

Eliciting Subjective Probabilities with Binary Lotteries

by

Glenn W. Harrison^a, Jimmy Martínez-Correa^{b,*} and J. Todd Swarthout^c

February 2014

ABSTRACT.

We evaluate a binary lottery procedure for inducing risk neutral behavior in a subjective belief elicitation task. Prior research has shown this procedure to robustly induce risk neutrality when subjects are given a *single* risk task defined over *objective* probabilities. Drawing a sample from the same subject population, we find evidence that the binary lottery procedure also induces linear utility in a *subjective* probability elicitation task using the Quadratic Scoring Rule. We also show that the binary lottery procedure can induce direct revelation of subjective probabilities in subjects with popular Non-Expected Utility preference representations that satisfy weak conditions.

Keywords: Subjective probability elicitation, binary lottery procedure, experimental economics, risk neutrality

JEL classification: C81, C91.

[†] We are grateful for comments from two referees and participants at the Foundations and Applications of Utility, Risk and Decision Theory Conference, Georgia State University, Atlanta, 2012. We are also grateful to the Society of Actuaries and the Center for Actuarial Excellence Research Fund for financial support.

^a Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, P.O. Box 4036, Atlanta, Georgia 30302-4036, USA.

^b Department of Economics, Copenhagen Business School, Porcelaenshaven 16A, 3.57, DK-2000 Frederiksberg, Denmark.

^c Department of Economics, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, Georgia, 30302-3992, USA.

* Corresponding author. Tel.: + 45 3815 2349.

E-mail addresses: gharrison@gsu.edu (G.W. Harrison), jima.eco@cbs.dk (J. Martínez-Correa) swarthout@gsu.edu (J. T. Swarthout).

The Quadratic Scoring Rule (QSR) directly elicits subjective probabilities so long as subjects behave as if they are risk neutral. However, there is systematic evidence that subjects behave as if they are risk averse, even when facing the relatively small stakes normally used in the laboratory. We consider a procedure that should, theoretically, induce linear utility in applications of the QSR to directly elicit subjective probabilities. Our controlled experiment leads us to conclude that this procedure does indeed induce linear utility and accurately reveal latent subjective probabilities.

There are several ways to control for risk attitudes and recover underlying subjective probabilities. Andersen et al. (2014) illustrate how one can jointly estimate risk attitudes and subjective beliefs using a structural estimation approach.¹ There also exist other mechanisms for eliciting subjective probabilities without correcting for risk attitudes, such as the procedures proposed by Grether (1992), Köszegi and Rabin (2008, p.199), Offerman et al. (2009), Karni (2009) and Holt and Smith (2009).

We consider the use of proper scoring rules, such as the popular QSR, combined with a Binary Lottery Procedure (BLP) to induce linear utility in subjects. Binary lottery procedures to induce linear utility functions have a long history, with the major contributions being Smith (1961), Roth and Malouf (1979) and Berg et al. (1986).² The consensus appears to be that these procedures may be fine in theory, but behaviorally they just do not work as advertised (e.g., Cox and Oaxaca, 1995, and Selten et al., 1999). However, Harrison et al. (2013) show that the BLP induces linear utility over objective probabilities when contaminating factors such as strategic equilibrium concepts and traditionally used payment protocols are avoided. They find that the BLP induces risk neutrality when subjects are given *one* binary lottery choice over objective probabilities, and that it also works

¹ The need to do this jointly is in fact central to the operational *definition* of subjective probability provided by Savage (1954): under certain postulates, he showed that there existed a subjective probability *and a utility function* that could rationalize observed choices under subjective risk.

² See Harrison et al. (2013) for a detailed discussion on the literature of the BLP.

well when subjects are given more than one binary choice.

Our goal is to determine whether the BLP induces risk neutrality in simple binary choices defined over *subjective probability* elicitation tasks. We examine the ability of the QSR, combined with the BLP, to directly elicit subjective probabilities without controlling for risk attitudes. The first statements of this mechanism, joining the QSR and the BLP, appear to be Allen (1987) and McKelvey and Page (1990).³ Schlag and van der Weel (2013) and Hossain and Okui (2013) examine the same extension of the QSR, along with certain generalizations, calling it a “randomized QSR” and “binarized scoring rule,” respectively. Hossain and Okui (2013) test their theoretical results on scoring rules and *subjective probability* elicitation with an experimental design that has the drawback that they elicit beliefs about the composition of an urn that has probabilities that are *objectively known* by the subjects. A proper test of the ability of BLP to recover subjective probabilities should involve random processes that are not objectively known by subjects, and of course that provides a methodological challenge since one must simultaneously control subjective probabilities about some event *and* test if elicited responses are closer to these probabilities with the help of the BLP. Our experimental design captures this critical feature.

Using non-parametric statistical tests we find evidence that the BLP mitigates the distortion in reports introduced by risk aversion. Inferred subjective probabilities under the BLP robustly shift in the direction predicted under the plausible, empirically supported assumption that subjects are risk averse and that the BLP reduces the contaminating factor of risk aversion on optimal reports. Structural econometric estimations provide further evidence consistent with the hypothesis that the

³ There is no shortage of theoretical procedures to elicit subjective probabilities, and one natural question is which procedure generates them most reliably from a behavioral perspective. For example, Trautmann and van de Kuilen (2011) compare several incentivized procedures in the context of eliciting own-beliefs in a two-person game, and find few behavioral differences between the procedures. That elicitation context, while important, is complex, as stressed by Rutström and Wilcox (2009).

risk-attitudes-adjusted underlying subjective probabilities of subjects not exposed to the BLP are equal to the raw average reports of subjects exposed to the BLP.

1. Theoretical Issues

We review below the background theory that explains why the BLP, in conjunction with the QSR, induces subjects to directly report subjective probabilities. Our theoretical treatment of belief elicitation is framed in the terms used in the actual experiment and therefore provides a more direct translation from the theory to the experiment. Moreover, it allows us to derive conclusions that should prove useful for belief elicitation applications.

1.1. The Binary Lottery Procedure Under Subjective Expected Utility

The BLP induces linear utility functions in subjects if one assumes Subjective Expected Utility (SEU). The central insight when operationalizing the BLP is to define the payoffs in the QSR as points that define the probability of winning either a high or a low amount of money in some subsequent lottery defined over two known prizes. For example, set the high and the low payoff of this binary lottery to be \$50 and \$0. In theory the BLP induces subjects to report the true subjective probability of some event independently of the shape of the utility function.

In our experiment there are two events: a ball drawn from a Bingo Cage is either Red (R) or White (W). The subject bets on R and W. A subject betting on event R might estimate that it occurs with subjective probability π_R , and that W will occur with subjective probability $\pi_W = (1-\pi_R)$.⁴ The

⁴ The elicitor or experimenter does not need to know the latent subjective probability in order to define and pose a lottery that uses it. For instance, if I tell you that you can bet on whether you have gained or lost weight overnight, and that you get \$100 if you are correct and \$0 otherwise, I have defined a lottery whose valuation depends on your subjective probability about having gained weight. Your response to this single question will only tell me the sign of your subjective probability, not its value. For that one needs several well-defined lotteries, determined by appropriate scoring rules.

popular QSR for binary events determines the reward the subject gets and a penalty for error. Assume that θ is the reported probability for R, and that Θ is the true binary-valued outcome for R. Hence $\Theta=1$ if R occurs, and $\Theta=0$ if W occurs. The QSR will pay subjects $S(\theta | R) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(1-\theta)^2$ if event R occurs and $S(\theta | W) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(0-\theta)^2$ if W occurs. Additionally, set the parameters of the QSR to be $\alpha = \beta = 100$.

Since we are using the BLP, subjects earn points in the QSR that give higher chances of winning the binary lottery. If a subject reports θ and event R is realized, he wins $S(\theta | R) = 100 - 100(1-\theta/100)^2$ points. For practical purposes in the experiment subjects can make reports in increments of single percentage points. Similarly, if event W is realized and a subject reports θ , he wins $S(\theta | W) = 100 - 100(0-\theta/100)^2$ points.

Suppose a subject reports $\theta=30$. This implies that he would win 51 points if event R is realized and 91 points if event W is realized. The BLP implies that a subject would then play a binary lottery where the probabilities of winning are defined by the points earned. If the realized event is R, then the individual would play a lottery that pays \$50 with probability 0.51 and \$0 with probability 0.49.⁵ Define $p_R(\theta) = S(\theta | R)/100$ as the objective probability of winning \$50 in the binary lottery induced by the points earned in the scoring rule task when the report is equal to θ and event R is realized. The objective probability $p_W(\theta) = S(\theta | W)/100$ is similarly defined for event W. In the example, $p_R(30) = 0.51$ and $p_W(30) = 0.91$. Figure 1 represents the subjective compound lottery and the actuarially-equivalent simple lottery induced by report $\theta = 30$.

In the QSR a subject may choose any possible Subjective Compound Lotteries (SCL) of the type depicted in Figure 1. In these SCLs, the first stage involves subjective probabilities while the second stage involves objective probabilities defined by the points earned in the first stage.

⁵ These were the actual stakes in our experiment where a subject could win up to \$50 in the task.

If the subject maximizes SEU, and therefore satisfies the Reduction of Compound Lotteries (ROCL) axiom, the valuation of each report θ will be given by

$$\begin{aligned} \text{SEU}(\theta) = & \pi_R \times \{p_R(\theta) \times U(\$50) + (1-p_R(\theta)) \times U(\$0)\} \\ & + (1-\pi_R) \times \{p_W(\theta) \times U(\$50) + (1-p_W(\theta)) \times U(\$0)\} \end{aligned} \quad (1)$$

and the subject chooses the report θ^* that maximizes (1) conditional on the subjective belief π_R .

Because $U(\cdot)$ is unique up to an affine positive transformation under SEU we can normalize

$U(\$50)=1$ and $U(\$0)=0$. Thus the SEU(θ) in (1) can be simplified to

$$\text{SEU}(\theta) = \pi_R \times p_R(\theta) + (1-\pi_R) \times p_W(\theta) = Q(\theta). \quad (1')$$

We rename SEU(θ) as $Q(\theta)$ to emphasize that the subject's valuation of the SCL induced by θ can be interpreted as the subjective average probability $Q(\theta)$ of winning the high \$50 amount in the binary lottery.

One can easily show, using the first order and second order conditions $Q'(\theta)=0$ and $Q''(\theta)<0$, that the report that maximizes (1') is $\theta^* = \pi_R \times 100$, which implies that the QSR combined with the BLP provides incentives to report the true subjective probability directly. The existence of a unique maximum is guaranteed because the function $Q(\cdot)$ is strictly concave, due to the strict concavity of the QSR, because these objective probabilities are a function of the scoring rule.

For comparison, with the QSR payouts defined directly in dollars, denoted by $\$\$S(\theta|\cdot)$, the subject would choose a report θ that maximizes the following valuation

$$\text{SEU}(\theta, U(\cdot)) = \pi_R \times U(\$\$S(\theta|R)) + (1-\pi_R) \times U(\$\$S(\theta|W)). \quad (3)$$

A sufficiently risk averse individual would be drawn to make a report of 50, independent of his subjective probability, because this report provides the same payoff under each event. In other words, the subject has incentives to *smooth out* the potential outcomes.

A proper scoring rule provides incentives to a subject to optimally choose one distinct report equal to her subjective probability. The uniqueness of the optimal report induced by the scoring rule

can be achieved by guaranteeing that the scoring rule induces strict quasi-concavity in the subject's valuation of choices in the belief elicitation task. This principle can be used to show that under certain conditions the BLP can help the QSR to elicit subjective probabilities of subjects with certain non-EUT preferences.

1.2. The Binary Lottery Procedure for Non-Expected Utility Individuals

We can show that a subject who follows the Rank-Dependent Utility (RDU) model will directly report his subjective probability under relatively weak conditions. These conditions are (i) ROCL for binary lotteries, (ii) probabilistic sophistication as defined by Machina and Schmeidler (1992, 1995), (iii) uniqueness of $U(\cdot)$ up to an affine positive transformation and $U(\cdot)$ increasing, (iv) a strictly increasing probability weighting function, and (v) a strictly concave scoring rule. Of course, there are many other ways in which SEU can be violated other than RDU, but RDU is an obvious and important alternative to consider.⁶ SEU is violated under RDU, because the Compound Independence Axiom is violated, but none of the axioms needed for the BLP to induce linear utility are violated provided that the prizes of the BLP are non-stochastic as is customary.

