

# Hypothetical Bias Over Uncertain Outcomes

by

Glenn W. Harrison<sup>†</sup>

October 2005

Forthcoming in J.A. List (ed).,  
*Using Experimental Methods in Environmental and Resource Economics*  
(Northampton, MA: Elgar, 2005)

*Abstract.* The evidence for hypothetical bias over uncertain outcomes is reviewed. Consistent with the evidence for deterministic outcomes, it appears that subjects respond differently to risky prospects when they face real economic consequences of their choices instead of hypothetical economic consequences. Implications for contingent valuation survey design and behavioral economics are discussed.

<sup>†</sup> Department of Economics, College of Business Administration, University of Central Florida, USA. E-mail: GHARRISON@BUS.UCF.EDU. I am grateful to John Kagel and Susan Laury for making detailed experimental results available, to the Danish Social Science Research Council for research support under project #24-02-0124, to Steffen Andersen, Morten Lau and Elisabet Rutström for discussions, and an anonymous referee for helpful suggestions. All data and statistical code are stored in the ExLab Digital Library at <http://exlab.bus.ucf.edu>.

One of the major contributions of experimental methods to environmental economics has been the characterization of hypothetical bias. A long series of experiments has established evidence of differences in responses to tasks that involve real economic commitments when compared to comparable tasks involving hypothetical economic commitments. In addition, there have been constructive attempts to use the laboratory environment to design instruments to mitigate the extent of the bias or to correct for it.<sup>1</sup> One gap in the previous literature has been the examination of hypothetical bias for outcomes that are uncertain. Although some of the commodities used in previous studies may have had some subjectively uncertain characteristics, those were not controlled for or explicit. This study fills that gap, by reviewing evidence for differences in responses to outcomes that are explicitly uncertain, focusing specifically on exogenous lotteries where the uncertainty is controlled and known *a priori*. In effect, we ask if estimates of risk attitudes defined over monetary outcomes suffer from hypothetical bias.

The relevance of characterizing hypothetical bias over uncertain outcomes should be apparent in the context of environmental valuation. Virtually every important environmental project includes some scientific or perceptual uncertainty. Many of the scenarios that are presented to subjects try to artificially remove any uncertainty, but often this entails less control than one might hope for since subjects are then likely to doubt the credibility of the artificially certain scenario. The danger is that they might then employ subjective assumptions that cannot be controlled for in the experiment. The implication is that one might elicit very different valuations if the scenario was presented openly as a policy lottery.

Sections 1 and 2 consider two series of experiments that considered the issue of hypothetical bias over uncertain outcomes: Battalio, Kagel and Jiranyakul [1990] and Holt and Laury [2002][2005]. Other experiments provide indirect opportunities for checking for hypothetical biases,

---

<sup>1</sup> See Cummings and Harrison [1994] and Shogren [2004]. The methods for correcting for hypothetical bias include *ex ante* approaches and *ex post* approaches, and are discussed extensively in Harrison [2005]. The former refer to efforts to design hypothetical instruments that better approximate the responses of instruments with salient incentives, perhaps by the use of “cheap talk” that brings the problem of hypothetical bias to the respondent’s attention. The latter refer to statistical methods for adjusting hypothetical responses to reflect systematic biases identified in comparable settings.

but these studies had this as one of their primary treatments.<sup>2</sup> Section 3 presents the results of a new experiment considering the effects of hypothetical bias, using subjects and procedures that match the salient experiments of Harrison, Johnson, McInnes and Rutström [2005]. Section 4 considers the sensitivity of inferences about hypothetical bias to alternative specifications of the underlying decision process, by allowing for probability weighting of choices. Section 5 draws implications from the results for the design and interpretation of contingent valuation surveys, and section 6 discusses implications for related debates in “behavioral economics.”

## **1. Battalio, Kagel and Jiranyakul**

### *A. Overall Design*

Battalio, Kagel and Jiranyakul [1990] and Kagel, MacDonald and Battalio [1990], hereafter BKJ and KMB, use a similar experimental design to collect information on human lottery choices. The subject is given a number of choice tasks, and told that one will be selected at random for payment at the end. Each subject received a \$30 endowment, and since only one choice will be paid out and the losses never exceed \$20, the subject knows that they will always leave the experiment with a gain of at least \$10. Some of the lotteries involve gains, and some involve losses, all relative to the initial stake: we refer to these as a gain-frame or a loss-frame. The experiments of BKJ and KMB span prizes of \$10, \$16, \$18, \$30, \$44 and \$50, roughly equally.

In some cases expected utility theory (EUT) makes predictions over triples of choice pairs,

---

<sup>2</sup> For example, Camerer [1989] included tests of hypothetical bias in his design. One of his payment treatments was to have subjects receive \$2 for completing the task, but no payments. Another treatment was to let subjects play out one of the small gain gambles, chosen at random. Finally, subjects in the small loss treatment were given \$10 and required to play out one of their choices, again chosen at random. Each subject made 12 choices, but only 4 of these were marked as eligible for selection to be played out. (Subjects were also asked to pick again for one of the 12, so there were actually 13 choices. In addition, half of the subjects that faced salient choices were given the option to change their choice before the final determination of their payoffs.) Furthermore, 4 of the choices involved large hypothetical gains, and the subjects were told explicitly that they were not to be played out. So there were, in fact, only 8 choices eligible for real payoffs. Thus the subjects knew in advance which 4 of this 8 were salient in the last two payoff treatments. There are three ways to check for hypothetical bias in this design. One is to compare the results from the responses to the large gains task, which were explicitly declared to be hypothetical (and the stakes would have implied that as well). Another is to compare the results from the responses to the small gains or small loss task that was implicitly not salient, assuming the subjects figured this out. The third is to compare the results from the subjects in the first payment treatment with those in the second. Each of these three has some advantages and disadvantages, but none provides a clean comparison without additional assumptions.

and in some cases it makes predictions of doubles of choice pairs. Those predictions are of no immediate import for our use of these data to characterize risk attitudes of the sample, other than for the fact that the data is reported in terms of frequencies of choice patterns over these triples or doubles, and not over the constituent choice pairs.

### *B. Hypothetical Bias Treatments*

BKJ included a controlled test of hypothetical bias in their design. Their “Series 1 design” in fact consisted of *in-sample* comparisons of hypothetical and real responses to the same lotteries: there were 41 hypothetical choices in the loss (gain) frame, matched with 15 real choices in the loss (gain) frame. The only substantive difference, apart from the salience of the consequences of the choices, was that the subjects in the hypothetical experiments did not receive any endowment.<sup>3</sup> Since there were some “prizes” in the hypothetical loss *and gain* frame experiments that entailed significant losses, we could proceed by assuming that the subjects behaved as if those losses would be covered by the experimenter out of an initial endowment.<sup>4</sup> Doing so, and pooling responses across individuals, we confirm the conclusion of BKJ that there is no qualitative effect of using hypothetical responses instead of real responses.<sup>5</sup> Since their conclusion has been widely cited, it is worth stating explicitly: “... despite these systematic and at times significant quantitative differences between responses to real versus hypothetical payoffs, qualitative conclusions regarding differences in risk attitudes over gains and losses were quite similar across both real and hypothetical choices;” (p.28). Their support for this conclusion consists of examination of hypothetical and real responses

---

<sup>3</sup> The instructions were also different in many other ways. In tests of hypothetical bias in valuation settings the experimental instructions have been generally designed to be virtually identical apart from the use of subjunctive language to describe the hypothetical task.

<sup>4</sup> Apart from ensuring comparability with the actual design of the real experiments, this avoids the problem of zero “contingent liability” in experiments. This refers to the problem of getting subjects to pay net losses in experiments, and the probability that subjects would then rationally behave as if risk-loving since they are not liable for the losses. This issue was raised by Hansen and Lott [1991] in the context of bidding behavior in experiments with common values.

<sup>5</sup> This conclusion derives from examination of the statistical significance of a dummy variable for hypothetical choices in a maximum-likelihood estimation of the coefficient in a constant relative risk aversion specification for all subjects in their Series 1. The same conclusion is drawn if estimation is solely over choices in the loss frame and choices in the gain frame.

on a *between-sample* basis.

However, the *in-sample* comparisons allowed by their design reveal that there is indeed a significant difference between risk attitudes in hypothetical and real settings. Figure 1 reports these comparisons. The left panels refer to choices in the loss frame, and the right panel to choices in the gain frame.<sup>6</sup> The top panels show the fraction of choices that were different when the same subject was asked in hypothetical or real mode.<sup>7</sup> These are all well above zero, and a *t*-test on each question confirms this conclusion. The direction of the change in risk preferences is also quite clear. In the loss frame more than 50% of the changes were in the direction of the subject expressing a reduction in risk aversion (or increase in risk loving), with one solitary exception. In the gain frame the reverse pattern obtains, with hypothetical responses being more risk averse (or less risk loving). Both sets of differences are again statistically significant from 0.5 using *t*-tests for each paired comparison.

