

Risk Aversion and Incentive Effects: Comment

By GLENN W. HARRISON, ERIC JOHNSON, MELAYNE M. MCINNES, AND
E. ELISABET RUTSTRÖM*

Charles A. Holt and Susan K. Laury (2002) develop an experimental design to determine the risk attitude of an individual. They use their observations to argue that increased incentives appear to change risk attitudes, leading to greater risk aversion. Popular utility functions that do not allow for such effects are therefore misspecified. Building on this finding, they estimate a flexible utility function that characterizes their aggregate data well, but that does not assume constant (absolute or relative) risk aversion.

The basic Holt and Laury (2002) (HL) design should become an important tool for the interpretation and design of experiments in which risk attitudes could play a role. However, their most important result—showing the effect of scaling up the stakes of the lottery choice task—is confounded by a possible order effect. An order effect occurs when prior experience with one task affects behavior in a subsequent task. The primary methodological contribution of experiments in economics is to enhance control. Ideally, such control makes the explanatory variables of interest, in this case scale, orthogonal to other explanatory variables such as order, allowing clearer inferences about behavior than one could get from field econometric data. We argue that not controlling for order effects

results in a misspecification of utility functions that is as important as that of scale.

The subjects in the HL design were given sequences of three or four tasks, as shown in Table 1. Task #1 involved choices over lotteries with a baseline level of prizes, which we will refer to as the “1× scale.” Task #2 involved hypothetical choices over prizes with a scaled-up level of prizes, either 20×, 50×, or 90×. Task #3 repeated task #2, but with choices that involved real payoffs. Task #4 was a return to the baseline task with real 1× prizes. In some sessions, subjects were not given task #2 or task #3.¹

What could one infer from the comparison of measured risk attitudes in the 1× and 20× tasks in HL? Unfortunately, any observed difference could be due either to the scale of the prizes involved or the task order, or some combination of both. Thus, the effect of scale is intrinsically confounded with the possible effect of task order.² This is a logical flaw in their design, which

¹ HL recognize the possibility that wealth effects could also confound the effects of scale in an in-sample design. To handle this they use a clever device: when the subjects proceed from task #1 to task #3, they are 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 income built into the design.

² HL are aware (p. 1647, fn. 4) of the possible effect of task order, albeit in the context of the real choices in task #3 being affected by the immediately preceding hypothetical task #2 with identical prize values. HL also included task #4 as a check for consistency of order effects for the 1× lottery. Nevertheless, the lack of a significant difference in choices between task #1 and task #4 does not prove conclusively that there are no order effects, since it is possible that subjects gave the same responses due to some sense of obligation or preference for being consistent. Moreover, there are no controls for wealth effects in their task #4. Our design does not suffer from such a confound.

* Harrison: Department of Economics, College of Business Administration, University of Central Florida, Orlando, FL 32816 (e-mail: gharrison@bus.ucf.edu); Johnson: Department of Economics, Moore School of Business, University of South Carolina, Columbia, SC 29208 (e-mail: ejohnson@moore.sc.edu); McInnes: Department of Economics, Moore School of Business, University of South Carolina, Columbia, SC 29208 (e-mail: mcinnes@moore.sc.edu); Rutström: Department of Economics, College of Business Administration, University of Central Florida, Orlando, FL 32816 (e-mail: erutstrom@bus.ucf.edu). We thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/MRI 9871019, and NSF/POWRE 9973669 to Rutström and grant NSF/SES 021429.0 to McInnes.

TABLE 1—DESIGN OF THE HOLT AND LAURY RISK AVERSION EXPERIMENTS

A. Task order										
Task #	Payoffs						Scale			
1	Low, real						1×			
2	High, hypothetical						20×, 50×, or 90×			
3	High, real						20×, 50×, or 90×			
4	Low, real						1×			

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

Note: The last three columns in this table, showing the expected values of the lotteries, were not shown to subjects.

can be fixed either by assuming away the possibility of order effects or redesigning their experiment and controlling for them. One may argue that the effect of increasing stakes from 20× to 50× (and 90×) represents unconfounded evidence in favor of a stake effect,³ but this assumes that there is no interaction effect between order and stake. If the order effect increases with the stakes, then it is possible that much of the change in preferences between 20× and 50× is due to order, not scale. Further, if the effect of scale on relative risk aversion is diminishing, the confound will become more pronounced as the stakes are increased.

