Eliciting Subjective Probability Distributions with Binary Lotteries

by

Glenn W. Harrison, Jimmy Martínez-Correa, J. Todd Swarthout and Eric R. Ulm †

July 2014

ABSTRACT.

We consider the elicitation of subjective belief distributions over continuous events using scoring rules with incentives. The theoretical literature suggests that risk attitudes have a surprisingly small role in distorting reports from true belief distributions. We use this theoretical prediction to test the effect of eliciting subjective belief distributions using a binary lottery procedure that should, in theory, lead to truthful reporting irrespective of the risk attitudes of the subject. In this instance this procedure leads to a prediction of "no effect" compared to using direct monetary payoffs to reward subjects. Of course, it is always possible that there is a behavioral effect from using the binary lottery procedure, contrary to the theoretical prediction. We demonstrate that the available controlled laboratory evidence is consistent with theory in this instance. If this result is true in general, then it expands the applicability of tools for eliciting subjective belief distributions.

Keywords: subjective belief distributions, risk attitudes, experiments

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, Copenhagen Business School, Denmark (Martínez-Correa); Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout); and Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, USA (Ulm). E-mail contacts: gharrison@gsu.edu, jima.eco@cbs.dk, swarthout@gsu.edu and eulm@gsu.edu.

Experimental economists would love to have a procedure to induce linear utility functions. Many inferences in economics depend on risk premia and the extent of aversion to risk. In fact, the settings in which these do not play a confounding role are the special case. Procedures to *induce* risk neutrality, and make this confound disappear by the intelligent use of theory and experimental design, have a long history. Unfortunately, these Binary Lottery Procedures (BLP) came under attack on behavioral grounds and the consensus appears to be that they may be fine in theory, but just do not work as advertized.

This consensus has recently been challenged. Starting with the use of the BLP for inducing risk neutral behavior in risky decisions over lotteries with *objective probabilities*, Harrison, Martínez-Correa and Swarthout (HMS) [2013] review the literature and note that there are many confounds in the majority of tests. For instance, several tests of the BLP are embedded in studies of strategic bidding in first-price sealed-bid auctions, requiring strong auxiliary assumptions about Nash equilibria. New experimental tests in non-strategic settings of individual choice over risky lotteries show support for the BLP. This starting point is important since it is easy to detect risk neutral behavior in settings with objective probabilities.

Tests of the BLP in settings in which *subjective probabilities* over binary events are elicited are much harder, since there is no simple way to infer the linearity of the utility function independently of inferences about the subjectively held probability (Savage [1971][1972]). However, HMS [2014] present experimental evidence that even in this setting there is a clear effect of the BLP to induce behavior consistent with linear utility functions. In this case the popular scoring rules for eliciting subjective probabilities imply a clear prediction if someone is risk averse: that reports will be closer to 50:50, in order to reduce the variability of payoffs from the two possible events. The extent of the pull towards a 50:50 report, relative to the true, latent subjective probability, depends on the curvature of the utility function under Subjective Expected Utility (SEU). This logic allows inference of the latent subjective probability if one knows the utility function of the subject, as demonstrated by Andersen, Fountain, Harrison and Rutström [2014]. Subjects who are risk averse will have a sizeable, first-order difference in their reports and inferred subjective probabilities, and subjects who are risk neutral will have no difference in their reports and the inferred subjective probabilities. Using an experimental procedure similar to the one described below, HMS [2014] show that the BLP does indeed generate different reports of subjective *probabilities* on a between-subjects basis, even though the subjects otherwise faced the same scoring rule and saw the same physical stimuli generating subjective probabilities.

We extend this evaluation of the BLP to the elicitation of *subjective belief distributions* over continuous events.