A decision maker with RDU preferences assigns to the higher prize a decision weight $w(p)$, where p is the probability of the higher prize, and the lower prize receives decision weight $1-w(p)$. In our experiment the high prize was \$50 and the low prize was \$0. Notice that binary ROCL implies that the probability weighting is applied to the reduced compound probability $Q(\theta)$, the subjective

⁶ We certainly encourage examination of other non-EUT models. One alternative might be “reference dependent” models of behavior towards risk, such as the Disappointment Aversion model of Gul (1991). Since the reference point for this model is the Certainty-Equivalent, it is not obvious how it would generate any biases in elicited subjective beliefs. Of course, if one is instead free to pick a reference point to explain any observed data, nothing is gained from such models in this application, or in general.

average probability of winning \$50 in the QSR task that uses the BLP.⁷ Therefore the weights for the high and the low prizes in the belief elicitation task are $w(Q(\theta))$ and $(1-w(Q(\theta)))$, respectively.

Consequently, an individual with RDU preferences will have a QSR valuation induced by any report θ of

$$\text{RDU}(\theta) = w(Q(\theta)) \times U(\$50) + (1-w(Q(\theta))) \times U(\$0). \quad (4)$$

Since $U(\cdot)$ is unique up to an affine positive transformation and increasing in the RDU model, we can also normalize $U(\$50)=1$ and $U(\$0)=0$ and the valuation of the individual is reduced to

$$\text{RDU}(\theta) = w(Q(\theta)) \quad (4')$$

An individual with RDU preference will then maximize (4') by choosing an optimal report θ^* that we characterized below. An RDU maximizer and a SEU maximizer, each with subjective probability π_R would have incentives, under certain conditions, to make exactly the same optimal report $\theta^* = \pi_R \times 100$. This follows from the first order condition on the subject's valuation of report θ ,

$$\text{RDU}'(\theta^*) = w'(Q(\theta^*)) \times Q'(\theta^*) = 0.$$

This condition is satisfied when $Q'(\theta^*)=0$, exactly equal to the first order condition of an SEU maximizer, given that $w'(Q(\theta))>0$ is assumed. The solution to the maximization of valuation $\text{RDU}(\theta)$ in (4') is unique because it is assumed that $w(\cdot)$ is strictly increasing and $Q(\cdot)$ is strictly concave, which implies that $w(Q(\cdot))$ is a strictly quasi-concave function.⁸ Therefore the RDU maximizer would optimally make the same report as a SEU maximizer with the same beliefs, and both would have incentives to directly report their true subjective probability. Consequently, the QSR using the BLP is proper also in the case of an RDU maximizer when conditions (i)-(v) are

⁷ The probabilistic sophistication assumption is used here because we are assuming that there exist a probability measure $Q(\theta)$ over the possible outcomes of the scoring rule. This is a plausible assumption since the QSR is designed, after all, to elicit subjective probabilities, and therefore it is implicitly assuming that this probability measure exists.

⁸ This is a simple mathematical fact for which we provide a simple proof in Appendix D.

satisfied.

Hossain and Okui (2013) independently proved a related result with respect to non-expected utility models, but our version of the result gives guidelines on how to choose scoring rules to provide better incentives in belief elicitation tasks. For instance, our analysis suggests that a Linear Scoring Rule (LSR)⁹ cannot be used in conjunction with the BLP if the probability weighting function of an RDU maximizer is strictly increasing. The reason is that the BLP induces risk neutrality over money in subjects and a LSR does not allow one to directly identify subjective probabilities from reports because the optimal report would be either 0 or 100, depending on whether the true latent subjective probability was less than $\frac{1}{2}$ or greater than $\frac{1}{2}$, respectively.¹⁰ Therefore, one has to make sure that $w(Q(\theta))$ is strictly quasiconcave for the scoring rule using the BLP to be proper. A similar conclusion applies with greater strength to the elicitation of beliefs of SEU maximizers¹¹ because the BLP converts them into expected value maximizers, and hence the LSR would only provide incentives to make reports of 0 or 1.

In summary, our analysis implies that appropriate scoring rules have the ability to induce concavity¹² in the subject's belief task valuation, since the SEU and RDU valuation functions of the elicitation task (i.e., $SEU(\theta) = Q(\theta)$ and $RDU(\theta) = w(Q(\theta))$) are functions of the scoring rule

⁹ A Linear Scoring Rule defines the scores for events A and B as $S(\theta|A) = \alpha - \beta |1 - \theta|$ if event A occurs and $S(\theta|B) = \alpha - \beta |0 - \theta|$ if B occurs.

¹⁰ However, the LSR can be used to elicit subjective probabilities when a subject displays risk aversion through a concave utility function. The LSR results in a valuation that is concave in the report if the utility function is concave. Andersen et al. (2013) show how one can infer true subjective probabilities with the LSR if one also knows the risk attitudes of subjects, even when they behave consistently with RDU.

¹¹ Notice they can be equivalently analyzed as RDU maximizers with a linear probability weighting function.

¹² In our arguments we assume strict quasi-concavity. However, one could also assume, the desirable for many applications but stronger, strict concavity which is guaranteed by the following second order condition:

$$RDU''(\theta) < 0.$$

The conclusions would remain the same except that strict quasi-concavity would be replaced by strict concavity. Therefore, in this final paragraph of the section, we loosely use the term concavity to refer to the mathematical definitions of strict quasi-concavity and strict concavity.

themselves. Therefore, one can potentially use this property to find scoring rules that induce enough concavity in the valuation function, and therefore provide more salient incentives for non-trivial belief elicitation problems such as the elicitation of very low subjective probabilities. One of the challenges with low subjective probabilities is that commonly used scoring rules do not provide enough incentives at the corners of the probability interval. One could potentially design scoring rules that sufficiently *concavify* the valuation of the subject such that incentives at these extremes are more powerful.¹³

2. Experiment

2.1. Experimental Design

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 before making choices.

Table 1 summarizes our experimental design and sample sizes.¹⁴

A critical element of our experimental design was for subjects in each treatment to see the same observable physical stimuli (i.e., the compositions of red and white balls in a bingo cage).

¹³ More generally, our theoretical treatment suggests that belief elicitation with scoring rules can be studied from a functional analysis point of view. This could potentially be an interesting avenue of research. For instance, one could study the general existence of proper scoring rules for a particular type of preference representation. More pragmatically, one could study scoring rules that provide salient incentives to deal with non-trivial belief elicitation problems such as low probability elicitation.

¹⁴ Given concerns about the Random Lottery Incentive Method, documented in Cox et al. (2011) and Harrison and Swarthout (2012), each subject was presented with only one belief elicitation decision to make, and subjects made no other decision in the experiment. In brief, these concerns are that the payment protocol in which one asks the subject to make $K > 1$ choices, and pick 1 of the K at random for payment at the end, requires that the Independence axiom apply. But then one cannot use those data to estimate models of decision-making, such as RDU, that assume the invalidity of that axiom. The only reliable payment protocol in this case is to ask subjects to only make one choice, and pay them for it.

Since the stimulus shown in a given session was the outcome of a random event, we could not control the stimuli across sessions. Instead, we split subjects in a given session into each treatment, so that the session-specific stimulus was shown to subjects in each treatment.

Subjects were seated randomly within the laboratory at numbered stations, signed the informed consent document, and were given printed introductory instructions.¹⁵ A Verifier was randomly selected for the sole purpose of verifying that the procedures of the experiment were carried out according to the instructions, and was paid only a fixed amount for this task.

Each subject was assigned to one of two treatments depending on whether her seat number was even or odd.¹⁶ Treatment groups took turns to go out of the laboratory under experimenter supervision, and waited for the other group to go over their treatment-specific instructions. Subjects waiting outside the laboratory had the chance to read the instructions individually. This procedure avoids subject confusion that may arise from hearing the other treatment's instructions. Once all instructions were finished, we proceeded with the remainder of the experiment with all subjects from both treatments in the laboratory together.

We implemented two treatments. In **treatment M** subjects are presented with only one belief elicitation question using the QSR with monetary scores. In **treatment P** subjects are also presented with only one belief elicitation question using the QSR, but the scores are denominated in points that subsequently determined the objective probability of winning a binary lottery. So treatment P implements the QSR with the BLP, and treatment M is the control.¹⁷

¹⁵ Complete subject instructions are provided in Appendix A.

¹⁶ We assign subjects to treatments according to their station number in the laboratory to avoid potential confounds due to subjects in each treatment having very different vantage points from which to observe the stimulus.

¹⁷ A referee also suggests an additional control, in which subjects play for points but these are converted at the end of the experiment using some fixed, announced exchange rate. This suggestion has the advantage of controlling for any possible framing effects from using points, as distinct from the joint effect of using points and the binary lottery procedure for converting points into money.

We used two bingo cages: Bingo Cage 1 and Bingo Cage 2. Bingo Cage 1 was loaded with balls numbered 1 to 99 in front of everyone.¹⁸ A numbered ball was drawn from Bingo Cage 1, but the draw took place behind a divider. The outcome of this draw was not verified in front of subjects until the very end of the experiment, after their decisions had been made. The number on the chosen ball from Bingo Cage 1 was used to construct Bingo Cage 2 behind the divider. The total number of balls in Bingo Cage 2 was always 100: the number of red balls matched the number drawn from Bingo Cage 1, and the number of white balls was 100 minus the number of red balls. Since the actual composition of Bingo Cage 2 was only revealed and verified in front of everybody at the end of the experiment, the Verifier's role was to confirm that the experimenter constructed Bingo Cage 2 according to the randomly chosen numbered ball. Once Bingo Cage 2 was constructed, the experimenter put the chosen numbered ball in an envelope and affixed it to the front wall of the laboratory, in full view of all subjects at all times.

Bingo Cage 2 was covered and placed on a platform in the front of the room. Bingo Cage 2 was uncovered for subjects to see, spun for 10 turns, and covered again. Subjects then made their decisions about the number of red and white balls in Bingo Cage 2. After choices were made and subjects completed a non-salient demographic survey, the experimenter drew a ball from Bingo Cage 2. Then the sealed envelope was opened and the chosen numbered ball was shown to everyone, and the experimenter publicly counted the number of red and white balls in Bingo Cage 2.

The final step during the session was to determine individual earnings. An experimenter approached each subject and recorded earnings according to the betting choices made and the ball drawn from Bingo Cage 2. If subjects were part of treatment M their earnings were directly

¹⁸ When shown in public, Bingo Cage 1 and 2 were always displayed in front of the laboratory where everyone could see them. We also used a high resolution video camera to display the bingo cages on three flat screen TVs distributed throughout the laboratory, and on the projection screen at the front of the room. Our intention was for everyone to have a roughly equivalent view of the bingo cages.

determined by the report and the corresponding score in dollars of the QSR. If subjects were in treatment P the number of points they earned in the belief elicitation task was recorded. The subjects in treatment P then rolled two 10-sided dice, and if the outcome was less or equal to the number of points earned they won \$50, otherwise they earned \$0. Finally, all subjects left the laboratory and were privately paid their earnings, and also a \$7.50 participation payment. Each verifier was paid \$25 plus the participation payment. Subjects earned \$45.60 on average, including the participation payment.

Several of our procedures are designed specifically to avoid trust issues the subjects may have with the experimenter, which can be source of significant potential confounds in subjective belief elicitation tasks. Our random selection of a Verifier makes it transparent to the subjects that any one of them could have been selected, and thus we are not employing a confederate. Further, by taking the time at the end of the session to publicly verify the previous private random draws, we are able to more credibly emphasize in the instructions that any composition of Bingo Cage 2 was equally likely, thus minimizing any prior beliefs that particular compositions may be more likely.

We used software we created to present the QSR to subjects and record their choices. Figures 2 and 3 illustrate the scoring rule task faced by subjects in treatments M and P, respectively, which are variants on the “slider interface” proposed by Andersen et al. (2014). Subjects can move one or other of two sliders, and the other slider changes automatically so that 100 tokens are allocated. The main difference between Figures 2 and 3 is that the payoffs of the scoring rule are denominated in dollars in the case of Figure 2, and determined in points in Figure 3. Subjects can earn up to \$50 in treatment M and either \$50 or \$0 in treatment P.