Fortunately, order effects are easy to control for by design. We undertook a new series of experiments that build closely on the basic design features of HL, but allow an identification of the extent to which the apparent scale effects on risk aversion are actually order effects. We focus on the two real decisions from HL in

which there is a change in payoff scale. In our first treatment, the subjects were asked to make the choices given in panel B of Table 1, and then provided the opportunity to give up those earnings in return for the chance to participate in choices with payments scaled up by 10.⁴ Scaling the base payoffs, which have prizes ranging between \$0.10 and \$3.85, provides responses that span prizes between \$1.00 and \$38.50. This comfortably covers the range of prizes needed to apply the measures of risk aversion to most experiments. We call this the 1×10× treatment. In the second treatment, a different sample of subjects was given only the 10× task. This allows us to test whether the responses in the 10× task are affected by the prior experience of having seen the 1× task. Thus, our design allows us to disentangle the effects of order and scale by conducting a between-subjects analysis of the 10× responses from the 10× and 1×10× treatments for pure order effects, and a within-subjects analysis of

³ Using nonparametric tests, the only significant scale effect across the 20×, 50×, and 90× tasks is between 20× and the two higher scales. There is no significant effect between 50× and 90×. This may be due to the small sample sizes in the 50× and 90× treatments.

⁴ We chose the 10× scaling, rather than the 20× scaling used by HL, since we were measuring risk aversion for use in later experiments for which 10× spanned the range of payoffs.

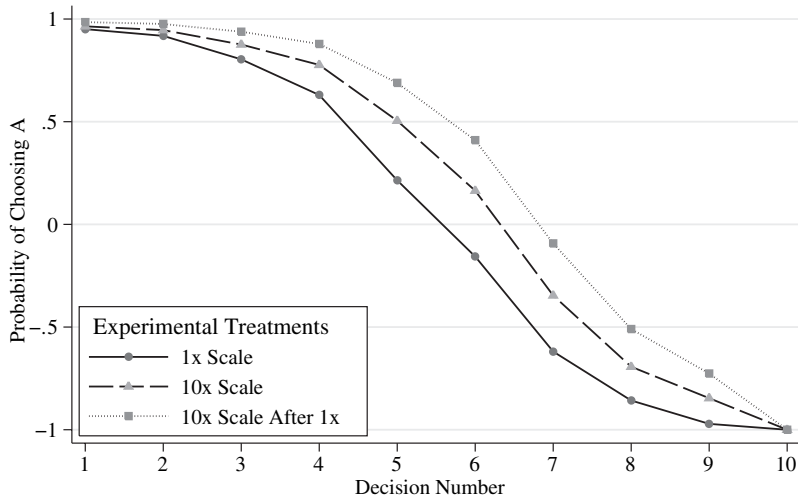


FIGURE 1. PREDICTED PROBABILITY OF A SAFE CHOICE
Ordered Probit Prediction of Choice by Order, Scale, and Decision Number

the $1 \times 10 \times$ responses for scale effects combined with order effects.

We recruited 178 subjects in 11 sessions in October and November 2002 at the University of South Carolina. Of these, 55 participated in the $10 \times$ experiments and 123 in the $1 \times 10 \times$ experiments.⁵

Like HL, we find that the majority of subjects choose the “safe” option (lottery A in Table 1) when the probability of the higher payoff is small and then switch over to the “risky” option (lottery B in Table 1) as the difference in expected value increasingly favors the risky option.⁶

Because the comparison of the $10 \times$ responses with and without prior experience is based on a between-subjects comparison, it is important to control for demographics. Differences in the composition of the subjects in the two treatment groups may otherwise confound inferences. To control for demographics, we

estimate an ordered probit regression model for the first risky response using the $10 \times$ data.⁷ We include a standard list of sociodemographic characteristics in the model. We also included a binary indicator for tasks that came second (the $10 \times$ task in the $1 \times 10 \times$ experiments), and dummy variables for each experimental session. Results from estimating the ordered probit model clearly show the importance of demographics and task order.

The ordered probit model allows us to construct Figure 1, which can be directly compared to the figures in HL, and shows the predicted probability of a safe choice for each of the treatments.⁸ To compare directly to HL, we

⁵ See project “Risk Aversion and Incentive Effects” in the *ExLab* Digital Library at <http://exlab.bus.ucf.edu>, for details of procedures and data analysis.

⁶ A small fraction, 16 percent at the $1 \times$ scale and 10 percent at the $10 \times$ scale, switch more than once. Most of these still choose safe for the first rows in Table 1, and then switch to risky for the last rows with some noise in the middle. Excluding these multiple switchers from the analysis made no difference to our conclusions.

⁷ The ordered probit specification has three advantages here. First, it recognizes the natural ordering of the ten decisions, which is a central feature of the experimental design. Second, it recognizes that the ten decisions we observe for each individual are not ten independent observations: the probability that an individual chooses a safe choice drops dramatically once a risky choice has been made. The third advantage of the ordered probit is that we can remain agnostic about the functional form of the utility function. We have also undertaken statistical analyses that do not rely on using the first risky choice, and draw the same conclusions about order effects.

⁸ For each individual and decision, we obtain the predicted probability of a safe choice and then sum over the sample to obtain the cumulative probabilities. We first estimated the model on the $10 \times$ data alone. By setting the order dummy equal to 0 or 1, we obtained predictions for

should compare the $1\times$ responses to the $10\times$ responses that follow the $1\times$ responses. The increase in risk aversion in this comparison is due to both order *and* scale effects. When we compare the $1\times$ responses to the $10\times$ responses with no prior experience, however, the pure scale effect is seen to be significantly smaller.