1. Theoretical Predictions

Let the decision maker report his subjective beliefs in a discrete version of a QSR. Partition the domain into K intervals, and denote as r_k the report of the density in interval $k = 1, 2,..., K$. The full report consists of a series of reports for each interval, $\{ r_1, r_2, ..., r_k ,..., r_K \}$ such that $r_k \geq 0\ \forall k$ and $\sum_{i = 1...K} (r_i) = 1$. If k is the interval in which the actual value lies, then the payoff score is from Matheson and Winkler [1976; p.1088, equation (6)]: $S = (2 \times r_k) - \sum_{i = 1...K} (r_i)^2$. So the reward in the score is a doubling of the report allocated to the true interval, and the penalty depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval the punishment is severe. The punishment includes all possible reports, including the correct one. A *risk neutral* decision maker would report his true subjective probability distribution when faced with this scoring rule.

To avoid any decision maker facing losses, allow some endowment, α, and scaling of the score, β. We then have the generalized scoring rule $\alpha + \beta\, [(2 \times r_k) - \sum_{i=1...K} (r_i)^2]$, where we initially assumed α=0 and β=1. We can assume α>0 and β>0 to get payoffs to any positive level we want.

Theoretical predictions for SEU decision makers who are *risk averse* are developed by Harrison, Martínez-Correa, Swarthout and Ulm [2012] (HMSU). In this case there is a striking difference in the theoretical predictions of using the BLP, compared to the case of subjective probabilities for a binary event: there is *no significant effect of "plausible" levels of risk aversion* on optimal reports compared to true latent subjective belief distributions. The qualitative effect of greater risk aversion is to cause the individual to report a "flatter" distribution than the true distribution, in order to reduce the variability of payoffs under events that are given positive true subjective probability of occurring. For the levels of risk aversion commonly observed in laboratory and field experiments, however, the effect is virtually imperceptible. Moreover, if one can assume that the latent, true subjective belief distribution is symmetric, risk averse decision makers will report their true average probability, even if there is some minuscule flattening of reports compared to the true distribution.

These theoretical properties of the QSR imply the prediction that *the BLP should have no perceptible effect* on elicited beliefs in this setting. This prediction is obviously qualitatively different than the theoretically predicted effect of the BLP in risky decisions over objective probabilities or over subjective probabilities for a binary event.

2. Experimental Design

Figures 1 and 2 illustrate the scoring rule for the case in which $K = 10$, $\alpha = \beta = 25$. Figure 1 shows the interface implementing the BLP, and Figure 2 the interface showing displays directly in money. Subjects could move the sliders at the bottom of the screen interface to re-allocate the 100 tokens as they wished, ending up with some preferred distribution. The instructions for the scoring rule defined directly in monetary payoffs explained that they could earn up to $50, but only by allocating all 100 tokens to one interval *and* that interval containing the true percent: if the true percent was just outside the selected interval, they would in that case receive $0.

Our experiment elicits beliefs from subjects over the composition of a bingo cage containing both red and white ping-pong balls. Subjects did not know with certainty the proportion of red and white balls, but they did receive a noisy signal from which to form beliefs. Table 1 summarizes our experimental design for each of 4 laboratory sessions at Georgia State University.

We implement two between-subjects treatments within each of sessions 1-4 so that both groups are *presented with the same randomly chosen and session-specific stimulus*. Thus we are able to compare treatment effects while conditioning on a specific realized stimulus. In **treatment 10m** we elicit subjective belief *distributions* about the true fraction of red balls in the bingo cage by using a generalized QSR with monetary outcomes (Figure 2). In **treatment 10p** we do the same thing but use an interface that rewards subjects with points that convert into increased probability of winning the better prize in a separate binary lottery (Figure 1).

Each session was conducted in the manner described in HMSU. Bingo Cage 1 was loaded with balls numbered 1 to 99 in front of everyone. A numbered ball was drawn from Cage 1, but the draw took place behind a divider. The outcome of this draw was not verified in front of subjects until after all decisions had been made. The number on the chosen ball from Cage 1 was used to construct Cage 2 behind the divider. The total number of balls in Cage 2 was always 100: the number of red balls matched the number on the ball drawn from Cage 1, and the rest were white balls. Once Cage 2 was constructed, the experimenter put the chosen numbered ball in an envelope and affixed it to the front wall of the laboratory.