A natural question to ask is whether subjects understood the BLP in treatment P? Of course, this question can and should be asked for any new procedure, particularly if the BLP has historically been “under suspicion” by experimental economists. In large measure we regard the tests of the

BLP with *objective* probabilities, reported in Harrison et al. (2013), as convincing evidence that it was understood in general by our population of subjects. Indeed, we would not have considered its application to subjective probabilities if it had not passed those tests with objective probabilities. Some experimenters have de-briefing questions after the task, as well as some test questions at the outset. We are not confident that any hypothetical questions after the session could reliably recover understanding, rather than subjects telling the experimenter what they think the experimenter wants to hear, and prefer to use hypotheses applied to incentivized data as the test of understanding. We were also concerned that adding some test questions about the BLP prior to the task might have an effect on behavior itself, just as one always wants to avoid instructions that guide subjects to behave in a certain way.

2.2. Evaluation of Hypotheses

We want to test if the BLP induces linear utility, providing incentives for subjects to report truthfully and directly their underlying subjective probabilities. In our tests we assume homogeneity of risk attitudes and subjective probabilities across subjects in the two treatments. Therefore any observed difference in reports would be a result of BLP affecting subjects' behavior.¹⁹ We have three ways of testing our hypothesis: the first two are non-parametric statistical tests designed to find

¹⁹ We developed our experimental procedures to avoid as much as possible perceptual confounds and to minimize perceptual differences across treatments. In each session we were careful to have roughly a 50/50 distribution among our two treatments, so our procedure of odd/even set number assignments would result in an alternating seating arrangement with an M-subject beside a P-subject, who was in turn seated beside another M-subject, and so on. Our computer stations have high division panels that prevent a subject looking at another subject's computer. Also, we had enough high-resolution displays such that no matter where one was seated in the laboratory, there was a clear view of the Bingo cages. These measures were taken to ensure that the experimental protocol did not create any perceptual advantage to any treatment and that if there is any difference in subject's vantage point then this difference was common across treatments. However, we do have to assume that if there is heterogeneity in perceptual abilities that can create noise in the reports, then those abilities are similarly distributed in both treatments. We believe that this is a reasonable assumption since all subjects are randomly sampled from the same population and subjects are then randomly assigned to each treatment.

treatment effects, and the third is a structural econometric approach that recovers the underlying subjective probabilities of M subjects that are then compared with the raw average reports of P subjects.

Hypothesis 1: *If subjects are risk averse and the BLP induces linear utility, then subjects in treatment M make reports closer to 50, on average, than subjects in treatment P.* To evaluate this hypothesis we need not know anything about the underlying subjective probabilities of the subjects. Rather, we need only to know that they are risk averse. And indeed, our subject population is risk averse.²⁰ A risk averse subject, independent of her subjective probability, will be drawn to make reports closer to 50 than a risk neutral individual with the same subjective belief. We test this hypothesis by calculating the absolute value of the distance between each report and 50. If the underlying subjective probability is close to 50%, there would be an identification problem because subjects in both treatments have strong incentives to make a report close to 50. This is likely in situations where the Bingo cage composition of red and white balls is close to 50:50, which was indeed the case in one of our sessions. Similarly, if the underlying subjective probability is close to 0% or 100% we would also have an identification issue. This was a risk that we had to take in order to ensure transparency of the process generating the random stimuli.

Hypothesis 2: *If subjects are risk averse and the BLP induces linear utility, then subjects in treatment P make reports that are closer, on average, to the true number of red balls in Bingo Cage 2 than subjects in treatment M.* To evaluate this test we again employ a measure of distance but, instead of using 50 as a point of reference, we use the true number of red balls in the Bingo Cage 2 each subject faced. We assume

²⁰ Evidence from previous experiments with subjects from the same population, such as Holt and Laury (2002) and Harrison et al. (2013), show that the subject pool at Georgia State University is indeed risk averse, on average, over money. Therefore, we use this knowledge to hypothesize how subjects behave in each treatment. Similarly, Hypothesis 2 also takes advantage of what we know about our subjects' risk attitudes.

this same value, which we know as experimenters, as a proxy for the underlying subjective probability. The comparison across treatments of this measure of distance, which we call report distance, also provides a test of the relative accuracy of reports. Even though it is interesting in its own right, it is not our primary objective to assess the perceptual accuracy of subjects.²¹

An ideal test of hypothesis 2 would involve comparing the reports in treatment P with the underlying subjective probabilities of subjects in treatment M, and we do this in the following structural econometric test.

Hypothesis 3: *If the BLP induces linear utility then the risk-attitudes-adjusted subjective probabilities in the M treatment are equal to the raw average report of subjects in treatment P.* If BLP induces linear utility in the P subjects, they should directly report their true underlying subjective probability, which should in turn be equal to the underlying estimated probabilities for M subjects. We estimate a structural model, both assuming EUT and RDU preferences, that *jointly* estimates risk attitudes and the underlying subjective probabilities of subjects in treatment M, and then we compare those estimated subjective probabilities with the raw average reports in treatment P. We pool data across sessions, but also analyze data from each individual session.

3. Results

3.1. Hypothesis 1: The BLP Mitigates the Effects of Risk Aversion.

We find evidence of a treatment effect which supports the hypothesis that the BLP induces linear utility in our belief elicitation tasks. Pooling across sessions, there were 68 subjects in treatment M and 70 in treatment P. Figure 4 shows the frequency of reports in each treatment, by

²¹ In fact there might be some visual saliency of red balls that might have induced subjects to make reports higher than if we had used balls of different colors.

session. Figure 5 displays, again by session and for illustrative purposes, the estimated densities of the reports, the correct number of red balls in Bingo Cage 2, and the mean report in each treatment. In session 1 the average reports for treatments M and P are 34.2 and 30.8, respectively. Excluding one subject who gave an idiosyncratic report of 100 red balls, the average report for treatment P in session 1 is decreased to 26.8.²² The average reports from treatments M and P are 59.7 and 65.8, respectively, for session 2. The panels for sessions 1 and 2 in Figure 5 are illustrative, although not statistical evidence yet, of a treatment effect consistent with risk aversion: the mass of the estimated densities for treatment M is closer to the middle of the report interval than for the case of treatment P. This feature is not readily seen in the case of sessions 3 and 4, and that is why we present below non-parametric statistics to test the statistical significance of this treatment effect across sessions.

In support of Hypothesis 1, we find evidence that subjects in treatment M tend to make reports closer to 50 than subjects in treatment P. Across all sessions, the average of the distance reports is 14.2 and 18.7 for treatments M and P, respectively. On average subjects in treatment M tend to make reports closer to 50 as suggested by the p -value of 0.02 of the one-sided Fisher-Pitman permutation test. We present non-parametric test results with and without session 3 since the ratio of red and white balls in Bingo Cage 2 was randomly chosen close to 50:50, precisely where we predict this treatment effect test would have low power. Figure 6 shows the cumulative distribution of our measure of distance of reports from 50 for sessions 1, 2 and 4 and for all sessions pooled. When we exclude session 3, we find evidence that subjects in treatment M tend to provide reports closer to 50 as suggested by a p -value of 0.02 of the one-sided (directional) Kolmogorov-Smirnov test.²³

²² This subject's reporting behavior was certainly puzzling and idiosyncratic, but can be rationalized by non-EU preferences. A simpler explanation is that the subject had a strong preference for the color red.

²³ When we pool all the sessions, the p -value increases to 0.23, which was expected given that this test of the hypothesis has low power in cases where the composition of the Bingo cage is close to 50:50.

3.2. Hypothesis 2: The BLP Improves Accuracy.

Subjects from treatment P tend to make reports closer to the correct number of red balls in Bingo Cage 2. This could be interpreted as better accuracy from the part of subjects in treatment P. However, since our experiment was designed to avoid creating any perceptual advantages for either treatment, there are no *a priori* reasons to believe that P subjects were superior at identifying the right number of red balls. Thus we interpret this result as evidence that the BLP induces linear utility in subjects and provides incentives to reveal directly the true latent subjective probability, thus mitigating the distortion in reports introduced by risk attitudes.

There were 68 subjects in treatment M whose average report distance was 15.2, while there were 70 subjects in treatment P whose average report distance was 12.8. Excluding the subject in session 1 that idiosyncratically reported 100 red balls when the stimuli was 17 red balls, the average report distance for treatment P was instead 11.8. On average subjects in treatment P tend to make reports closer to the actual number of red balls in the Bingo cage as suggested by the p -value of 0.02 of the one-sided Fisher-Pitman permutation test. Figure 7 shows the distribution of the absolute value of differences between reports and the correct number of red balls, pooling over all sessions. The cumulative distribution of treatment P is dominated by the distribution of treatment M, which implies that distances are smaller in treatment P. This is supported by a p -value of 0.04 of the one-sided (directional) Kolmogorov-Smirnov test, which is consistent with the Fisher-Pitman permutation test.

3.3. Hypothesis 3: Subjects Report Underlying Subjective Probabilities with the BLP.

To reinforce the above results, and as a robustness check, we estimate the subjective probabilities of M subjects and compare them to the raw reports of the P subjects both assuming EUT and RDU, which should be equal if BLP works as suggested by theory.

3.3.1. Subjective Expected Utility Case

Following Andersen et al. (2013), we develop a structural econometric model to estimate the underlying subjective probabilities of subjects in the M treatment and compare these estimates with the raw reports of P subjects. We find that after controlling for risk attitudes, the adjustment on probability reports in the M treatment is in the expected direction (closer to the mean reports in treatment P), and we cannot reject the hypothesis that the underlying subjective probabilities of M subjects are equal to the mean reports of P subjects.

The objective of the structural estimation is to jointly estimate risk attitudes and the underlying subjective probabilities in the M treatment. Responses from the M treatment in the belief elicitation task identify the subjective probabilities. Since we do not collect lottery choices under risk from these subjects, we cannot identify their risk attitudes. Instead, we pool observations from prior experiments²⁴ with different subjects to identify risk attitudes. These pooled observations were sampled from the same subject population and in the same manner as the present study.²⁵

Conditional on EUT and the assumption of a CRRA utility function being the model that characterizes individual decision under risk in our sample,²⁶ we maximize the *joint* likelihood of

²⁴ We use choices from two other experiments, reported in Harrison and Swarthout (2012) and Harrison et al. (2012), that collect responses to binary choices between lotteries with objective probabilities. Each subject in these two studies made one, and only one, choice and was paid for it. Subjects in all tasks were sampled from the same population and there were 160 observations. Payoffs were roughly the same.

²⁵ We use the same subject pool and sampling procedures for all experiments under discussion, resulting in samples with very similar characteristics. For instance, the female to male ratio was 61:39 in the the risk attitudes experiments and the ratio was 62:38 in the present belief elicitation experiment. This is to be expected, since each session was recruited as follows: 1. a random sample was drawn from a database of over 2,000 Georgia State University undergraduate students who had previously registered in the web-based recruiting system; 2. an invitation email was sent to each student in the random sample; and 3. the first 40 students who logged into the system and confirmed their availability were the people ultimately selected to participate in the given session.

²⁶ Our objective is simply to find a way of characterizing risk attitudes to illustrate how the estimated and risk-attitudes-adjusted subjective probabilities in the M treatment compare to the average raw elicited reports in the P treatment. We can therefore remain agnostic as to the “true” model of behavior towards risk. Nevertheless, we also estimate the structural model in the next subsection assuming that risk attitudes are characterized by the RDU model.

observed choices in the risk task and the belief elicitation task in the treatment M by pooling the observations. The solution to this maximization yields estimates of risk attitudes and subjective probabilities that best explain observed choices in the belief elicitation task as well as observed choices in the lottery tasks.²⁷ We are mainly concerned with inferences about *average* risk attitudes and *average* subjective probabilities, therefore we present results assuming homogeneity across subjects in both dimensions.²⁸ Nevertheless, we can control for demographic characteristics to provide a richer characterization of risk attitudes and subjective probabilities, and we obtain virtually the same results as in the homogenous case.

In sessions 1 and 4 the estimated subjective probabilities were, respectively, equal to 30.2% with a p -value of 0.011 and 70.7% with p -value less than 0.001.²⁹ We conclude that we cannot reject the hypothesis that each of these *estimates* is equal to the average *raw* reports of P subjects. In session 1 (4) a test for the null hypothesis that the probability estimate is equal to the average report of P subjects of 26.8% (75.8%) results in a p -value equal to 0.77 (0.66). Similarly, in session 2 a test for the hypotheses that the estimated subjective probability of 62.1% for M subjects is equal to the P subjects' average report of 65.8 P results in a p -value equal to 0.74. Finally, the p -value for the

²⁷ Appendix B provides a more detailed explanation of the estimation procedures and Appendix C shows the estimation results. We estimate two models, one for sessions 1 and 4 and another for sessions 2 and 3. In sessions 1 and 4 the stimuli were closer to 0 and 100, respectively, while in sessions 2 and 3 the stimuli were clearly closer to 50. Assuming homogenous preferences and EUT preferences, the estimated risk aversion parameter was virtually the same in both models and equal to 0.61 with p -values on the null hypothesis of risk-neutrality less than 0.001.