To measure the economic significance of the scale and order effects, we also estimate an interval regression model under the assumption of constant relative risk aversion (CRRA). This allows us to treat observations of subjects who choose A throughout with appropriate agnosticism, not imposing an arbitrary upper bound on their risk preference parameter. It also allows us to interpret observations of subjects who switch back and forth across several rows as having a wider switching interval, imposing a statistical interpretation of this uncertainty over their risk attitudes. Finally, it allows a direct interpretation of the magnitude of the effects in terms that are familiar to most economists, namely the CRRA coefficient.⁹ The interval regression model predicts that the average CRRA coefficient for the $1\times$ scale is 0.37, that it is 0.74 for the $10\times$ scale when there are both order and scale effects present, and that it is 0.57 in the $10\times$ scale when there are only scale effects. The coefficients on dummy variables capturing order and scale effects are each statistically significant, with *p*-values less than 0.05. Hence the order effect in the HL design confounds the inference about scale effects, such that the true scale effect is a little over one-half of the apparent effect when scale and order are confounded.

We therefore reaffirm the primary conclusion of HL, that risk aversion varies over the income range found in typical experiments. The effect is significantly smaller than they estimate, but the presence of a basic confound in their design does not lead one to reject their qualitative conclusion. Nevertheless, we conclude that order effects are significant and almost as large as scale effects, so that they can lead to misspeci-

fications of utility functions that are as serious as those of scale effects unless properly controlled for.

Practical implications arise for both academic researchers and policy analysts. First, since we find that pure scale effects cause smaller increases in risk aversion than suggested by HL, auction theorists may in many cases be able to continue using the CRRA specification as a local approximation to more general functional forms. The empirical task now is to define better the domains over which CRRA remains a valid approximation, and that is likely to vary with the population sampled and the context of the task. Similarly, in tests of Expected Utility Theory (EUT) certain combinations of risk attitudes and lottery parameter values lead to indifference between lottery pairs, implying that any choice pattern could be consistent with EUT if risk attitudes were uncontrolled.¹⁰ In order to make such tests of EUT operationally meaningful, one therefore has to select the lottery parameters conditionally on the risk attitudes of the respondents, and one has to have relatively precise estimates of risk attitude to do that.

Finally, because the predicted impacts of large-scale policy changes are uncertain,¹¹ policy analysts must allow for the risk attitudes of households when evaluating welfare changes. Controlling for known systematic effects in risk elicitation, such as order, is easy and removes an important bias. While it is certainly true that other variations in procedures may affect responses in the risk elicitation task, it makes little sense to fail to control for the procedural effects we can easily design away.

We are cautiously optimistic that risk preferences elicited with salient incentives and proper controls for order are robust,¹² particularly when analysts are armed with the flexible utility specification employed by HL to evaluate scale effects. We believe that careful experimental design and

the two $10\times$ treatments shown in Figure 1. We then re-estimated the model on the $1\times$ data alone to obtain the predictions for the $1\times$ treatment.

⁹ See Harrison et al. (2004, 2005) for further discussion of methodological issues in the design and analysis of elicitation designs such as these.

¹⁰ This issue is discussed in Harrison et al. (2003).

¹¹ In the field of computable general equilibrium models, there has long been a recognition that systematic sensitivity analysis of simulations conditioned on uncertain parameters implies uncertain policy impacts. See Harrison and H. D. Vinod (1992) for example.

¹² For example, Harrison, Johnson, McInnes, and Rutström (2005) show that risk preferences appear to be stable over time horizons of several months.

implementation can provide the needed control for most methodological and policy applications.

REFERENCES

- Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M. and Rutström, E. Elisabet.** "Individual Choice and Risk Aversion in the Laboratory: A Reconsideration." University of Central Florida, Department of Economics Working Paper: No. 3-18, 2003.
- Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M. and Rutström, E. Elisabet.** "Temporal Stability of Estimates of Risk Aversion." *Applied Financial Economics Letters*, 2005, *1*(1), pp. 31–35.
- Harrison, Glenn W.; Lau, Morten I. and Rutström, E. Elisabet.** "Estimating Risk Attitudes in Denmark: A Field Experiment." University of Central Florida, Department of Economics Working Paper: No. 04-07, 2004.
- Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet and Sullivan, Melonie B.** "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in Jeffrey Carpenter, Glenn W. Harrison and John A. List, eds., *Field experiments in economics*. Greenwich: JAI Press, 2005.
- Harrison, Glenn W. and Vinod, H. D.** "The Sensitivity Analysis of Applied General Equilibrium Models: Completely Randomized Factorial Sampling Designs." *Review of Economics and Statistics*, 1992, *74*(2), pp. 357–62.
- Holt, Charles A. and Laury, Susan K.** "Risk Aversion and Incentive Effects." *American Economic Review*, 2002, *92*(5), pp. 1644–55.