How can one reconcile this conclusion with the one stated by BKJ (p.28)? The answer is that their conclusion was based on between-sample responses,<sup>8</sup> and referred to their conclusions about specific violations of EUT. Just because there is a change in the degree of risk aversion, there may not be a change in the extent to which the aggregate sample exhibits EUT violations. Unfortunately, the carefully worded conclusion of BKJ, which is correct as far as it goes, has been mis-characterized in some widely cited surveys of the effects of hypothetical response on lottery choices. For example, Camerer [1995; p.634] notes that several “... studies have compared hypothetical choices with real choices (in which one choice was played). They found either no effect or a slight tendency for playing gambles to yield more risk aversion.” Although this conclusion admittedly refers to several studies all at once, it over-states the inferences appropriate from the BKJ data. There is an effect on

---

<sup>6</sup> In the experimental *session* labeled “loss frame” subjects were also asked some *questions* in a gain frame, hence there are more paired comparisons in the latter than in the former. These are paired comparisons 6, 8, 9, 10 and 12 in Figure 1.

<sup>7</sup> To be conservative, and literal, any choice that a subject expressed indifference over is viewed as consistent with any other choice in the other mode. Thus, if a subject expressed indifference over a hypothetical pair and then expressed a strict preference in the paired real pair, we count that as consistent (since it is).

<sup>8</sup> Between-sample estimates do not allow one to control for unobserved individual effects with the statistical power that within-sample estimates do.

risk attitudes, and it differs in sign as one changes from a loss frame to a gain frame.<sup>9</sup>

## 2. Holt and Laury

### *A. Overall Design*

Holt and Laury [2002] (HL) provide a relatively transparent task for eliciting risk attitudes using a Multiple Price List (MPL).<sup>10</sup> Each subject is presented with a choice between two lotteries, which we can call A or B. Table 1 illustrates the basic payoff matrix presented to subjects. The first row shows that lottery A offers a 10% chance of receiving \$2 and a 90% chance of receiving \$1.60. The expected value of this lottery,  $EV^A$ , is shown in the third-last column as \$1.64, although the EV columns were not presented to subjects.<sup>11</sup> Similarly, lottery B in the first row has chances of payoffs of \$3.85 and \$0.10, for an expected value of \$0.48. Thus the two lotteries have a relatively large difference in expected values, in this case \$1.17. As one proceeds down the matrix, the expected value of both lotteries increases but the expected value of lottery B becomes greater than the expected value of lottery A.

The subject chooses A or B in each row, and one row is later selected at random for payout for that subject. The logic behind this test for risk aversion is that only risk-loving subjects would take lottery B in the first row, and only risk-averse subjects would take lottery A in the second last row.<sup>12</sup> Arguably, the last row is simply a test that the subject understood the instructions, and has no

---

<sup>9</sup> Our focus is solely on risk attitudes. It would be interesting to tabulate the in-sample choices of the BKJ subjects and re-do their tests of the effect of real rewards on the extent of their violations of EUT, rather than doing that on a pooled basis.

<sup>10</sup> The earliest use of the MPL design in the context of elicitation of risk attitudes is, we believe, Miller, Meyer and Lanzetta [1969]. Their design confronted each subject with 5 alternatives that constitute an MPL, although the alternatives were presented individually over 100 trials. It was subsequently used by Murnighan, Roth and Schoumaker [1988], although they only used the results to sort subjects into one group that was less risk averse than the other. Beck [1994] utilized it to identify risk aversion in subjects, prior to them making group decisions about the dispersion of everyone else's potential income. This allowed an assessment of the extent to which subjects in the second stage chose more egalitarian outcomes because they were individually averse to risk or because they cared about the distribution of income. The use of the MPL also has a longer history in the elicitation of hypothetical valuation responses in contingent valuation survey settings, as discussed by Mitchell and Carson [1989; p. 100, fn. 14].

<sup>11</sup> There is an interesting question as to whether they should be provided. Arguably the subjects are trying to calculate them anyway, so providing them avoids a test of the joint hypothesis that "the subjects can calculate EV in their heads and will not accept a fair actuarial bet." On the other hand, providing them may cue the subjects to adopt risk-neutral choices. The effect of providing EV information deserves empirical study.

<sup>12</sup> Friedman [1981] argues that subjects should never exhibit risk-loving behavior in the laboratory even if they are risk lovers, since they have cheaper ways to purchase uncertainty in the field (e.g., purchase of lottery tickets, or trips

relevance for risk aversion at all. A risk neutral subject should switch from choosing A to B when the EV of each is about the same, so a risk-neutral subject would choose A for the first four rows and B thereafter.

### *B. Experimental Design*

HL examine two main treatments with 212 subjects. The first is the effect of incentives. They vary the scale of the payoffs in the matrix shown in panel A of Table 1, which we take to be the scale of 1. Every subject was presented with the first matrix of choices shown in Table 1, and with the exact same matrix at the end of the experiment. These two choices were always given to all subjects, and we refer to them as task #1 and task #4. All subjects additionally had one *or* two intermediate choices, referred to here as task #2 and task #3. The question in task #2, if asked, was a *higher-scale, hypothetical* version of the initial matrix of payoffs. The question in task #3, if asked, was the *same* higher-scale version of payoffs but with *real* payoffs. Some subjects were asked one of these intermediate task questions; most subjects were asked both of them.<sup>13</sup> Thus we obtain the tabulation of individual responses shown in panel B of Table 1.

We see from panel B of Table 1 how each subject experienced different scales of payoffs in task #2 and/or task #3. This provides in-sample tests of the hypothesis that risk aversion does not vary with wealth, an important issue for those that assume specific functional forms such as Constant Relative Risk Aversion (CRAA) or Constant Absolute Risk Aversion (CARA), where the “constant” part in CRRA or CARA refers to the scale of the choices. A rejection of the “constancy” assumption is not a rejection of expected utility theory in general, of course, but just these particular (popular) parameterizations.

The second treatment in the HL design is the effect of hypothetical payoffs, which is why the questions in task #2 are included. We focus on these treatments below in more detail.

---

to a casino). This amounts to “field censoring” of lab response by field substitutes.

<sup>13</sup> Hence for *some* subjects task #4 was actually their third and last task.

Although having in-sample responses is valuable, it comes at a price in terms of control since there may be wealth effects from the subjects having earned some profit in the previous choice. To handle this potential problem HL use a nice trick: when the subjects proceed from task #1 to task #3, they are first asked if they are willing to give up their earnings in task #1 in order to play task #3. Since the stakes are so much higher in task #3, all subjects chose to do so. This means that the subjects face tasks #1 and #3 with no prior earnings from these experiments, although they do have experience with the type of task when facing task #3. No such trick can be applied for task #4, since the subjects would be unlikely to give up their earnings in task #3 in this instance. Thus the responses to task #4 have no controls for wealth built in to the design. However, we do know the actual earnings of the subjects from the experimental data.<sup>14</sup>

HL also ask each subject to fill out a detailed question of individual demographic information, so their data include a rich set of controls for differences in risk preferences due to these characteristics.

Figure 2 shows the main responses in the HL experiments with real responses. Consider the top left panel, which shows the average number of choices of the “safe” option A in each problem. Thus in problem 1, which is row 1 in Table 1, virtually everyone chooses option A (the safe choice). By the time the subjects get to problem 10, which is the last row in Table 1, virtually everyone has switched over to problem B, the “risky” option. The dashed line marked RN shows the prediction if each and every subject were risk-neutral: in this case everyone would choose option A up to problem 4, then everyone would choose option B thereafter. The solid line marked 1x shows the observed behavior in task #1, the low-payoff case. The solid line marked 20x shows the observed behavior in task #3, the high-payoff case that scales up the values in Table 1 by 20. The top right panel in Figure 2 shows comparable data for the 50x problems, and the bottom left panel shows

---

<sup>14</sup> All of the risk aversion experiments in Holt and Laury [2002][2005] appear to have been preceded by other experiments that generated income for the subjects. This could have added some noise to responses, although in principle one could condition on previous earnings.

comparable data for the 90x problems.<sup>15</sup> We examine the bottom-right panel later.

HL proceed with their analysis by looking at the first three pictures and drawing two conclusions. First, that one has to introduce some “noise” into any model of the data-generation process, since the observed choices are “smoother” than the risk neutral prediction. A more general way of saying this is to allow subjects to have a specific degree of risk aversion, but to assume that they all have exactly the same degree of risk aversion. Thus, if subjects were a little risk averse the line marked “Risk Neutral” (RN) would shift to the right and drop down a bit to the right, perhaps at problem 6 or 7 instead of problem 5.<sup>16</sup> Of course, it would no longer represent risk-neutral responses, but it would still drop sharply, and that is the point being made by HL when arguing for a noise parameter. Second, and related to the previous explanation, the best-fitting line that assumes homogenous risk preferences would have to be a bit to the right of the risk neutral line marked “Risk Neutral.” So some degree of risk aversion, they argue, is needed to account for the *location* of the observed averages, quite apart from the need for a noise parameter to account for the *smoothness* of the observed averages.