²⁸ It is not possible to estimate risk attitudes for an individual and then use those estimates to condition the inferences about subjective probabilities of the same individual. First, the same individual did not participate in both the risk aversion and subjective probabilities tasks. Second, each individual only made a single choice in each, so there are no “degrees of freedom” to estimate for that individual alone. We must use estimation for the pooled sample. This is a deliberate design choice, given our theoretical and behavioral concerns with the Random Lottery Incentive Method, and not something we decided to do after the fact.

²⁹ For the estimation we drop two subjects from session 1, where the number of red balls in the Bingo Cage was 17. One of the subjects was from the M treatment who made a report of 60, and the other was from the P treatment and made a report of 100. Our overall conclusions are not affected by dropping these outliers.

equivalent null hypothesis for session 3 is equal to 0.84; however, although consistent with our overall conclusions, the choices from this session are not particularly informative because subjects in both treatments had strong incentives to make a report of 50 given that the stimulus was very close to this number.

3.3.2. Rank-Dependent Utility Case

If risk attitudes are characterized by the RDU model there is probability weighting, we showed earlier that the BLP should still induce truthful revelation under some weak conditions. The reason is that when the BLP is applied, the RDU model collapses to the Dual Theory model of Yaari (1987), since utility is linear in the monetary payoffs of the binary lottery. Hence the responses in the P treatment will, in theory, generate directly the true, latent subjective probability. When we risk-calibrate the responses in the M treatment, however, we need to correct for probability weighting and for the utility curvature. We estimate this model assuming a CRRA utility function and the probability weighting function popularized by Tversky and Kahneman (1992), with curvature parameter γ , $w(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma}$.³⁰

We cannot reject the hypothesis that the *estimated subjective probability* for each session is equal to the session's average raw report of the P subjects. In sessions 1 and 2 the estimated subjective probabilities are, respectively, equal to 22.3% with a p -value of 0.077 and 64.9% with p -value less than 0.001. Meanwhile, in sessions 3 and 4 the subjective probabilities are, respectively, 52.5% and

³⁰ As required by the BLP to work under non-EUT preferences, this probability weighting function is strictly increasing for the levels of the parameter γ observed in the laboratory. In fact, this function is strictly increasing in the domain of the probability interval and of parameter γ except for a small region in which γ is roughly below 0.28. Our estimates of γ are outside this region. When we pool sessions 1 and 2, the estimates for the utility and the probability weighting function parameters were equal to 0.26 (p -value = 0.12) and 0.49 (p -value < 0.001), respectively. For sessions 3 and 4, the same estimates were equal to 0.25 (p -value = 0.19) and 0.47 (p -value < 0.001), respectively.

81.5% with p -values less or equal to 0.001 . In session 1 (2) a test of the null hypothesis that the probability estimate is equal to the average report of P subjects of 26.8% (65.8%) results in a p -value equal to 0.72 (0.95). Similarly, in session 4 a test of the hypotheses that the estimated subjective probability of 81.5% for M subjects is equal to the P subjects' average report of 75.8% results in a p -value equal to 0.64. Finally, the p -value for the equivalent null hypothesis for session 3 is equal to 0.94. These results emphasize that the BLP is robust to certain violations of the axioms of SEU.

4. Conclusions

The BLP has been shown by Harrison et al. (2013) to work robustly to induce risk neutrality when subjects are given one risk task defined over *objective* probabilities. Motivated by the success of the BLP in that setting, and using individuals sampled from the same pool of subjects, we find evidence that the BLP also induces linear utility in a *subjective* probability elicitation task when using the popular Quadratic Scoring Rule.

An important feature of the BLP in this application is that it theoretically provides incentives for subjects to directly report underlying latent subjective probabilities. This applies for individuals with SEU preference representations and, under certain weak conditions, for individuals with RDU preference representations as well.

Figure 1: Binary Scoring Rule Using the Binary Lottery Procedure

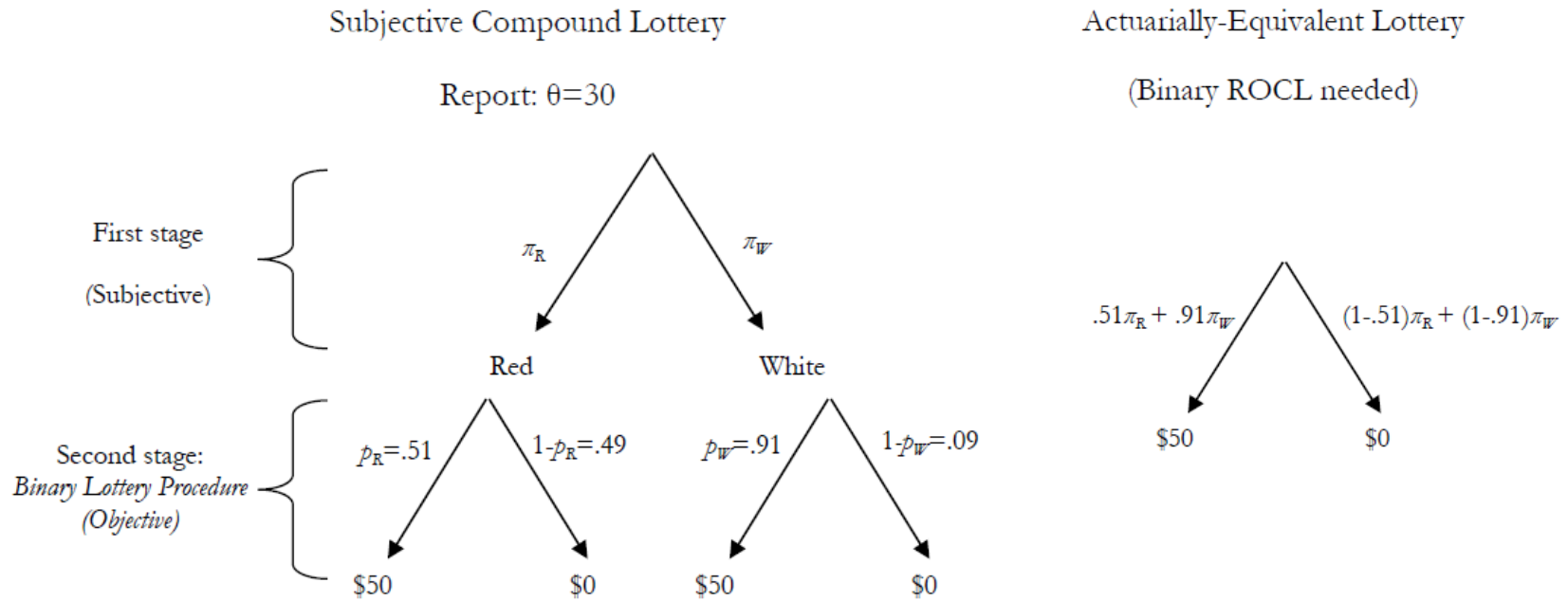


Table 1: Experimental Design

	Treatment M	Treatment P	Total
Session 1	17	18	35
Session 2	16	17	33
Session 3	17	16	33
Session 4	18	19	37
Total	68	70	138