Cage 2 was then covered with a black blanket and placed on a platform in the front of the room. When Cage 2 was then uncovered for subjects to see, it was spun for 10 turns, and covered again. This visual display was the information that each subject received. Subjects then made their decisions based on this information about the number of red and white balls in Cage 2. The sealed envelope was then opened, the chosen numbered ball was shown to everyone, and the experimenter publicly counted the number of red and white balls in Cage 2.

The stimulus, the number of red balls in Cage 2, was different in each session since we wanted the true number of red balls to be generated in a credible manner, to avoid subjects second-guessing the procedure. This credibility comes at the risk that the stimulus is extreme and uninformative; as it happens, we had a good variety of realizations over the 4 sessions.

3. Results

We have independent evidence that the subjects from our population do "robustly" exhibit risk aversion over stakes comparable to those used in the present experiment: see Holt and Laury [2002] and Harrison and Rutström [2008], for instance. Thus any success of the BLP is not due to the pre-existing risk neutrality of the subjects over these stakes.

Our maintained joint hypothesis is that subjects behave consistently with SEU *and* that their subjective belief distributions are distributed around the true population average that provides the common stimulus they all observe. Figure 3 reports the results across all sessions. With one exception, the elicited averages closely track the true averages.[1]

We formally statistically test the hypothesis that the elicited averages from treatments **10m** and **10p** in Figure 3 are equal to the true percent by estimating an interval regression model in which the intervals are the bin "labels" in Figures 1 and 2, and the tokens allocated to each bin are frequency weights for each subject. We also cluster the standard errors on each subject. If we estimate this model with only a constant term and no covariates, we can directly test the hypothesis that the estimate of the constant term is equal to the true percent. We find that the true percent accounts for 99.1% of the observed responses, and one cannot reject the hypothesis that it accounts for 100% of them (p-value = 0.70). Hence we cannot reject the null hypothesis that average elicited

[1] The is session 3, in which the true number of red balls was 11% and the elicited average using treatment **10m** was 25%. This disparity is due to three outliers, as explained in HMSU; we believe *a priori* these subjects did not understand the task.

beliefs are the same as the true percent.

Turning to the main hypothesis, we further find that the elicited beliefs from treatment **10m** and **10p** are not statistically different. Pooling the data over all 4 sessions with an interval regression, we estimate the elicited average to be 96% of the true average, and the responses with BLP to elicit responses that are 2.07 percentage points higher than the responses with direct monetary incentives. The 96% is not statistically different from 100% (p-value = 0.19), and the 2.07 is not statistically different from 0 (p-value = 0.23). This is consistent with our hypothesis that the BLP, if effective, should *not* make a difference to elicited beliefs in this setting (and, again, in contrast to the setting in which one elicits subjective *probabilities*). We interpret these results as evidence for the truthful elicitation of subjective belief distributions

4. Conclusion

These results provide clear support for the use of practical methods for eliciting subjective belief *distributions* over *continuous* events. We find that the binary lottery procedure does not distort elicited mean subjective beliefs in an experiment, consistent with theoretical expectations.