Both conclusions depend on the assumption that every subject in the experiment has the same preferences over risk. The smoothness of the observed averages is easily explained if one allows heterogenous risk attitudes and no noise at all at the individual level: some people drop down at problem 4, some more at problem 5, some more at problem 6, and so on. The smoothness that the eyeball sees in Figure 1 is then just a counterpart of averaging over this heterogeneous process. The fact that *some* degree of risk aversion is needed for *some* subjects is undeniable, from the positive area above the RN line and below the other lines from problems 5 through 10. But it simply does not follow without further statistical analysis that all subjects, or even the typical subject, exhibits

---

<sup>15</sup> The control data in these three panels, for the 1x problem, are pooled across all task #1 responses. That is, the task #1 responses in the bottom left panel of Figure 2 are not just the task #1 responses of the individuals facing the 90x problem. Nothing essential hinges on this at this stage of exposition.

<sup>16</sup> One concern with the use of MPL procedures is that subjects might gravitate to the middle of the table, and hence appear to be more risk-neutral than they really are. One can test these “framing” effects easily by experimental variations. Andersen, Harrison, Lau and Rutström [2004] reports such experiments, and note that there are some small framing effects with the MPL procedures employed by HL.

significant amounts of risk aversion.

These conclusions follow from inspection of each of the first three panels, and just the RN and 1x lines in each for that matter. Now turn to the comparison of the lines of observed choices *within* each of the first three panels. The eyeball suggests that the 20x, 50x and 90x lines are to the right of the 1x lines, which implies that risk aversion increases as the scale of payoffs increases. But this conclusion requires some measures of the uncertainty of these averages. Not surprisingly, the standard deviation in responses is the largest around problems 5 through 7, suggesting that the confidence intervals around these lines of observed choices could easily overlap. Again, this is a matter for an appropriate statistical analysis, not eyeball inspection of the averages.

Finally, compare the differences between the lines of observed choices as one scans across the first three panels in Figure 2. As the payoff scale gets larger, from 20x to 50x and then to 90x, it appears that the gap widens. That is, if one ignores the issue of standard errors around these averages, it appears that the degree of risk aversion increases. This leads HL to reject CRRA and CARA, and to consider generalized functional forms for utility functions that admit of increasing risk aversion. However, as panel B of Table 1 shows, the sample sizes for the 50x and 90x treatments were significantly smaller than those for the 20x treatment: 38 and 36 subjects, respectively, compared to 268 subjects for the 20x treatments. So one would expect that the standard errors around the 50x and 90x high-payoff lines would be much larger than those around the 20x high-payoff lines. This could make it difficult to statistically draw the eyeball conclusion that scale increases risk aversion.

Finally, one needs to account for the fact that all of the high-payoff data in the HL experiments were obtained in a task that followed the low-payoff task. Income effects were controlled for, in an elegant manner described above. But there could still be *simple order effects* due to experience with the qualitative task. HL recognize the possibility of order effects when discussing why they had the high hypothetical task before the high real task: “Doing the high hypothetical choice task before high real allows us to hold wealth constant and to evaluate the effect of using real

incentives. For our purposes, it would not have made sense to do the high real treatment first, since the careful thinking would bias the high hypothetical decisions.” The same (correct) logic applies to comparisons of the second real task with the first real task.

The bottom, right panel of Figure 2 examines the data collected by HL in task #1 and task #4, which have the same scale but differ only in terms of the order effect and the accumulated wealth from task #3. These lines appear to be identical, suggesting no order effect, but a closer statistical analysis that conditions on the two differences shows that there is in fact an order effect at work.

Figures 3, 4 and 5 show in detail the paired real and hypothetical responses for each of the scales used by HL. The same general conclusion emerges from each comparison: the real responses exhibit greater risk aversion than the hypothetical responses.

### *C. Design Issues*

An obvious design issue with the HL experiments is that inferences about scale and hypothetical bias are confounded by order. This effect is quite distinct from the effect of order on “income effects,” although that is also an issue for some of the responses. Harrison, Johnson, McInnes and Rutström [2005] demonstrated that the order effects in the real responses of HL were in fact statistically significant, and reduced by about one-half the effects of scale on risk aversion.

Holt and Laury [2005] agreed with the potential and estimated effects of order, and extended their earlier experiments to consider the effects of hypothetical bias without any potential confounds from order. Specifically, they conducted four sessions. One session had 1x payoffs with real rewards, and one session had 1x payoffs with hypothetical rewards. The other two sessions were the same but with 20x payoffs. Each session used different subjects, so the comparisons are all between-subjects. Each of the sessions with real rewards used 48 subjects, and each of the sessions with hypothetical rewards used 36 subjects.<sup>17</sup> We therefore use these new data to consider the effect of hypothetical

---

<sup>17</sup> An additional treatment was to control for the order of presentation of the task within each MPL table.

bias in their design, since they do not suffer from order effects.

#### *D. Hypothetical Bias Treatments*

HL were concerned with two issues at once: the constancy of risk aversion over the income domain that they scaled payoffs over, and the effect of hypothetical responses compared to real responses. To allow for the possibility that relative risk aversion is not constant we follow HL and estimate a flexible functional form, such as the Expo-Power (EP) function proposed by Saha [1993]. The EP function can be defined as  $u(y) = [1 - \exp(-\alpha y^{1-r})] / \alpha$ , where  $y$  is income and  $\alpha$  and  $r$  are parameters to be estimated using maximum likelihood methods. Relative risk aversion (RRA) is then  $r + \alpha(1-r)y^{1-r}$ . So RRA varies with income if  $\alpha \neq 0$ . This function nests CARA (as  $r$  tends to 0), but is not defined for  $\alpha$  equal to 0.

Maximum likelihood estimates of the EP model can be used to calculate the RRA for different income levels. The likelihood function we use here employs the same function used by Holt and Laury [2002] to evaluate their laboratory data, and indeed we replicate their estimates exactly.<sup>18</sup> Their likelihood function takes the ratio of the expected utility of the safe option to the sum of the expected utility of both options, where each expected utility is evaluated conditional on candidate values of  $\alpha$  and  $r$ . Their likelihood specification also allow for a “noise parameter,”  $\mu$ , to capture stochastic errors associated with the choices of subjects.

One important econometric extension of their approach is to allow each parameter,  $r$  and  $\alpha$ , to be a separate linear function of the task controls and individual characteristics, where we can estimate the coefficients on each of these linear functions. We also allow for the responses of the same subject to be correlated, due to unobserved individual effects. The data from Holt and Laury [2005] do not include information on individual characteristics, which is unfortunate since the

---

<sup>18</sup> Alternative statistical specifications might be expected to lead to different estimates of risk attitudes, although one would not expect radically different estimates. On the other hand, alternative specifications that deviate from traditional EUT, such as allowance for probability weighting, might lead to very different inferences about hypothetical bias.

treatments involve between-subject comparisons for which it is particularly important to control for observable differences in samples.

Table 2 displays the results from maximum likelihood estimation of the EP model. Treatment dummies are included for the tasks in which the order of presentation of the lotteries was reversed (variable “reverse”). This model allows for the possibility of correlation between responses by the same subject, since each subject provides 10 binary choices.<sup>19</sup> Panel A includes the data pooled from the hypothetical and real samples, and panels B and C estimate the model on each sample.

In general the treatment of “reversing” the order of presentation has no statistically significant effect on any parameter.

Panel A of Table 2 indicates that the real responses differ from the hypothetical responses solely in terms of the  $\alpha$  parameter, which controls the non-constancy of RRA in this EP specification. Since CRRA emerges in the limit as  $\alpha$  tends to 0, the hypothetical responses are consistent with CRRA roughly equal to 0.38 (the constant term on the  $r$  parameter). That is also the value for RRA with real responses when income levels are sufficiently low, since RRA is equal to  $r$  at zero income levels. These inferences are confirmed in Figure 6, which displays the predicted RRA in each treatment, along with a 95% confidence interval. *At low levels of income there is virtually no discernible difference between RRA for the hypothetical and real responses, but at higher income levels the real responses exhibit much higher RRA.* Thus hypothetical rewards provide reliable results precisely when they save the least money in terms of subject payments.

---

<sup>19</sup> 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.

### **3. Harrison, Johnson, McInnes and Rutström**

Harrison, Johnson, McInnes and Rutström [2005] (HJMR) replicated the basic experimental procedures developed by Holt and Laury [2002] to elicit risk attitudes, but avoided the order effects in their original design. The experimental results they reported all used salient incentives, but they also conducted some hypothetical experiments using subjects drawn from the same population and the same instruments. The effect of hypothetical bias can therefore be evaluated using those published and unpublished experiments. The hypothetical experiments only used the 10x design, which is to say that they were not preceded by a 1x treatment. Thus there are no order effects, and the responses should be compared to the real 10x responses reported in HJMR. The results is a sample of 46 hypothetical responses and 55 real responses. One feature of these data is that a rich array of individual characteristics was collected, and can be used to condition responses in the two samples.