Figure 2: Subject Display for Treatment M

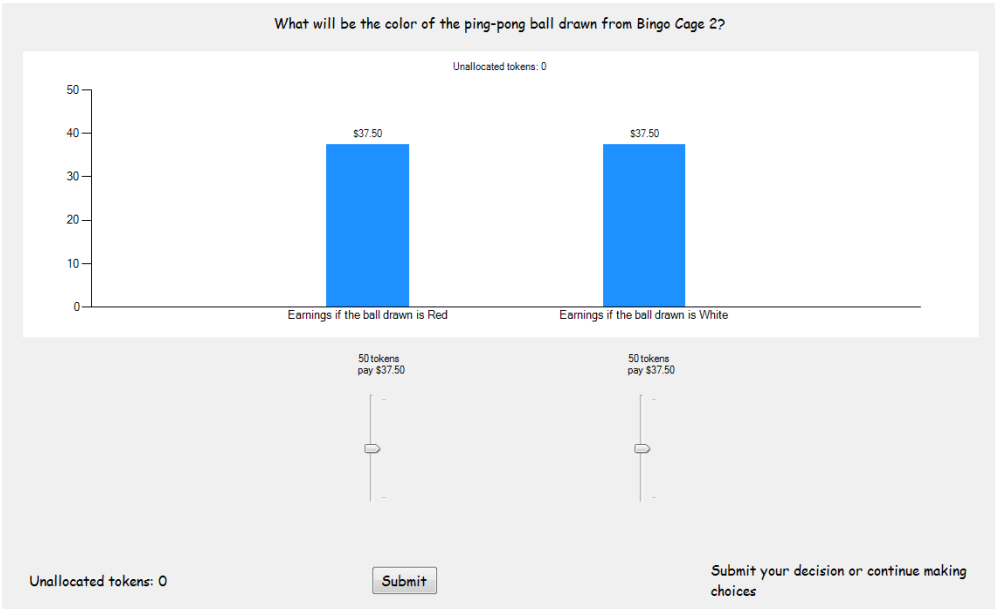


Figure 3: Subject Display for Treatment P

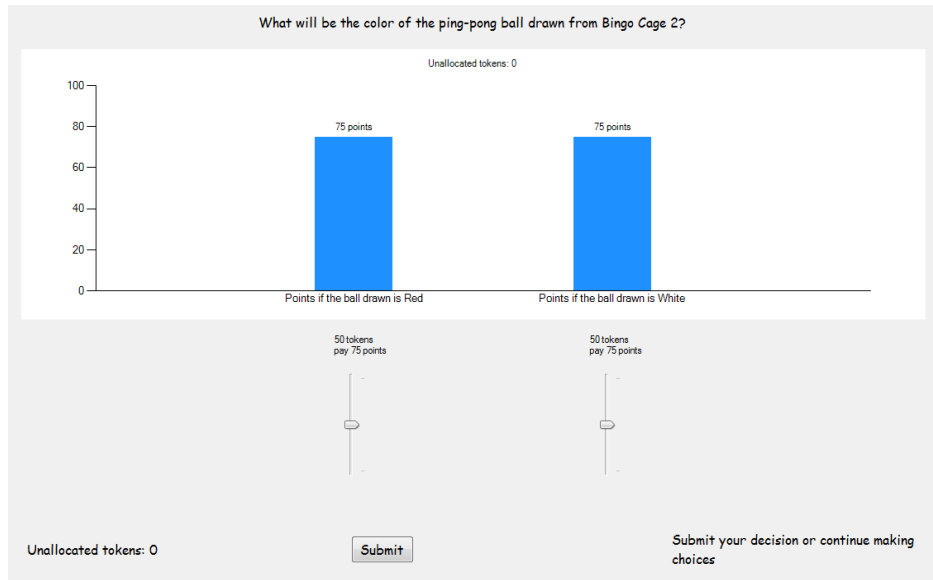


Figure 4: Frequency of Reports by Session

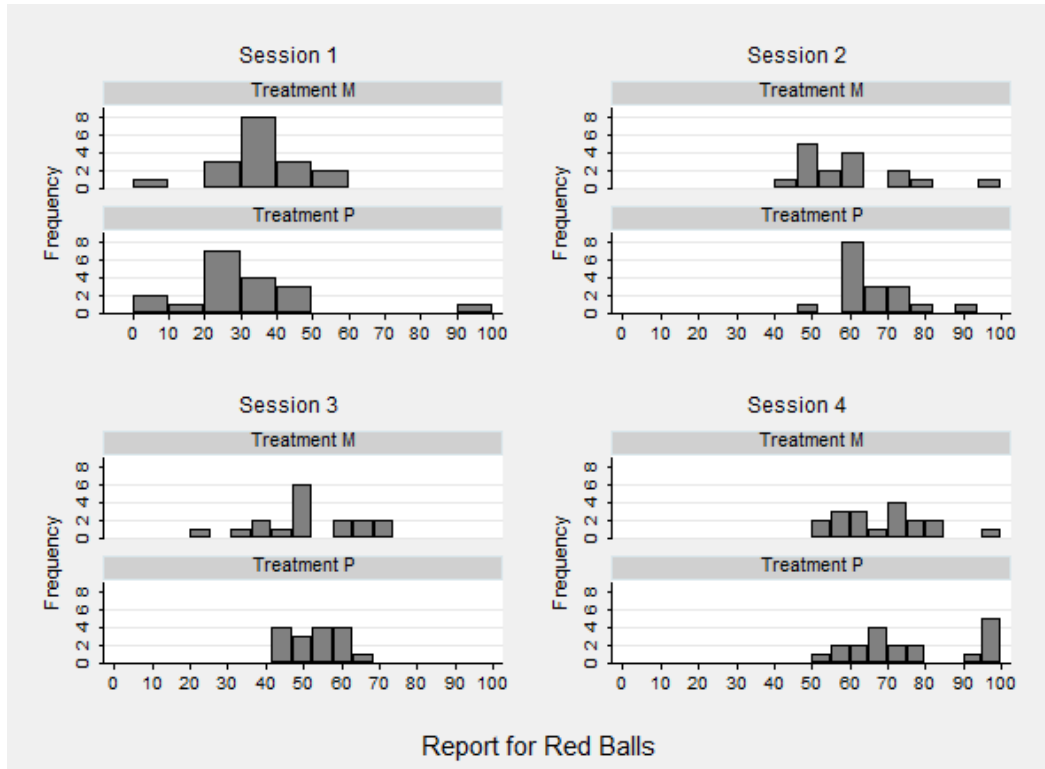


Figure 5: Estimated Densities of Reports by Session

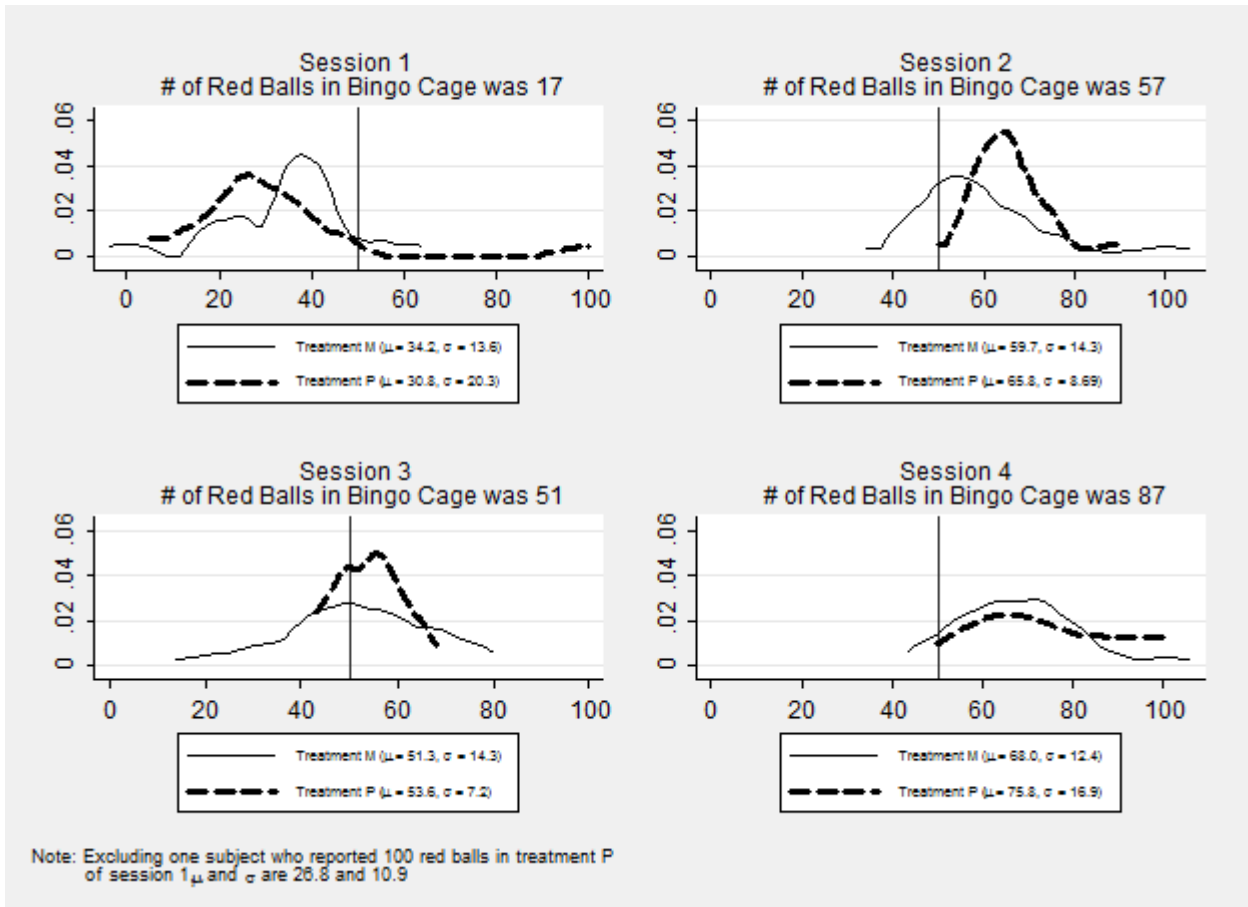


Figure 6: Empirical Cumulative Distribution of Distance of Reports from 50

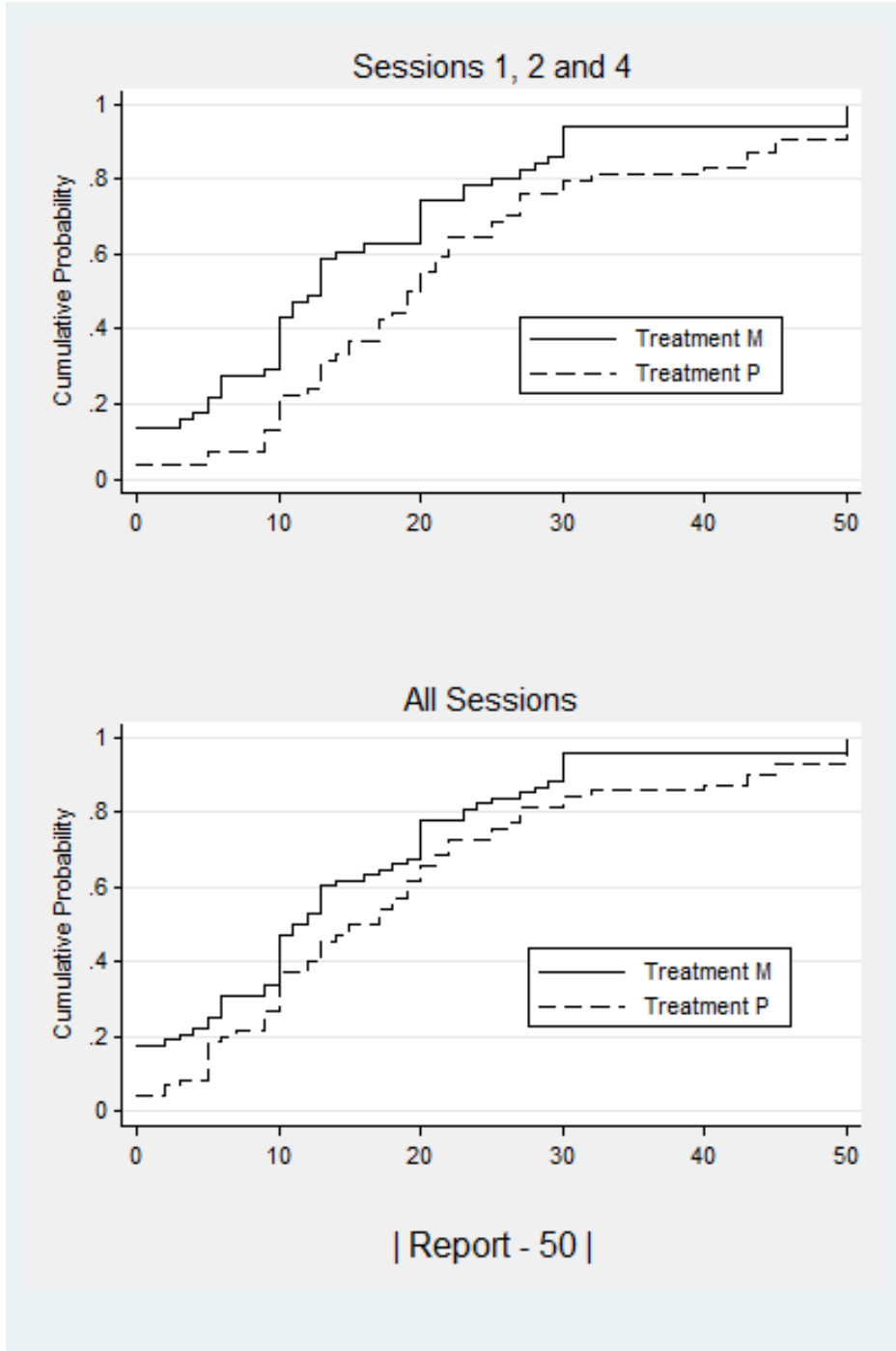
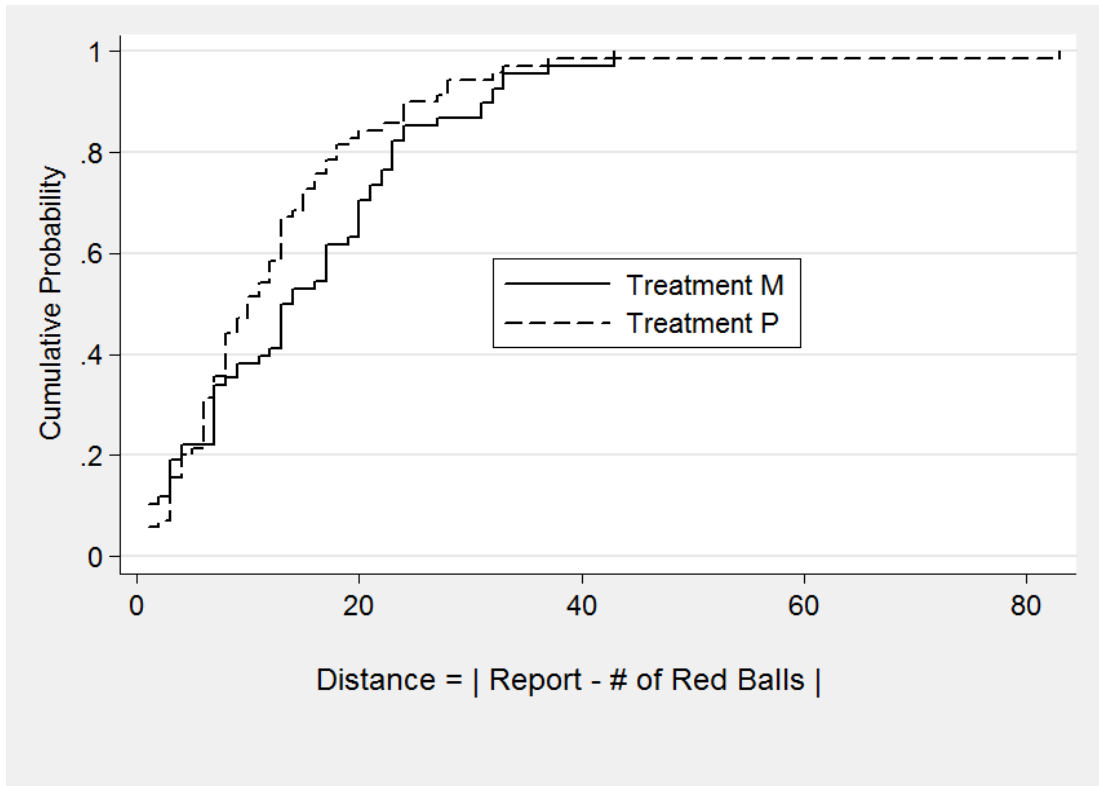


Figure 7: Empirical Cumulative Distribution of Distance Pooling Data From All Sessions



References

- Allen, F., 1987. Discovering personal probabilities when utility functions are unknown. *Management Science* 33, 452-454.
- Andersen, S., Fountain, J., Harrison, G. W., Rutström, E. E., 2014. Estimating subjective probabilities. *Journal of Risk and Uncertainty*, forthcoming.
- Berg, J. E., Daley, L. A., Dickhaut, J. W., O'Brien, J. R., 1986. Controlling preferences for lotteries on units of experimental exchange. *Quarterly Journal of Economics* 101, 281-306.
- Cox, J. C., Oaxaca, R. L., 1995. Inducing risk-neutral preferences: Further analysis of the data. *Journal of Risk and Uncertainty* 11, 65-79.
- Cox, J. C., Sadiraj, V., Schmidt, U., 2011. Paradoxes and mechanisms for choice under risk. Working Paper 2011-12, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.
- Grether, D. M., 1992. Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization* 17, 31-57.
- Gul, F., 1991. A theory of disappointment aversion. *Econometrica* 59, 667-686.
- Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., 2012. Reduction of compound lotteries with objective probabilities: Theory and evidence. Working Paper 2012-05, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.
- Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., 2013. Inducing risk neutral preferences with binary lotteries: A reconsideration. *Journal of Economic Behavior and Organization* 94, 145-159.
- Harrison, G. W., Swarthout, J. T., 2012. Experimental payment protocols and the bipolar behaviorist. Working Paper 2012-01, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.
- Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. *American Economic Review* 92, 1644-1655.
- Holt, C. A., Smith, A. M., 2009. An update on Bayesian updating. *Journal of Economic Behavior and Organization* 69, 125-134.
- Hossain, T., Okui, R., 2013. The binarized scoring rule. *Review of Economic Studies* 80, 984-1001.