Figure 1: Belief Distribution Elicitation with Binary Lottery Payments

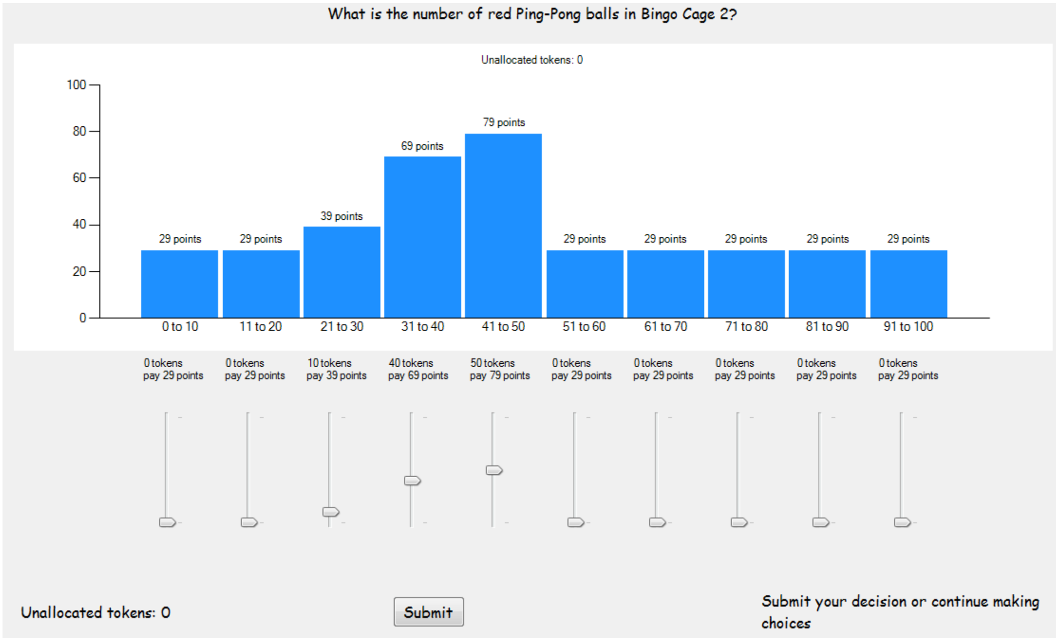


Figure 2: Belief Distribution Elicitation with Monetary Payments

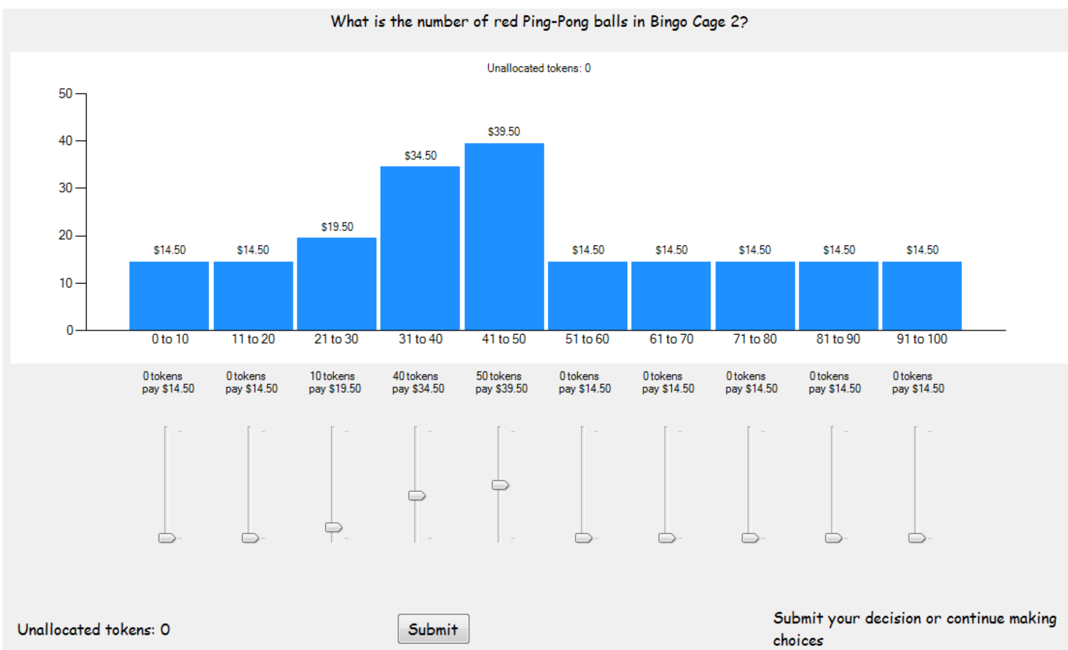


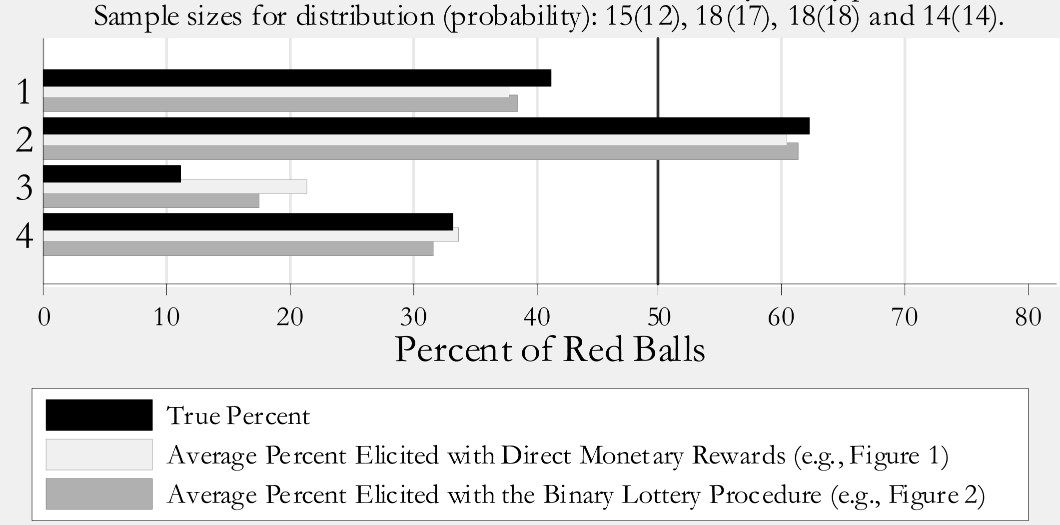
Table 1: Experiment Design and Sample Sizes

Session	Treatments		Total
	10m	10p	
1	15	12	27
2	18	17	35
3	18	18	36
4	14	14	28
Total	65	61	126

Notes: **treatment 10m** is elicitation of a distribution with the QSR defined directly over money, and **treatment 10p** is elicitation of a distribution with the QSR over binary lottery procedure "points" which convert into the probability of winning a high monetary prize.

Figure 3: Average of Elicited Subjective Belief Distributions using Money and the Binary Lottery Procedure

Pooled averages for each of 4 sessions, with treatments *within* each session.
Each session used the same random stimulus.
One treatment elicited beliefs with direct monetary payoffs,
and another treatment elicited beliefs with the binary lottery procedure.
Sample sizes for distribution (probability): 15(12), 18(17), 18(18) and 14(14).



References