Table 3 presents estimates from an interval regression model of the elicited CRRA interval. Since there is evidence that hypothetical responses may have different variances from real responses, as well as different means, these estimates also allow for multiplicative heteroskedasticity associated with the response being hypothetical or real.

The results provide further evidence that hypothetical responses are systematically different from real responses. The mean effect is significantly different, as is the variance. Relative risk aversion in the hypothetical setting is 0.21 lower than the real setting, and has a standard error that is 0.11 higher. Both effects are statistically significant, with  $p$ -values of 0.024 and 0.086 respectively.

### **4. Sensitivity to Alternative Formulations**

Behavior under uncertainty is an area in which there have been many alternative theories to standard EUT, starting with prospect theory and rank-dependent utility theory, and encompassing many other subsequent specifications. Starmer [2000] provides an excellent overview of the evolution of this literature, and its relation to experimental evidence.

To illustrate the sensitivity of inferences about hypothetical bias over risk attitudes to alternative specifications, consider the effect of allowing for “probability weighting” of outcomes. The idea of probability weighting was originally proposed by Edwards [1962], and was extended by other decision theorists in the 1960's and 1970's, and had a longer implicit tradition in psychometrics. Kahneman and Tversky [1979; p.280ff.] brought it to the attention of mainstream economists as one component of their prospect theory of choice under uncertainty.

To be specific, consider two functional forms that have been popular. Let  $w(p)$  denote the weighting function of probability  $p$ . The identity weighting function,  $w(p) = p$ , is employed by EUT and emerges as a special case of virtually all of the weighting functions employed. One functional form we use is the S-shaped function  $w(p) = p^\gamma / \{p^\gamma + (1-p)^\gamma\}^{1/\gamma}$  introduced by Tversky and Kahneman [1992], and widely used in other applied work using prospect theory. The other functional form is a generalization due to Prelec [1998], in which  $w(p) = \exp(-\delta(-\ln p))^\gamma$ . Previous statistical applications of these weighting functions have typically used restrictive functional forms for the utility (or value) function, such as assuming CRRA or even risk neutrality. We allow the utility function to be flexible, in the sense of using the same EP function employed in Section 2, along with these flexible<sup>20</sup> probability weighting functions.

Table 4 reports the results of maximum likelihood estimation using these probability weighting functions applied to the experimental data from Holt and Laury [2005]. Panel A uses the S-shaped function, and panel B uses the Prelec function.<sup>21</sup> Each lead to dramatically different inferences about the extent of hypothetical bias. Using the S-shaped function one would be led to infer that there is no hypothetical bias in the value function (née utility function), but that there is hypothetical bias in the probability weighting function. Using real rewards significantly shifts the estimates of  $\gamma$  down by 0.48 from the default value of 0.85. The effect on the parameter  $r$  of the EP value function remains negative, but it is smaller in size at only -0.13 and has a  $p$ -value of 0.32. On

---

<sup>20</sup> Flexible relative to the EUT specification of the identity function.

<sup>21</sup> The Prelec function is not defined for choices in which there is a probability of 0 or 1, so row 10 of the choices in Panel A of Table 1 is dropped when estimating it.

the other hand, using the Prelec function one would be led to infer that there is a significant effect of hypothetical bias on both the value function *and* the probability weighting function. In this case the  $r$  parameter of the EP value function is 0.46 lower with real rewards, and the estimates of the  $\delta$  parameter of the probability weighting function are much lower with real rewards. Since the Prelec functional form generalizes the S-shaped functional form in terms of the probability weighting behavior it admits, one might argue *a priori* that it is to be preferred.

These results are intended to be illustrative of the sensitivity of inferences about hypothetical bias to the precise specifications used. In maximum likelihood estimation of this kind one must expect some numerical instability, as three sets of parameters are “competing” to account for observed behavior: the parameters characterizing the value function, the noise parameter that can explain anything if it is large enough, and the parameters characterizing the probability weighting function. More precise characterizations of hypothetical bias that allow all three to play a role will require larger sample sizes and controls for observable sample differences (e.g., sex, race, etc.). It would be invalid to infer from the fragility of the inferences in Table 4 that there is no evidence that hypothetical bias matters in a systematic or robust manner. Instead, the correct conclusion is that the inferences one makes about hypothetical bias depend critically on what specifications of the underlying decision-making process one assumes.<sup>22</sup>

## 5. Implications for Contingent Valuation Surveys

In contingent valuation surveys a key issue is the credibility of the scenario. One factor encouraging incredibility and scenario rejection must be the need to make the scenario sound overly precise and known. Subjects are likely to respond naturally to scenarios presented with some uncertainty surrounding them. For this reason it becomes important to know that hypothetical bias

---

<sup>22</sup> It is difficult to come up with *a priori* arguments for the validity of probability weighting, or one or other functional form, being appropriate in specific settings. This may be possible as laboratory experiments provide a better sense of the performance of different assumptions in different settings, but that evidence is not yet available. In a related area, the choice of which “stochastic error story” to use when evaluating experimental data testing expected utility theory, Loomes, Moffatt and Sugden [2002] make the important point that there are likely to be significant interactions between model selection and stochastic specifications.

does exist for uncertain outcomes, and to design scenarios that include explicit statements about uncertainty.

For example, consider the scenario used in the contingent valuation survey undertaken for the state of Alaska after the *Exxon Valdez* oil spill (Carson et al. [1992]). The scenario asked subjects to consider willingness to pay for an escort ship and an emergency Norwegian sea net program. The subjects were told that until double hulling laws take effect in ten years they were to expect another large oil spill in Prince William Sound of the same size and potential scope as the *Exxon Valdez* oil spill. The escort ship and sea net scenario was presented as a way of reducing the chance of such an oil spill effectively to zero during the next ten years, upon which time the double hulling laws would presumably reduce the probability to zero. Many features of this scenario defy credibility. Anyone with a passing knowledge of the problems encountered after the actual oil spill would know that plans and implementation were two very different things, so a rational respondent that knew this or intuited it from common sense would correctly be asking what probabilities to attach to the claims of the scenario. Similarly, what probability is there that another large oil spill would occur within precisely 10 years, and how can one hope to state such a thing with any certainty? The subject is presumably filling in some probabilities to these outcomes, and the surveyor has no control over these subjective assessments.

In some settings there have been several scenarios presented to subjects that differ in terms of the “scope” of the environmental injury. The most notable contingent valuation study to do this was one undertaken for preservation of the Kakadu Conservation Zone in Australia (Imber, Stevenson and Wilks [1991]). In this case a between-subjects design was employed to reflect intrinsic scientific uncertainty about the likely ecological impacts of the proposed mining activity at the time that the survey needed to go out into the field. But each scenario was presented to subjects as if it were “the truth,” when in fact there was scientific uncertainty about which one would obtain. Of course, one could use those responses as the basis for an expected utility calculation of willingness to pay, imposing some probabilities on each conditional outcome. But it would be better to build such

inferences about uncertain outcomes into the design from the outset, rather than correcting for them by heroic assumption after the fact.

An ideal design would elicit valuations of final outcomes, subjective beliefs that each outcome would occur, and the risk attitudes of the respondent. These would preferably be elicited on a within-subjects basis, but one could use between-subjects designs with sufficient sample size and controls for observable sample characteristics. It would then be possible for the analyst to identify valuations that reflect the subjective uncertainty of outcomes, as well as to infer “corrected” probabilities of outcomes if the respondents made systematic errors in that area. It would, of course, be appropriate to report the original valuations as well as any that were corrected for errors in probability perception, to avoid concerns that voter sovereignty was being implicitly rejected.<sup>23</sup>

One might object that it is often hard to reduce “intrinsic uncertainty” about the credibility of features of a contingent valuation survey, just as it is almost always hard to make the good or policy “deliverable.” This is true, but does not mean that methods for identifying the effects of uncertainty and hypothetical bias cannot be developed, as reviewed in Harrison [2005]. It is an open question for future research if those methods will prove reliable in the case of hypothetical bias over uncertain outcomes.

In the case of uncertainty, one might also profit from using ideas from the statistical literature on “errors in variables” to identify the extent of hypothetical bias due to the uncertainty. For example, consider the simulation-extrapolation approach developed by Cook and Stefanski [1994] and Stefanski and Cook [1995]. Their idea is that one can always *add noise* to a variable, using a known random process, and that one can then use information on the relationship between the addition of noise and the coefficient of interest to extrapolate what the coefficient would be if the uncertainty were removed. To be concrete, they suggest generating additional noise with a Gaussian process with zero mean and known variance, where the variance is the estimated sample variance of

---

<sup>23</sup> One might also undertake corrections for the use of individual risk attitudes instead of social risk attitudes. Ongoing experimental work is examining the possible differences between these.

the variable suspected to have some errors of measurement. Add some noise to the suspect variable with variance scaled by a fixed parameter  $\lambda > 0$ , re-estimate the model, go back and change  $\lambda$  in a known way, add more noise to the original values of the variable, and repeat liberally as the computer runs over a long weekend. Then collect the estimates of the coefficients of interest in the model, and estimate a relationship between those estimates and the values of  $\lambda$  that were tried. Then solve this estimated relationship for  $\lambda = -1$  and one has an estimate of the effect of removing the measurement error. This idea is closely related to the notion of a “reduced-bias jackknife estimator,” introduced by Quenouille [1956], and can be readily applied in an experimental setting where one has control over the instruments. It is always easy to add noise to an instrument, and only slightly harder to add controlled and measurable noise.