- Karni, E., 2009. A mechanism for eliciting probabilities. *Econometrica* 77, 603-606.
- Kőszegi, B., Rabin, M., 2008. Revealed mistakes and revealed preferences. In: Caplin, A., Schotter, A. (Eds.). *The Foundations of Positive and Normative Economics: A Handbook*. New York: Oxford University Press, 193-209.
- Machina, M. J., Schmeidler, D., 1992. A More robust definition of subjective probability. *Econometrica* 60, 745-780.
- Machina, M. J., Schmeidler, D., 1995. Bayes without bernoulli: Simple conditions for probabilistically sophisticated choice. *Journal of Economic Theory* 67, 106-128.
- McKelvey, R. D., Page, T., 1990. Public and private information: An experimental study of information pooling. *Econometrica* 58, 1321-1339.
- Offerman, T., Sonnemans, J., van de Kuilen, G., Wakker, P. P., 2009. A truth-serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *Review of Economic Studies* 76, 1461-1489.
- Roth, A. E., Malouf, M. W. K., 1979. Game-theoretic models and the role of information in bargaining. *Psychological Review* 86, 574-594.
- Rutström, E. E., Wilcox, N. T., 2009. Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior* 67, 616-632.
- Savage, L. J., 1954. *The Foundations of Statistics*. New York: John Wiley and Sons.
- Schlag, K. H., van der Weele, J., 2013. Eliciting probabilities, means, medians, variances and covariances without assuming risk-neutrality. *Theoretical Economics Letters* 3, 38-42.
- Selten, R., Sadrieh, A., Abbink, K., 1999. Money does not induce risk neutral behavior, but binary lotteries do even worse. *Theory and Decision* 46, 211-249.
- Smith, C. A.B., 1961. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society* 23, 1-25.
- Trautmann, S. T., van de Kuilen, G., 2011. Belief elicitation: A horse race among truth serums. *Center Working Paper 2011-117*, Tilburg University.
- Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297-323.
- Yaari, M. E., 1987. The dual theory of choice under risk. *Econometrica* 55, 95-115.

Appendix A: Instructions

A. Introductory Instructions

sub-rocl

Introductory Instructions

You are now participating in a decision-making experiment. Based on your decisions in this experiment, you can earn money that will be paid to you in cash today. It is important that you understand all instructions before making your choices in this experiment.

Please turn to silent, and put away, your cell phone, laptop computer, or any other device you may have brought with you. Please do not talk with others seated nearby for the duration of the experiment. If at any point you have a question, please raise your hand and we will answer you as soon as possible.

The experiment consists of one decision-making task and a demographics survey. You have already earned \$7.50 for agreeing to participate in the experiment, which will be paid in cash at the end of the session. In addition to this show-up fee, you may earn considerably more from your choices in the decision-making task. This task and the potential earnings from it will be explained in detail as we proceed through the session.

This experiment requires us to do some things out of your sight. However, at the end of the experiment we will prove to you that we followed the procedures described in the instructions. Additionally, we will select one of you at random solely for the purpose of verifying that the steps of this experiment are done exactly as described in the instructions. As we proceed in the experiment, we will outline clearly the steps that this Verifier has to verify. In a moment we will select the Verifier by drawing a random number and matching the outcome with the appropriate seat number. The Verifier will be paid \$25 for this job on top of the \$7.50 show-up fee, and will not make any decisions in the experiment. The Verifier will join the experimenter, observe the procedures, and confirm that we are following the procedures explained in these instructions. The Verifier must not communicate with anyone in the room except the experimenter. Failure to do so will result in that person losing the promised amount, another person being chosen as Verifier, and a restart of the experiment.

Part of this experiment is to test different computer screens. Therefore, we will divide you into two groups, and each group will be presented with a slightly different instructions and computer screens. If you are sitting in a computer station that has an odd number on it, you are part of the Odd group. If you are sitting in a computer station that has an even number on it, you are part of the Even group.

Once the Verifier is chosen and joins the experimenter at the front of the room, we will hand out the rest of the instructions. We will then have one of the two groups leave the room for a few minutes, so that an experimenter can read the instructions aloud to the remaining group and answer any questions if necessary. Then the groups will swap places and an experimenter will read instructions to the other group and answer any questions if necessary. There will always be some

experimenters guiding you to get in or out of the room at the right moment.

Once all instructions are finished, and both groups are together in the room again, we will proceed with the experiment. Please remain silent during the experiment, and simply raise your hand if you have any question so that an experimenter will come to you.

Your Beliefs

This is a task where you will be paid according to how accurate your beliefs are about certain things. You will be presented with one and only one question of the type we will explain below. You will actually get the chance to play the question presented to you, so you should think carefully about your answer to the question.

You will make decisions about the color of a ball to be drawn from a bingo cage. This bingo cage will contain 100 balls colored red and white. The exact mix of red and white balls will be unknown to you, but you will receive information about the mixture. The following instructions explain in more detail how this experiment will work.

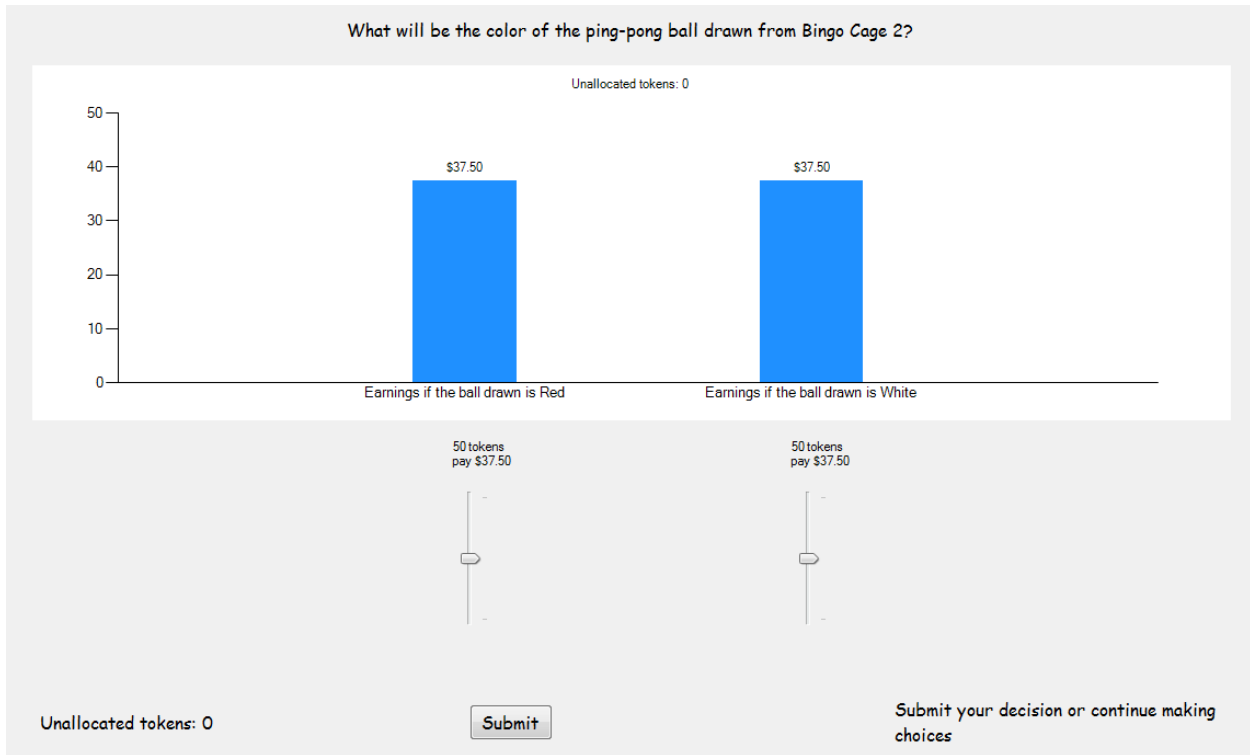
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab except the experimenter. Failure to do so will result in that person losing the promised amount, another person being chosen as verifier, and a restart of the experiment.

We have two bingo cages: Bingo Cage 1 and Bingo Cage 2. We will load Bingo Cage 1 with balls numbered 1 to 99. You will watch us do this, and be able to verify yourself that Bingo Cage 1 is loaded with the correct numbered balls. We will then draw a numbered ball from Bingo Cage 1. However, the draw of a numbered ball from Bingo Cage 1 will take place behind a divider, and you will not know the outcome of this draw until the very end of the experiment, after you have made your decisions. Any number between 1 and 99 is equally likely.

The number on the chosen ball from Bingo Cage 1 will be used to construct Bingo Cage 2 behind the divider. The total number of balls in Bingo Cage 2 will always be 100: the number of red balls will match the number drawn from Bingo Cage 1, and the number of white balls will be 100 minus the number of red balls. Since the actual composition of the Bingo Cage 2 will only be revealed and verified in front of everybody at the end of the experiment, the Verifier will confirm that the experimenter constructs Bingo Cage 2 according to the randomly chosen numbered ball. Once Bingo Cage 2 is constructed, the experimenter will put the chosen numbered ball in an envelope and affix it to the front wall above the white board.

Next, Bingo Cage 2 will be covered and placed on the platform in the front of the lab. Then, Bingo Cage 2 will be uncovered for you to see and spun for 10 turns. After this, we will again cover Bingo Cage 2. You will then make your decisions about the number of red and white balls in Bingo Cage 2. After you have made your choices, we will draw a ball from Bingo Cage 2 and your winnings will depend on your choices and the outcome of this draw. Finally, the sealed envelope will be opened and we will show the chosen numbered ball to everyone, and we will also publicly count the number of red and white balls in Bingo Cage 2. We go through with this verification process so that you can believe that the experiment will take place exactly as we describe in the instructions.

Now we will explain how you will actually make your choices. To make your choices, you will use a computer screen like the one shown below.