Andersen, Steffen; Fountain, John; Harrison, Glenn W., and Rutström, E. Elisabet, "Estimating Subjective Probabilities," *Journal of Risk & Uncertainty*, 2014 forthcoming.

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Harrison, Glenn W, Martínez-Correa, Jimmy and Swarthout, J. Todd, "Inducing Risk Neutral Preferences with Binary Lotteries: A Reconsideration," *Journal of Economic Behavior & Organization*, 94, 2013, 145-159.

Harrison, Glenn W., Martínez-Correa, Jimmy, and Swarthout, J. Todd, "Eliciting Subjective Probabilities with Binary Lotteries," *Journal of Economic Behavior & Organization*, 101, May 2014, 128-140.

Harrison, Glenn W, Martínez-Correa, Jimmy; Swarthout, J. Todd, and Ulm, Eric "Scoring Rules for Subjective Probability Distributions," *Working Paper 2012-10*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012.

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

Matheson, James E., and Winkler, Robert L, "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22(10), June 1976, 1087-1096.

Savage, Leonard J., "Elicitation of Personal Probabilities and Expectations," *Journal of American Statistical Association*, 66, December 1971, 783-801.

Savage, Leonard J., *The Foundations of Statistics* (New York: Dover Publications, Second Revised Edition, 1972).