## **6. Implications for Behavioral Economics**

There has been a parallel debate in behavioral economics over the validity of hypothetical bias in the context of lottery choices. For some reason, many proponents of behavioral economics insist on using task responses that involve hypothetical choices. One simple explanation is that many of the earliest examples in behavioral economics came from psychologists, who did not use salient rewards to motivate subjects, and this tradition just persisted. Another explanation is that an influential survey by Camerer and Hogarth [1999] is widely mis-quoted as concluding that there is no evidence of hypothetical bias in such lottery choices. Although one can dismiss this issue as a red herring in the context of debates over the validity of the empirical premisses of behavioral economics, such claims are important to evaluate in the context of environmental economics where there is a substantive issue at stake: the validity of choices elicited by hypothetical surveys, such as those employing the contingent valuation method.

What Camerer and Hogarth [1999] conclude, quite clearly, is that the use of hypothetical rewards makes a difference to the choices observed, but that it does not generally change the

inference that they draw about the validity of EUT.<sup>24</sup> Since the latter typically involve paired comparisons of response rates in *two lottery pairs* (e.g., in common ratio tests), it is logically possible for there to be (i) differences in choice probabilities in *a given lottery* depending on whether one use hypothetical or real responses, and (ii) no difference between the effect of the EUT treatment on lottery *pair* responses rates depending on whether one uses hypothetical or real responses.

Furthermore, Camerer and Hogarth [1999] explicitly exclude from their analysis the mountain of data from experiments on valuation<sup>25</sup> that show hypothetical bias. Their rationale for this exclusion was that economic theory did not provide any guidance as to which set of responses was valid. This is an odd rationale, since there is a well-articulated methodology in experimental economics that is quite precise about the motivational role of salient financial incentives (Smith [1982]). And the experimental literature has generally been careful to consider elicitation mechanisms that provide dominant strategy incentives for honest revelation of valuations, and indeed in most instances explain this to subjects since it is not being tested. Thus economic theory clearly points to the real responses as having a stronger claim to represent true valuations. In any event, the mere fact that hypothetical and real valuations differ so much tells us that at least one of them is wrong! Thus one does not actually need to identify one as reflecting true preferences, even if that is an easy task *a priori*, in order to recognize that there are differences in behavior between hypothetical and real responses.

---

<sup>24</sup> With one exception, I do not believe that this inference is supported by the existing data and experimental designs, but that is an issue well beyond the scope of the present study. That exception is Beattie and Loomes [1997], an excellent example of the type of controlled study of incentives that is needed to address these issues.

<sup>25</sup> The term “valuation” subsumes open-ended elicitation procedures as well as dichotomous choice, binary referenda, and stated choice tasks.

## References

- Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Elicitation Using Multiple Price Lists," *Working Paper 04-08*, Department of Economics, College of Business Administration, University of Central Florida, 2004.
- Battalio, Raymond C.; Kagel, John C., and Jiranyakul, K., "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty*, 3, 1990, 25-50.
- Beattie, Jane, and Loomes, Graham, "The Impact of Incentives Upon Risky Choice Experiments," *Journal of Risk and Uncertainty*, 14, 1997, 155-168.
- Beck, John H., "An Experimental Test of Preferences for the Distribution of Income and Individual Risk Aversion," *Eastern Economic Journal*, 20(2), Spring 1994, 131-145.
- Camerer, Colin F., "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 1989, 61-104.
- Camerer, Colin F., "Individual Decision Making," in Kagel, J.H., and Roth, A.E., (eds.), *The Handbook of Experimental Economics* (Princeton: Princeton University Press, 1995).
- Camerer, Colin, and Hogarth, Robin, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Framework," *Journal of Risk and Uncertainty*, 19, 1999, 7-42.
- Carson, Richard T.; Mitchell, Robert C.; Hanemann, W. Michael; Kopp, Raymond J.; Presser, Stanley; and Ruud, Paul A., *A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill* (Anchorage: Attorney General of the State of Alaska, November 1992).
- Cook, J.R., and Stefanski, L.A., "Simulation-Extrapolation Estimation in Parametric-Measurement Error Models," *Journal of the American Statistical Association*, 89, December 1994, 1314-1328.
- Cummings, Ronald G., and Harrison, Glenn W., "Was the Ohio Court Well Informed in Their Assessment of the Accuracy of the Contingent Valuation Method?," *Natural Resources Journal*, 34(1), Winter 1994, 1-36.
- Edwards, Ward, "Subjective Probabilities Inferred from Decisions," *Psychological Review*, 69, 1962, 109-135.
- Friedman, David, "Why There Are No Risk Preferrers," *Journal of Political Economy*, 89(3), 1981, 600.
- Hansen, Robert G., and Lott, John R., Jr., "The Winner's Curse and Public Information in Common Value Auctions: Comment," *American Economic Review*, 81, March 1991, 347-361.
- Harrison, Glenn W., "Experimental Evidence on Alternative Environmental Valuation Methods," *Environmental & Resource Economics*, 31, 2005 forthcoming.
- 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 Rutström, E. Elisabet, "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods," in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results* (North-Holland: Amsterdam, 2005).
- 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.
- Imber, David; Stevenson, Gay; and Wilks, Leanne, *A Contingent Valuation Survey of the Kakadu Conservation Zone* (Canberra: Australian Government Publishing Service for the Resource Assessment Commission, February 1991).
- Kagel, John H.; MacDonald, Don N., and Battalio, Raymond C., "Tests of 'Fanning Out' of Indifference Curves: Results from Animal and Human Experiments," *American Economic Review*, 80(4), September 1990, 912-921.
- Kahneman, Daniel, and Tversky, Amos, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47(2), March 1979, 263-291.
- Liang, K-Y., and Zeger, S.L., "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 1986, 13-22.
- Loomes, Graham; Moffatt, Peter G., and Sugden, Robert, "A Microeconomic Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 2002, 103-130.
- Miller, Louis; Meyer, David Edward, and Lanzetta, John T., "Choice Among Equal Expected Value Alternatives: Sequential Effects of Winning Probability Level on Risk Preferences," *Journal of Experimental Psychology*, 79(3), 1969, 419-423.
- Mitchell, Robert C., and Carson, Richard T., *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Baltimore: Johns Hopkins Press, 1989).
- Murnighan, J. Keith; Roth, Alvin E., and Shoumaker, Françoise, "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty*, 1(1), March 1988, 101-124.
- Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66(3), May 1998, 497-527.
- Quenouille, M.H., "Notes on Bias in Estimation," *Biometrika*, 43, 1956, 353-360.
- Rogers, W. H., "Regression standard errors in clustered samples," *Stata Technical Bulletin*, 13, 1993, 19-23.
- 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.
- Shogren, Jason F., "Experimental Methods and Valuation," in K-G. Mäler and J. Vincent (eds.), *Handbook of Environmental Economics. Volume 2: Valuing Environmental Changes* (Amsterdam: North-Holland, 2004).
- Smith, Vernon L., "Microeconomic Systems As An Experimental Science," *American Economic Review*,

72(5), December 1982, 923-955.

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

Stefanski, L.A., and Cook, J.R., "Simulation-Extrapolation: The Measurement Error Jackknife," *Journal of the American Statistical Association*, 90, December 1995, 1247-1256

Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 1992, 297-323.

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.