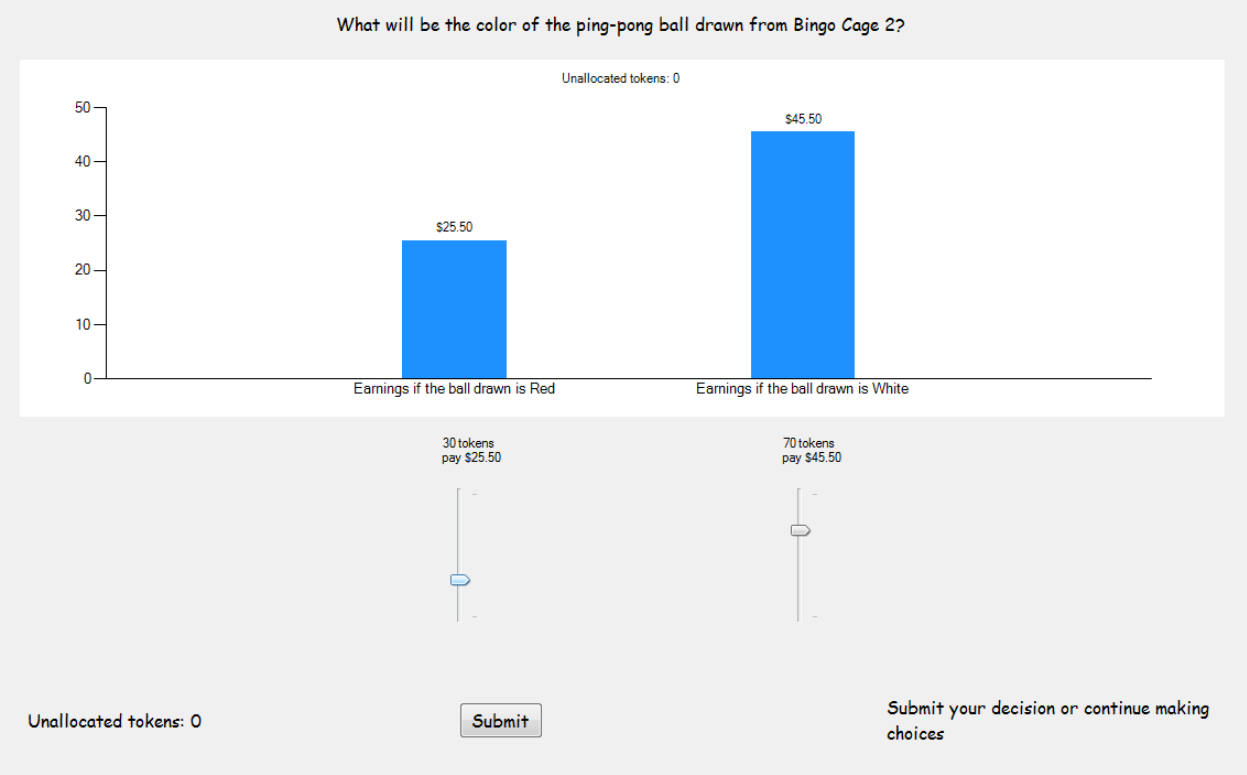


The display on your computer will be larger and easier to read. You have 2 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we always start with 50 tokens being allocated to each slider. The dollar payoffs shown on the screen only apply when you allocate all 100 tokens. As you allocate tokens, by adjusting sliders, the dollar payoffs displayed on the screen will change. Your earnings are based on the payoffs that are displayed after you have allocated all 100 tokens.

You can earn up to \$50 in this task.

Where you position each slider depends on your beliefs about the color of the Ping-Pong ball to be drawn from the bingo cage. The tokens you allocate to each bar will naturally reflect your beliefs about the number of red and white balls in Bingo Cage 2. The bar on the left depends on your beliefs that the ball to be drawn will be red and the bar on the right depends on your beliefs that the ball to be drawn will be white. Each bar shows the amount of money you earn if the ball drawn from the bingo cage is red or white.

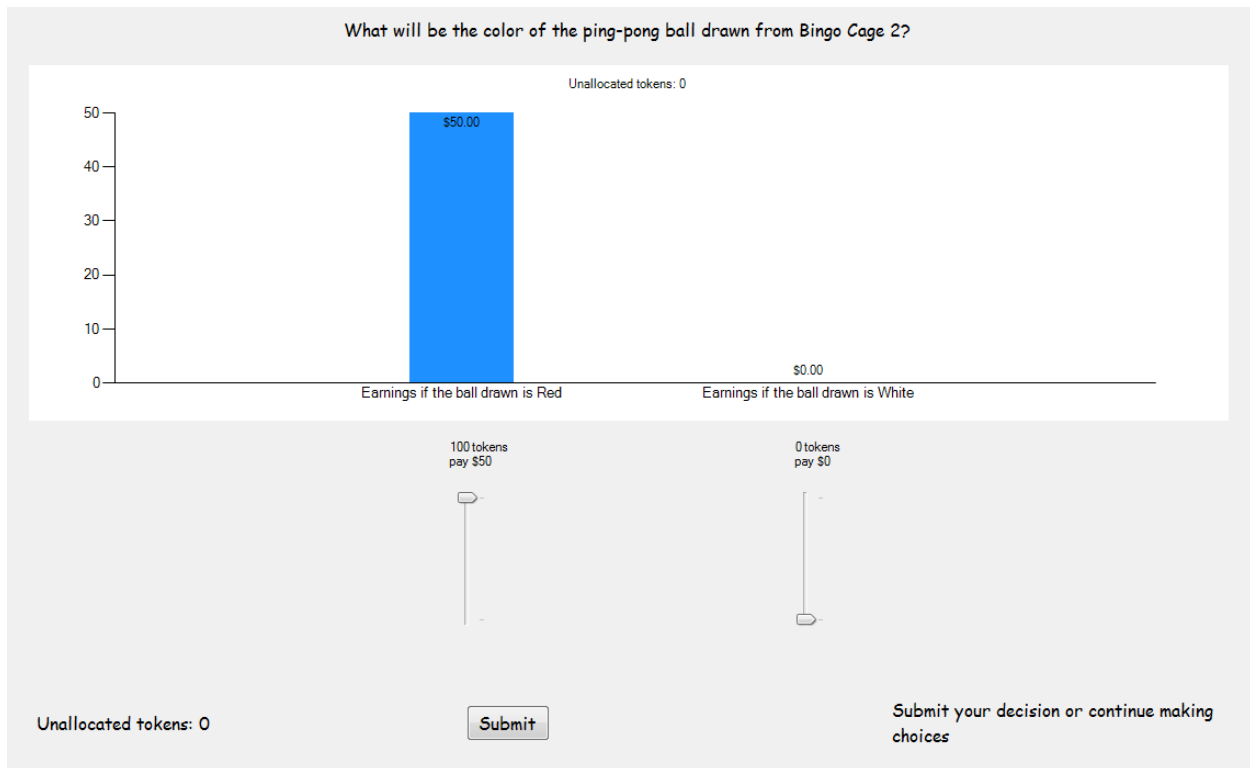
To illustrate how you use these sliders, suppose you think there is a fair chance that there are less red balls than white balls in Bingo Cage 2. Then you might allocate 30 tokens to the first bar, as shown below. Notice that the second bar will be automatically adjusted depending on the number of tokens you allocated on the first bar. Therefore, by allocating 30 tokens to the first bar you are allocating 70 tokens to the second. So you can see that if indeed the ball drawn is red you would now earn \$25.50. If the ball drawn is white instead you would earn \$45.50.



The above pictures show someone who allocated 30 tokens to red Ping-Pong balls and 70 tokens to white Ping-Pong balls. You can adjust this as much as you want to best reflect your personal beliefs about the composition of the bingo cage.

Your earnings depend on your reported beliefs and, of course, the ball drawn. Suppose that a red ball was drawn from Bingo Cage 2 and you reported the beliefs shown above. You would have earned \$25.50.

What if instead you had put all of your eggs in one basket, and allocated all 100 tokens to the draw of a red ball? Then you would have faced the earnings outcomes shown below.



Note the “good news” and “bad news” here. If the chosen ball is red, you can earn the maximum payoff, shown here as \$50. But if a white ball is chosen, then you would have earned nothing in this task.

It is up to you to balance the strength of your personal beliefs with the risk of them being wrong. There are three important points for you to keep in mind when making your decisions:

Your belief about the chances of each outcome is a personal judgment that depends on the information you have about the different events. Remember that you will have the chance to see Bingo Cage 2 being spun for ten turns before it is covered again. This is the information you will have to make your choices.

Depending on your choices and the ball drawn from Bingo Cage 2 you can earn up to \$50.

Your choices might also depend on your willingness to take risks or to gamble. There is no right choice for everyone. For example, in a horse race you might want to bet on the long shot since it will bring you more money if it wins. On the other hand, you might want to bet on the favorite since it is more likely to win something.

The decisions you make are a matter of personal choice. Please work silently, and make your choices by thinking carefully about the task you are presented with.

When you are happy with your decisions, you should click on the **Submit** button and confirm your choices. When everyone is finished we will uncover and spin Bingo Cage 2, and pick one ball at random in front of you. Then an experimenter will come to you and record your earnings according to the color of the ball that was picked and the choices you made.

All payoffs are in cash, and are in addition to the \$7.50 show-up fee that you receive just for being here. The only other task today is for you to answer some demographic questions. Your answers to those questions will not affect your payoffs.

Are there any questions?

Your Beliefs

This is a task where you will be paid according to how accurate your beliefs are about certain things. You will be presented with one and only one question of the type we will explain below. You will actually get the chance to play the question presented to you, so you should think carefully about your answer to the question.

You will make decisions about the color of a ball to be drawn from a bingo cage. This bingo cage will contain 100 balls colored red and white. The exact mix of red and white balls will be unknown to you, but you will receive information about the mixture. The following instructions explain in more detail how this experiment will work.

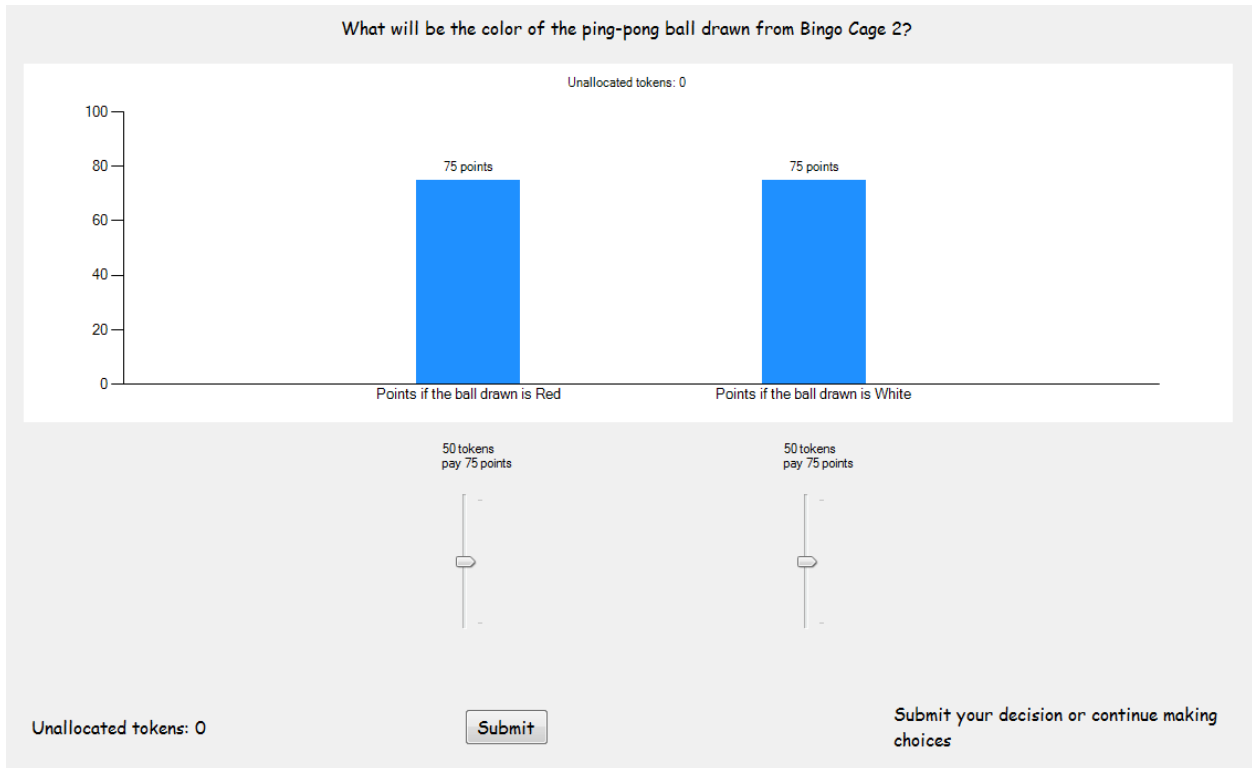
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab except the experimenter. Failure to do so will result in that person losing the promised amount, another person being chosen as verifier, and a restart of the experiment.

We have two bingo cages: Bingo Cage 1 and Bingo Cage 2. We will load Bingo Cage 1 with balls numbered 1 to 99. You will watch us do this, and be able to verify yourself that Bingo Cage 1 is loaded with the correct numbered balls. We will then draw a numbered ball from Bingo Cage 1. However, the draw of a numbered ball from Bingo Cage 1 will take place behind a divider, and you will not know the outcome of this draw until the very end of the experiment, after you have made your decisions. Any number between 1 and 99 is equally likely.

