	120	0.30	680	0.44	1240	0.22	1800	0.04	680
170	85%	30	0.67	625	0.33	228.3	90	0.30	685	0.44	1280	0.22	1875	0.04	685
180	90%	20	0.67	650	0.33	230.0	60	0.30	690	0.44	1320	0.22	1950	0.04	690
190	95%	10	0.67	675	0.33	231.7	30	0.30	695	0.44	1360	0.22	2025	0.04	695
200	**100%**	**0**	**0.67**	**700**	**0.33**	**233.3**	**0**	**0.30**	**700**	**0.44**	**1400**	**0.22**	**2100**	**0.04**	**700**

Note: p(LLL) = ⅔ × ⅔ × ⅔; p(LLG) = ⅔ × ⅔ × ⅓, and can occur in three equivalent ways (LLG, LGL and GLL), so the probability shown is ⅔ × ⅔ × ⅓ × 3; p(LGG) = ⅔ × ⅓ × ⅓, and can also occur in three equivalent ways; and p(GGG) = ⅓ × ⅓ × ⅓.

Table 2: Maximum Likelihood Estimates of Expo-Power Utility Function

Coefficient	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
	A. Gneezy and Potters [1997] – Estimates for All Bets by Dutch Students				
r	0.21	0.08	0.009	0.06	0.37
α	-0.02	0.03	0.463	-0.07	0.03
a	2.32	0.22	0.000	1.87	2.76
	B. Gneezy and Potters [1997] – Estimates for Interior Bets by Dutch Students				
r	1.12	0.25	0.000	0.61	1.63
α	-0.57	0.09	0.000	-0.74	-0.40
a	1.88	0.29	0.000	1.30	2.46
	C. Haigh and List [2005] – Estimates for All Bets by CBOT Traders				
r	0.36	0.05	0.000	0.26	0.46
α	-0.13	0.02	0.000	-0.16	-0.10
a	3.67	0.42	0.000	2.82	4.53
	D. Haigh and List [2005] – Estimates for Interior Bets by CBOT Traders				
r	0.67	0.04	0.000	0.60	0.74
α	-0.44	0.01	0.000	-0.46	-0.42
a	3.69	0.34	0.000	3.01	4.37
	E. Haigh and List [2005] – Estimates for All Bets by UMD Students				
r	-0.99	0.27	0.001	-1.54	-0.44
α	0.22	0.05	0.000	0.13	0.32
a	1.71	0.21	0.000	1.28	2.13
	F. Haigh and List [2005] – Estimates for Interior Bets by UMD Students				
r	1.24	0.10	0.000	1.04	1.44
α	-0.60	0.02	0.000	-0.64	-0.56
a	1.96	0.11	0.000	1.75	2.17

Table 3: Maximum Likelihood Estimates of Myopic Loss Aversion Utility Function

Estimates from responses in Gneezy and Potters [1997] experiments

Coefficient	Variable	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
α	Low frequency	1.48	0.04	0.000	1.40	1.55
	High frequency	1.38	0.10	0.000	1.18	1.59
β	Low frequency	0.03	0.07	0.689	-0.11	0.17
	High frequency	0.55	0.28	0.052	0.00	1.10
λ	Low frequency	1.90	0.08	0.000	1.74	2.07
	High frequency	4.28	0.64	0.000	2.99	5.56

References

Benartzi, Shlomo, and Thaler, Richard H., "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 111(1), February 1995, 75-92.

Cox, James C., and Sadiraj, Vjollca, "Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory," *Unpublished Manuscript*, Department of Economics, University of Arizona, May 2005; forthcoming, *Games & Economic Behavior*.

Frederick, Shane; Loewenstein, George; and O'Donoghue, Ted, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, XL, June 2002, 351-401.

Friedman, Milton, and Savage, Leonard J., "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 46, August 1948, 279-304.

Gneezy, Uri; Kapteyn, Arie, and Potters, Jan, "Evaluation Periods and Asset Prices in a Market Experiment," *Journal of Finance*, 58, 2003, 821-838.

Gneezy, Uri, and Potters, Jan, "An Experiment on Risk Taking and Evaluation Periods," *Quarterly Journal of Economics*, 112, 1997, 631-645.

Gollier, Christian, *The Economics of Risk and Time* (Cambridge, MA: MIT Press, 2001).

Haigh, Michael S., and List, John A., "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *Journal of Finance*, 60(1), February 2005, 523-534.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005, 897-901.

Harrison, Glenn W., and Lau, Morten I., "Is the Evidence for Hyperbolic Discounting in Humans Just An Experimental Artefact?" *Working Paper 05-03*, Department of Economics, College of Business Administration, University of Central Florida, 2005; forthcoming, *Behavioral & Brain Sciences*.

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., and Rutström, E. Elisabet, "Representative Agents in Lottery Choice Experiments: One Wedding and A Decent Funeral," *Working Paper 05-18*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

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.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, 902-912.

Kahneman, Daniel, and Lovallo, Da, "Timid Choices and Bold Forecasts: A Cognitive Perspective on

Risk Taking," *Management Science*, 39(1), January 1993, 17-31.

Kahneman, Daniel, and Tversky, Amos, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 1979, 263-291.

Köbberling, Veronika, and Wakker, Peter P., "An Index of Loss Aversion," *Journal of Economic Theory*, 122, 2005, 119-131.

Kocherlakota, Narayana R., "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, March 1996, 42-71.

Liang, K-Y., and Zeger, S.L., "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 1986, 13-22.

List, John A., "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118, February 2003, 41-71.

List, John A., "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, 72(2), March 2004, 615-625.

Mankiw, N. Gregory, and Zeldes, Stephen P., "The Consumption of Stockholders and Non-Stockholders," *Journal of Financial Economics*, 29(1), March 1991, 97-112.

Markowitz, Harry, "The Utility of Wealth," *Journal of Political Economy*, 60, April 1952, 151-158.

Mosteller, Frederick, and Nogee, Philip, "An Experimental Measurement of Utility," *Journal of Political Economy*, 59(5), October 1951, 371-404.

Papke, Leslie E., and Wooldridge, Jeffrey M., "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 1996, 619-632.

Perroni, Carlo, and Rutherford, Thomas F., "Regular Flexibility of Nested CES Functions," *European Economic Review*, 39(2), 1995, 335-343.

Quizon, Jaime B.; Binswanger, Hans P., and Machina, Mark J., "Attitudes Toward Risk: Further Remarks," *Economic Journal*, 94, March 1984, 144-148.

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

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

Rogers, W. H., "Regression standard errors in clustered samples," *Stata Technical Bulletin*, 13, 1993, 19-23.

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

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

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

Thaler, Richard H; Tversky, Amos; Kahneman, Daniel, and Schwartz, Alan, "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test," *Quarterly Journal of Economics*, 112, 647-661.

Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representations of Uncertainty," *Journal of Risk & Uncertainty*, 5, 1992, 297-323; references to reprint in D. Kahneman and A. Tversky (eds.), *Choices, Values, and Frames* (New York: Cambridge University Press, 2000).

Williams, Rick L., "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics*, 56, June 2000, 645-646.

Wooldridge, Jeffrey, "Cluster-Sample Methods in Applied Econometrics," *American Economic Review (Papers & Proceedings)*, 93, May 1993, 133-138.