**Table 1: Design of the Holt and Laury [2002] Experiments**

A. Lottery Payoff Alternatives at Payoff Scale 1 Level

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

B. Sample Sizes and Design

Scale of Payoffs	Task				Total
	1	2	3	4	
1	212			212	424
20		118	150		268
50		19	19		38
90		18	18		36
All	212	155	187	212	766
Hypothetical?	No	Yes	No	No	

**Table 2: Estimates of Expo-Power Model**

Maximum likelihood estimates of expo-power utility function  $u(y) = [1 - \exp(-\alpha y^r)] / \alpha$ , using data from the HL [2005] experiments

Utility Function Parameter	Parameter Covariate	Point Estimate	Standard Error	t	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
<b>A. Pooled Data</b> (N=1680 choices by 168 subjects)							
<i>r</i>	Real responses	-0.018	0.117	-0.155	0.877	-0.250	0.213
	Reverse column order	0.068	0.094	0.726	0.469	-0.117	0.253
	Constant	0.386	0.101	3.808	0.000	0.186	0.586
$\alpha$	Real responses	0.072	0.034	2.097	0.038	0.004	0.140
	Reverse column order	-0.004	0.028	-0.142	0.888	-0.058	0.051
	Constant	0.005	0.026	0.187	0.852	-0.047	0.057
$\mu$	Real responses	0.014	0.026	0.549	0.584	-0.037	0.066
	Reverse column order	-0.011	0.026	-0.439	0.661	-0.062	0.039
	Constant	0.108	0.015	7.043	0.000	0.078	0.138
<b>B. Real Responses Only</b> (N=960 choices by 96 subjects)							
<i>r</i>	Reverse column order	-0.027	0.148	-0.179	0.858	-0.321	0.268
	Constant	0.412	0.099	4.179	0.000	0.216	0.608
$\alpha$	Reverse column order	-0.015	0.037	-0.390	0.697	-0.089	0.060
	Constant	0.084	0.030	2.794	0.006	0.024	0.144
$\mu$	Reverse column order	-0.008	0.047	-0.180	0.857	-0.101	0.084
	Constant	0.121	0.033	3.661	0.000	0.055	0.187
<b>C. Hypothetical Responses Only</b> (N=720 choices by 72 subjects)							
<i>r</i>	Reverse column order	0.197	0.268	0.734	0.465	-0.337	0.730
	Constant	0.347	0.106	3.279	0.002	0.136	0.557
$\alpha$	Reverse column order	-0.056	0.200	-0.279	0.781	-0.455	0.344
	Constant	0.012	0.019	0.607	0.546	-0.027	0.050
$\mu$	Reverse column order	-0.014	0.027	-0.537	0.593	-0.068	0.039
	Constant	0.108	0.015	7.395	0.000	0.079	0.137

Note: each utility function parameter is estimated as a linear function of the covariates indicated. For example, in Panel A the utility function parameter *r* is estimated as  $0.386 - 0.018 \times \text{REAL} + 0.068 \times \text{REVERSE}$ , where REAL and REVERSE are binary dummy variables reflecting the use of real responses and the reverse column order, respectively.

**Table 3: Estimates of CRRA Interval Regression Model**

Maximum likelihood estimates of CRRA utility function  $u(y) = y^{1-r}/(1-r)$ , using published and unpublished data from the HJMR [2005] experiments

Parameter	Parameter Covariate	Point Estimate	Standard Error	<i>p</i> -value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r	Hypothetical responses	-0.208	0.092	0.024	-0.389	-0.027
	Female	0.102	0.078	0.188	-0.050	0.255
	Black	0.052	0.144	0.719	-0.230	0.334
	Age	0.009	0.012	0.434	-0.014	0.032
	Major is in business	-0.089	0.077	0.244	-0.240	0.061
	Sophomore in college	0.211	0.144	0.143	-0.072	0.494
	Junior in college	-0.035	0.133	0.795	-0.296	0.227
	Senior in college	-0.044	0.137	0.747	-0.313	0.225
	High GPA (greater than 3.75)	0.043	0.090	0.632	-0.134	0.221
	Low GPA (below 3.24)	-0.159	0.104	0.126	-0.363	0.045
	Graduate student	-0.100	0.171	0.561	-0.435	0.236
	Expect to complete a higher degree	-0.117	0.091	0.195	-0.295	0.060
	Father completed college	0.171	0.105	0.105	-0.036	0.377
	Mother completed college	-0.139	0.095	0.142	-0.325	0.046
U.S. citizen	-0.212	0.119	0.074	-0.445	0.020	
Constant		0.568	0.313	0.069	-0.045	1.181
$\sigma_r$	Hypothetical responses	0.110	0.064	0.086	-0.016	0.236
	Constant	0.286	0.033	0.000	0.221	0.351

**Table 4: Effects on Hypothetical Bias of Probability Weighting**

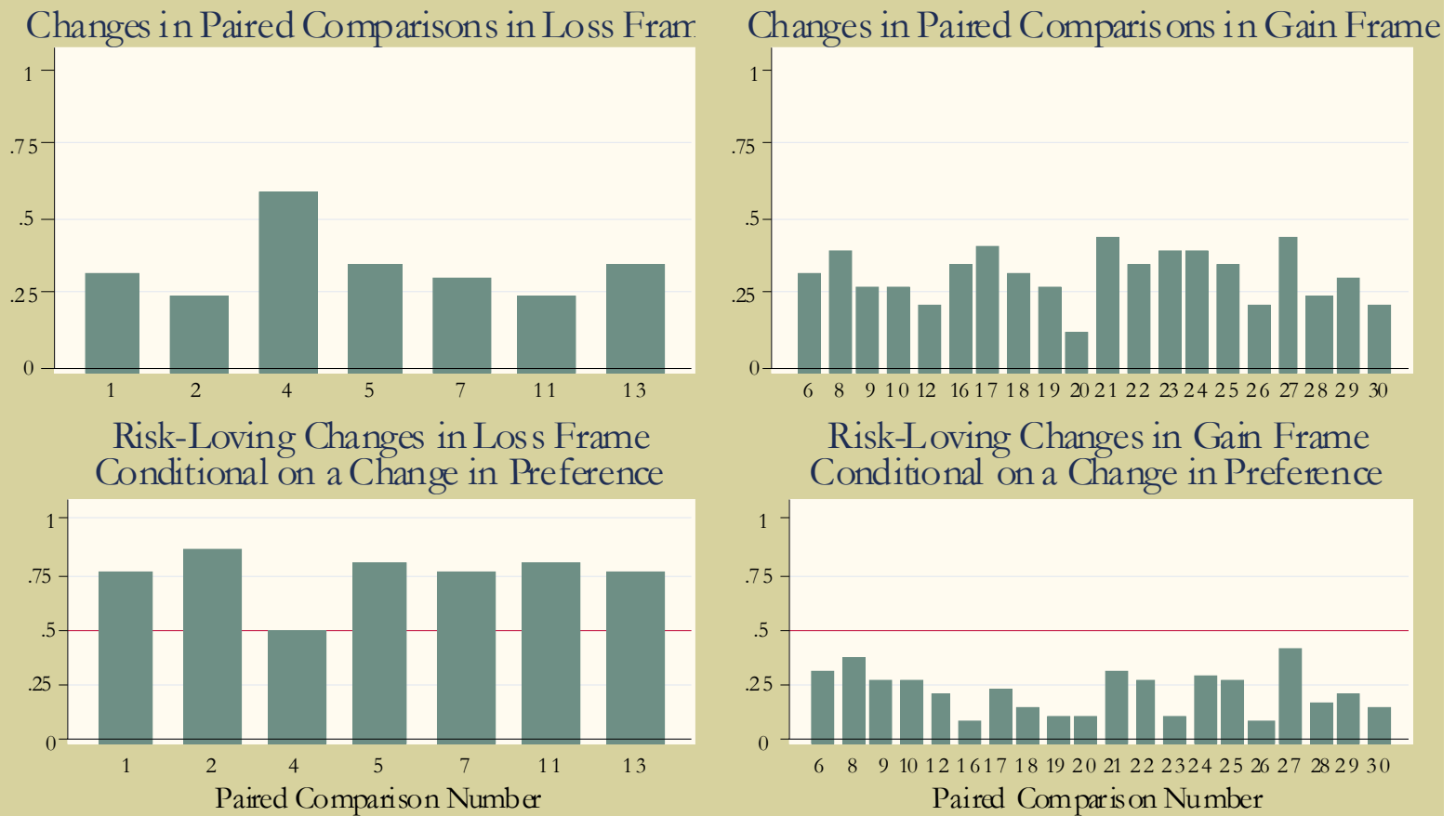
Maximum likelihood estimates of expo-power utility function  $u(y) = [1 - \exp(-\alpha y^{1-\gamma})] / \alpha$ , and specified probability weighting function, using pooled data from the HL [2005] experiments

Utility Function Parameter	Covariate	Point Estimate	Standard Error	t	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
<b>A. S-shaped Function <math>w(p) = p^\gamma / \{p^\gamma + (1-p)^\gamma\}^{1/\gamma}</math></b>		(N=1680 choices by 168 subjects)					
$r$	Real responses	-0.128	0.127	-1.004	0.317	-0.378	0.123
	Reverse column order	0.014	0.079	0.170	0.865	-0.143	0.170
	Constant	0.374	0.119	3.136	0.002	0.139	0.609
$\alpha$	Real responses	0.014	0.026	0.532	0.595	-0.037	0.065
	Reverse column order	-0.002	0.015	-0.167	0.868	-0.032	0.027
	Constant	0.004	0.021	0.215	0.830	-0.036	0.045
$\mu$	Real responses	-0.032	0.029	-1.124	0.263	-0.089	0.025
	Reverse column order	-0.009	0.022	-0.429	0.669	-0.053	0.034
	Constant	0.100	0.019	5.354	0.000	0.063	0.137
$\gamma$	Real responses	-0.483	0.258	-1.873	0.063	-0.993	0.026
	Reverse column order	-0.020	0.184	-0.107	0.915	-0.383	0.344
	Constant	0.855	0.208	4.103	0.000	0.444	1.267
<b>B. Prelec Function <math>w(p) = \exp(-\delta(-\ln p))^\gamma</math></b>		(N=1512 choices by 168 subjects)					
$r$	Real responses	-0.456	0.251	-1.815	0.071	-0.952	0.040
	Reverse column order	0.074	0.176	0.421	0.674	-0.273	0.421
	Constant	0.195	0.110	1.770	0.079	-0.023	0.413
$\alpha$	Real responses	-0.004	0.005	-0.879	0.381	-0.013	0.005
	Reverse column order	-0.002	0.003	-0.651	0.516	-0.008	0.004
	Constant	0.002	0.004	0.435	0.664	-0.006	0.009
$\mu$	Real responses	-0.218	0.115	-1.884	0.061	-0.445	0.010
	Reverse column order	0.014	0.065	0.221	0.826	-0.114	0.142
	Constant	0.028	0.057	0.493	0.623	-0.084	0.140
$\gamma$	Real responses	-0.041	0.175	-0.235	0.814	-0.386	0.304
	Reverse column order	0.047	0.112	0.417	0.677	-0.174	0.268
	Constant	0.460	0.149	3.092	0.002	0.166	0.754
$\delta$	Real responses	-2.016	1.003	-2.010	0.046	-3.995	-0.036
	Reverse column order	0.161	0.786	0.205	0.838	-1.390	1.713
	Constant	0.367	0.771	0.475	0.635	-1.156	1.890

Note: each utility function parameter and probability weighting function parameter is estimated as a linear function of the covariates indicated. For example, in Panel A the probability weighting function parameter  $\gamma$  is estimated as  $0.855 - 0.483 \times \text{REAL} - 0.020 \times \text{REVERSE}$ , where REAL and REVERSE are binary dummy variables reflecting the use of real responses and the reverse column order, respectively.

# Figure 1: Hypothetical Bias in BKJ Experiments

Fraction of Within-Sample Responses (N=33)



# Figure 2: Real Responses in HL [2002] Experiments

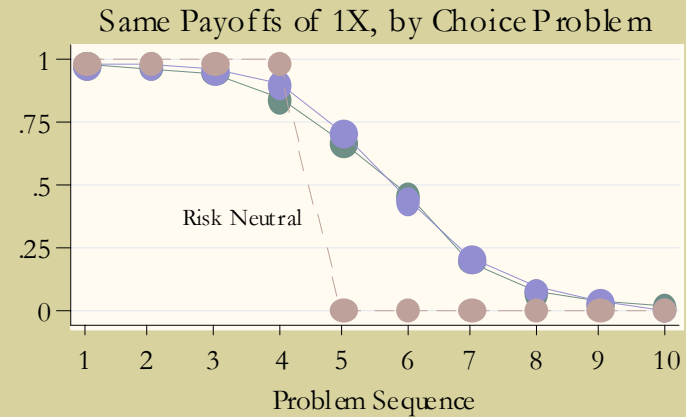
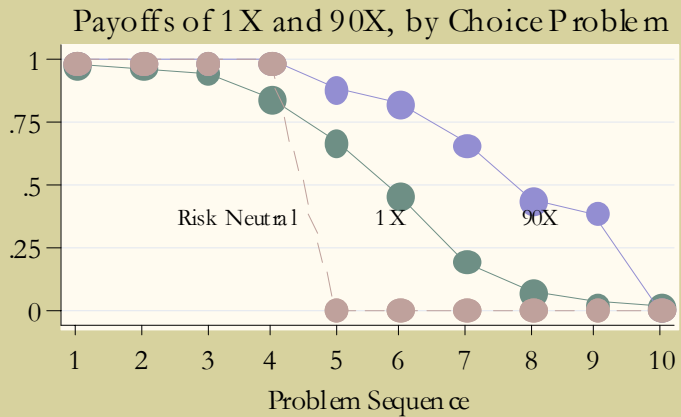
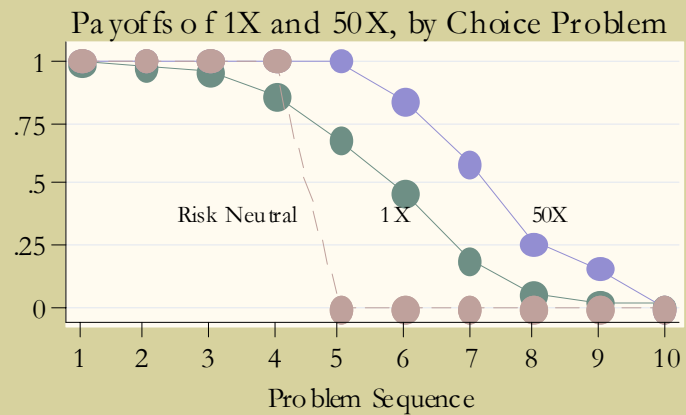
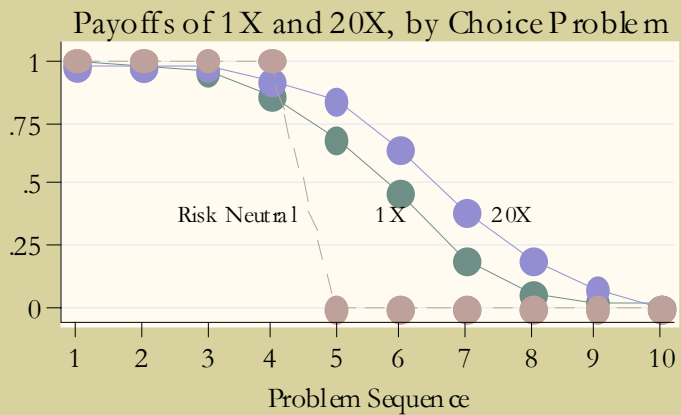


Figure 3: Hypothetical Bias in HL [2002]  
Experiments With 20x Payoffs

Fraction choosing safe option A in each choice problem

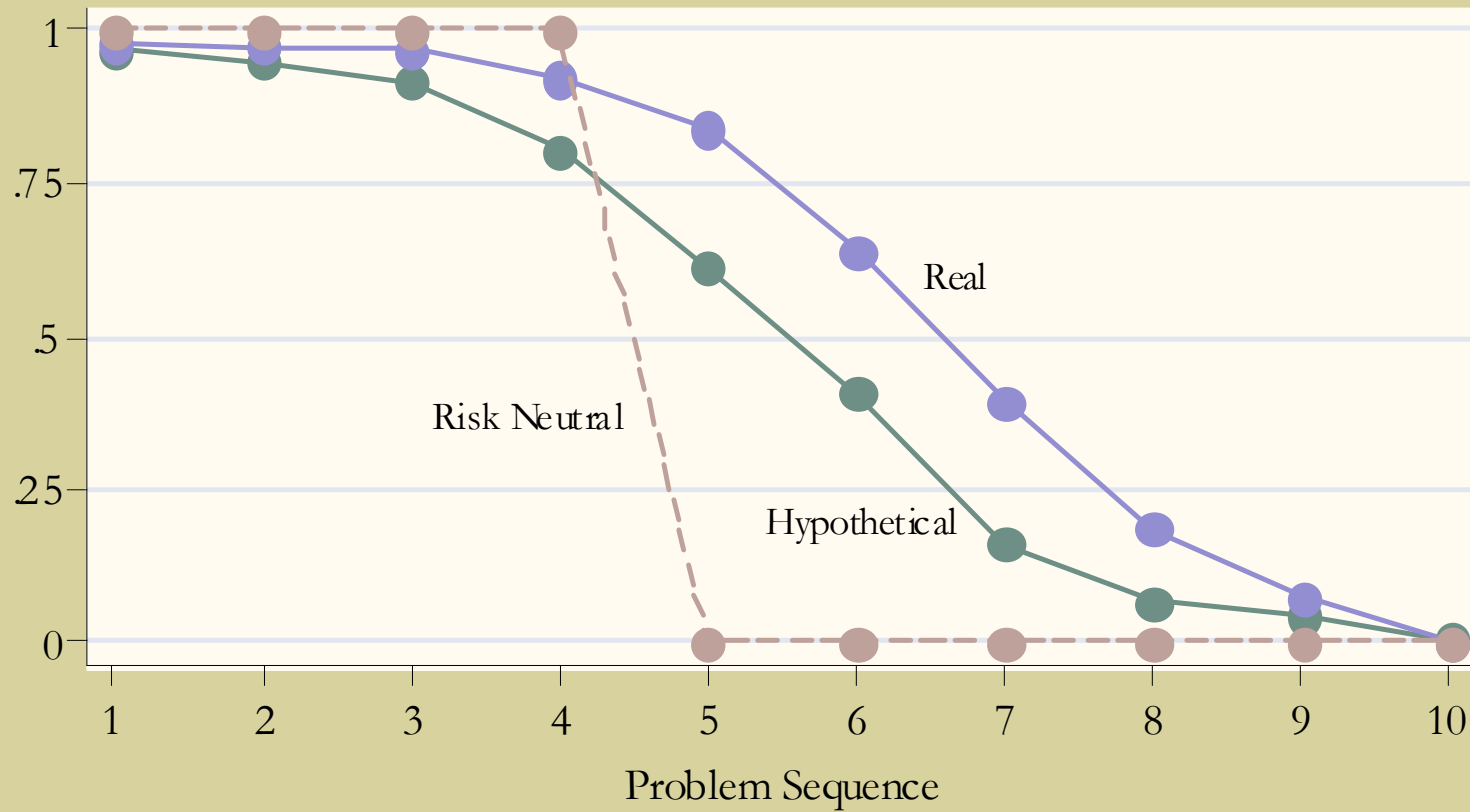


Figure 4: Hypothetical Bias in HL [2002]  
Experiments With 50x Payoffs

Fraction choosing safe option A in each choice problem

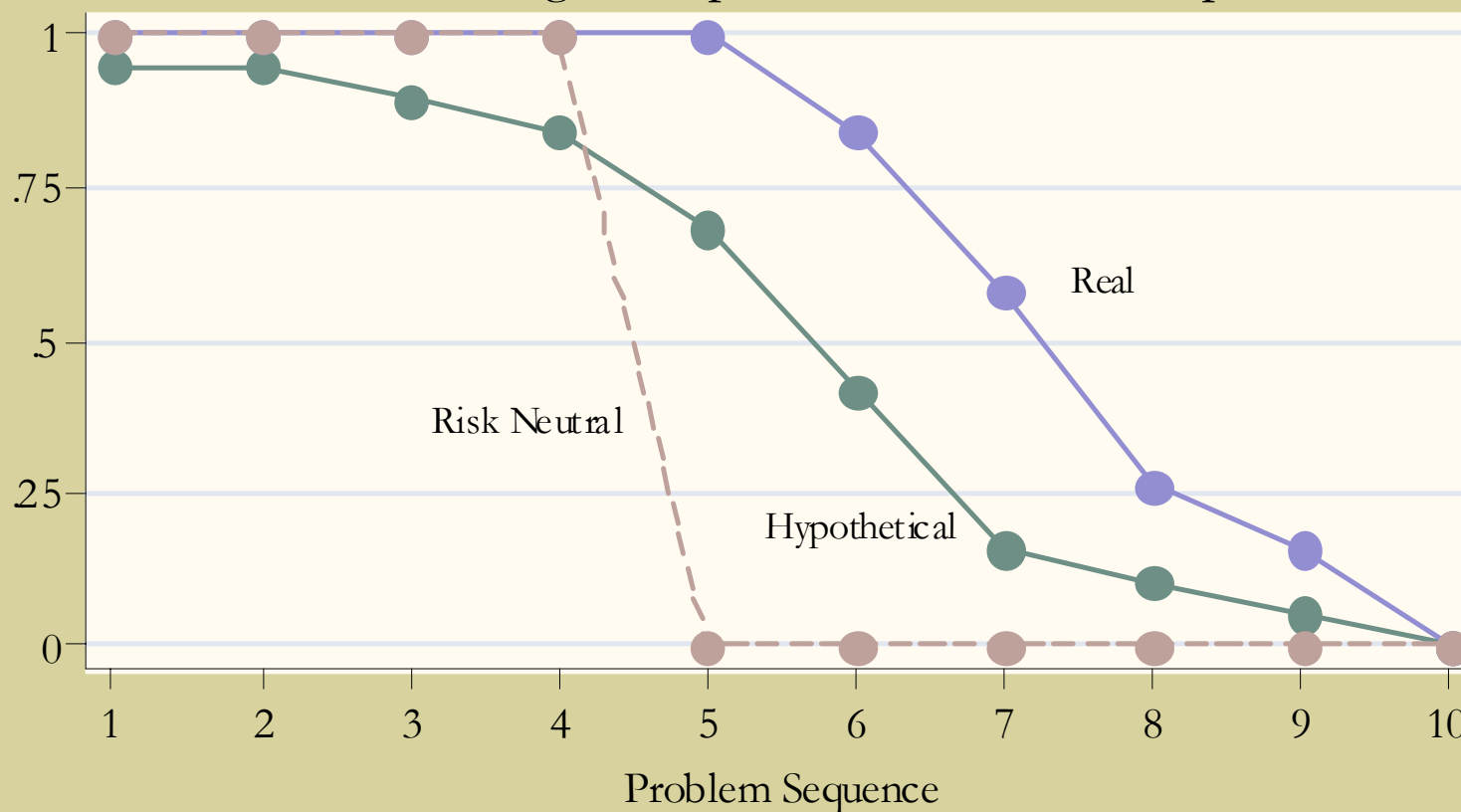


Figure 5: Hypothetical Bias in HL [2002]  
Experiments With 90x Payoffs

Fraction choosing safe option A in each choice problem

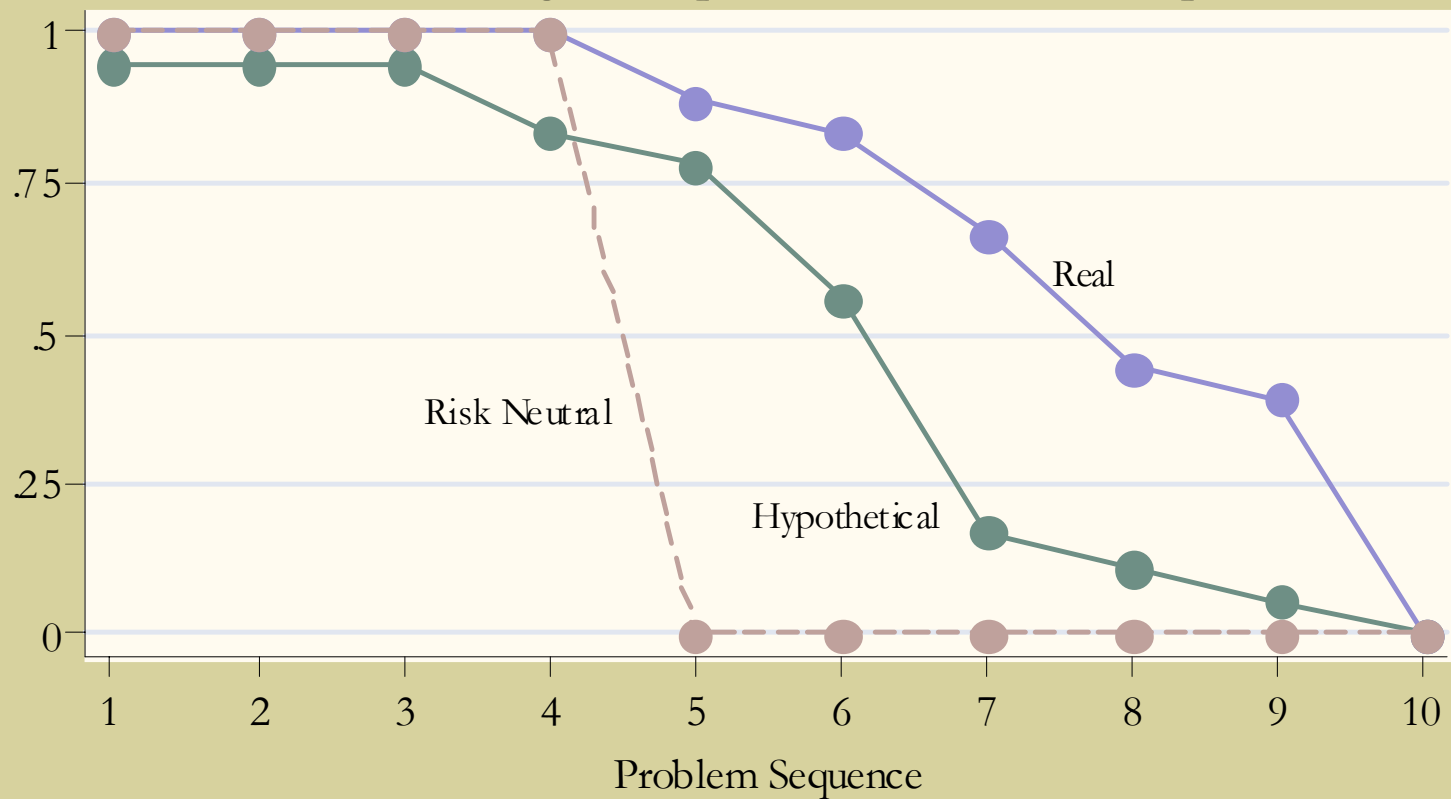


Figure 6: Hypothetical Bias in HL [2005] Experiments  
Predicted RRA from maximum likelihood expo-power model