The number on the chosen ball from Bingo Cage 1 will be used to construct Bingo Cage 2 behind the divider. The total number of balls in Bingo Cage 2 will always be 100: the number of red balls will match the number drawn from Bingo Cage 1, and the number of white balls will be 100 minus the number of red balls. Since the actual composition of the Bingo Cage 2 will only be revealed and verified in front of everybody at the end of the experiment, the Verifier will confirm that the experimenter constructs Bingo Cage 2 according to the randomly chosen numbered ball. Once Bingo Cage 2 is constructed, the experimenter will put the chosen numbered ball in an envelope and affix it to the front wall above the white board.

Next, Bingo Cage 2 will be covered and placed on the platform in the front of the lab. Then, Bingo Cage 2 will be uncovered for you to see and spun for 10 turns. After this, we will again cover Bingo Cage 2. You will then make your decisions about the number of red and white balls in Bingo Cage 2. After you have made your choices, we will draw a ball from Bingo Cage 2 and your winnings will depend on your choices and the outcome of this draw. Finally, the sealed envelope will be opened and we will show the chosen numbered ball to everyone, and we will also publicly count the number of red and white balls in Bingo Cage 2. We go through with this verification process so that you can believe that the experiment will take place exactly as we describe in the instructions.

Now we will explain how you will actually make your choices. To make your choices, you will use a computer screen like the one shown below.

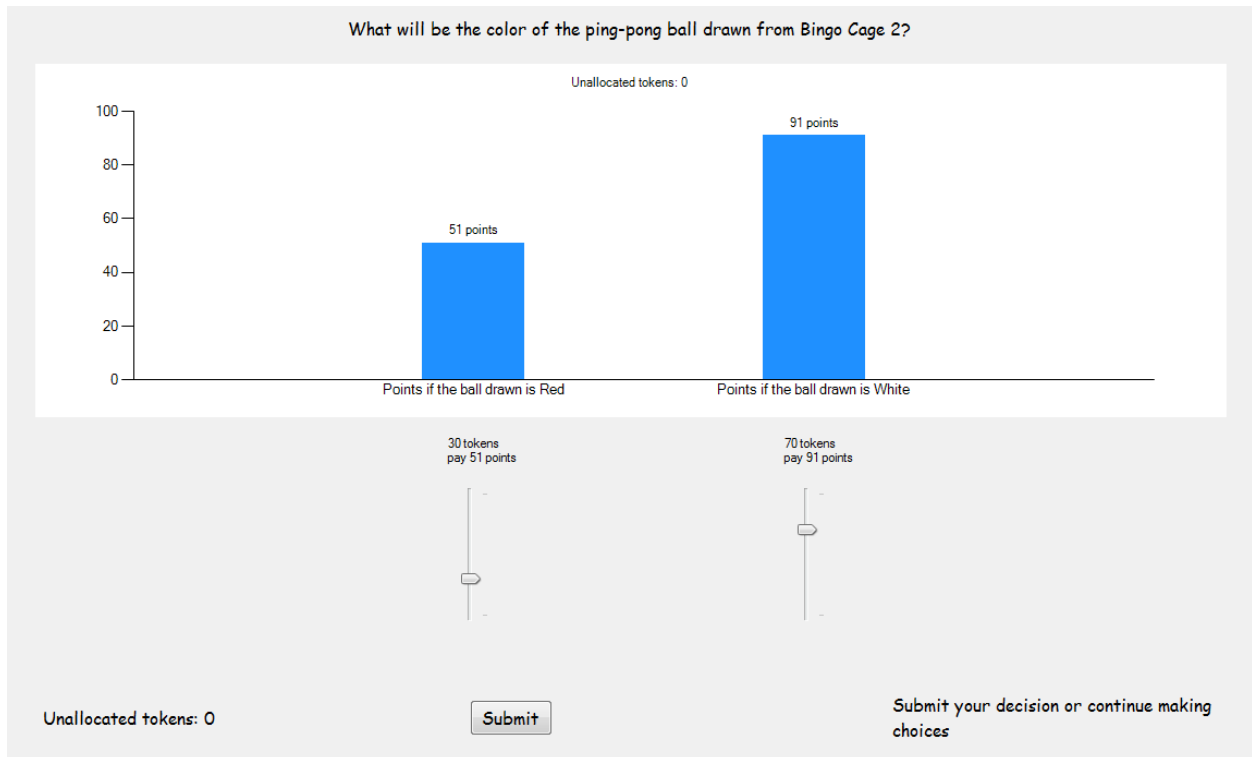


The display on your computer will be larger and easier to read. You have 2 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we always start with 50 tokens being allocated to each slider. The point payoffs shown on the screen only apply when you allocate all 100 tokens. As you allocate tokens, by adjusting sliders, the point payoffs displayed on the screen will change. Your earnings are based on the payoffs that are displayed after you have allocated all 100 tokens.

You earn points in this task. Every point that you earn gives you a greater chance of being paid \$50. To be paid for this task you will roll two 10-sided dice, with every outcome between 1 and 100 equally likely. If you roll a number that is less than or equal to your earned points, you earn \$50; otherwise you earn \$0.

Where you position each slider depends on your beliefs about the color of the Ping-Pong ball to be drawn from the bingo cage. The tokens you allocate to each bar will naturally reflect your beliefs about the number of red and white balls in Bingo Cage 2. The bar on the left depends on your beliefs that the ball to be drawn will be red and the bar on the right depends on your beliefs that the ball to be drawn will be white. Each bar shows the amount of points you earn if the ball drawn from the bingo cage is red or white.

To illustrate how you use these sliders, suppose you think there is a fair chance that there are less red balls than white balls in Bingo Cage 2. Then you might allocate 30 tokens to the first bar, as shown below. Notice that the second bar will be automatically adjusted depending on the number of tokens you allocated on the first bar. Therefore, by allocating 30 tokens to the first bar you are allocating 70 tokens to the second. So you can see that if indeed the ball drawn is red you would now earn 51 points. If the ball drawn is white instead you would earn 91 points.

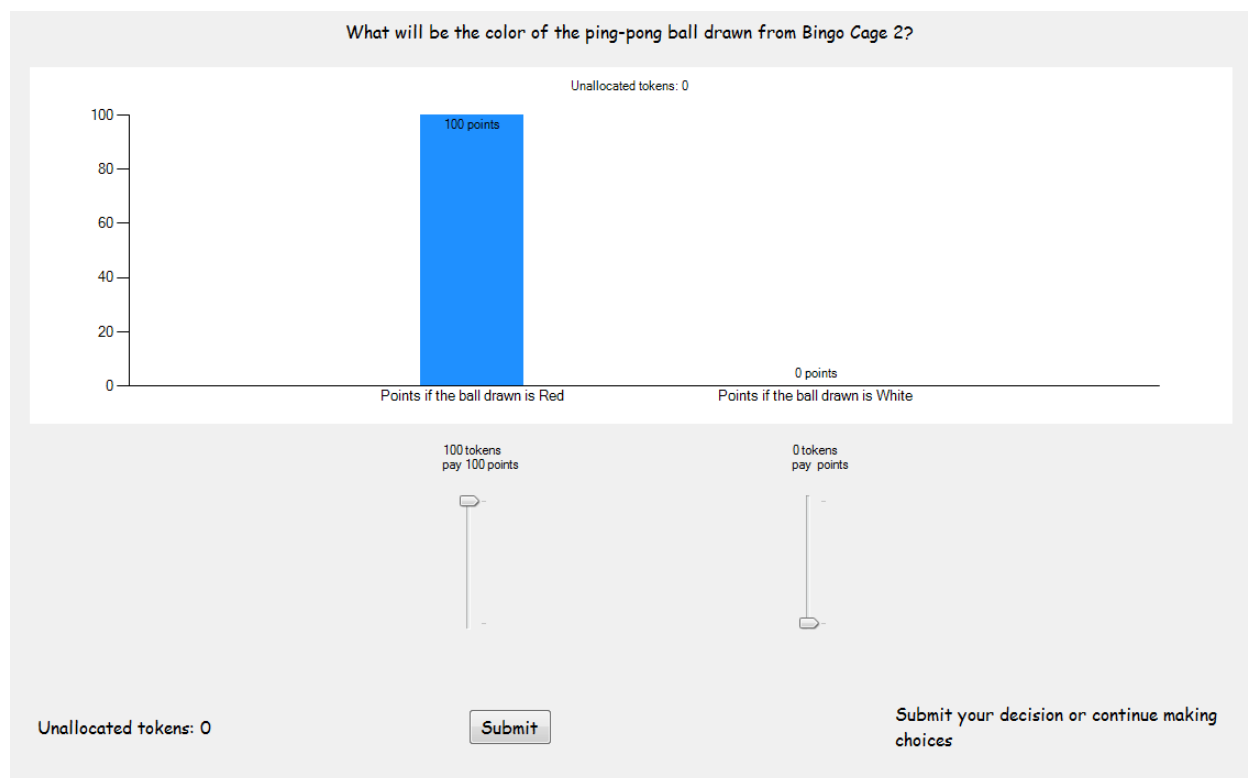


The above pictures show someone who allocated 30 tokens to red Ping-Pong balls and 70 tokens to white Ping-Pong balls. You can adjust this as much as you want to best reflect your personal beliefs about the composition of the bingo cage.

For instance, suppose you allocated your tokens as in the figure shown above. If a red ball is drawn from Bingo Cage 2, then you would earn 51 points. Then suppose that you rolled a 40 with the two 10-sided dice. In this case, you would be paid \$50 since your dice roll is less than or equal to your earned points. However, if your dice roll was some number greater than 51, say 60, you earn \$0. If you earn 100 points then you will earn \$50 for sure, since every outcome of your dice roll would result in a number less than or equal to 100.

If you do not earn \$50 you receive nothing from this task, but of course get to keep your show-up fee. Again, the more points you earn in the correct bar the greater your chance of getting \$50 in this task.

What if instead you had put all of your eggs in one basket, and allocated all 100 tokens to the draw of a red ball? Then you would have faced the earnings outcomes shown below.



Note the “good news” and “bad news” here. If the chosen ball is red, you can earn the maximum payoff, shown here as 100 points. But if a white ball is chosen, then you would have earned nothing in this task.

It is up to you to balance the strength of your personal beliefs with the risk of them being wrong. There are three important points for you to keep in mind when making your decisions:

Your belief about the chances of each outcome is a personal judgment that depends on the information you have about the different events. Remember that you will have the chance to see Bingo Cage 2 being spun for ten turns before it is covered again. This is the information you will have to make your choices.

Depending on your choices and the ball drawn from Bingo Cage 2 you can only earn either \$50 or \$0.

More points increase your chance of being paid \$50. The points you earn will be compared with the outcome of the roll of the two 10-sided dice to determine whether you win \$50 or \$0.

The decisions you make are a matter of personal choice. Please work silently, and make your choices by thinking carefully about the task you are presented with.

When you are happy with your decisions, you should click on the **Submit** button and confirm your choices. When everyone is finished we will uncover and spin Bingo Cage 2, and pick one ball at random in front of you. Then an experimenter will come to you and record your earnings according to the color of the ball that was picked and the choices you made.

All payoffs are in cash, and are in addition to the \$7.50 show-up fee that you receive just for being here. The only other task today is for you to answer some demographic questions. Your answers to those questions will not affect your payoffs.

Are there any questions?

Appendix B: Structural Estimation of Subjective Probabilities

We follow Andersen, Fountain, Harrison and Rutström (2010) and develop a structural econometric model to be estimated in two stages. The objective is to jointly estimate risk attitudes and the underlying subjective probabilities in the M treatment. We use choices from a risk task to identify risk attitudes, and the reports of subjects in the M treatment in the belief elicitation to then identify the subjective probabilities.

First we present the specification of risk attitudes assuming an EUT model of latent choice, where risk attitudes are entirely captured by the concavity of the estimated utility function. Second, we consider the joint estimation of risk attitudes and subjective probability, conditional on the EUT specification.

Estimation of Risk Attitudes

We assume the following constant relative risk aversion (CRRA) utility function

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

Risk neutrality is characterized by r equal to zero, risk aversion is characterized by positive values of r , and risk loving behavior by negative values of r . The parameter in the utility function (1) can be estimated by maximum likelihood estimators and a latent EUT structural model of choice. The functional form of utility employed here is of no importance, and any monotonic increasing function of $U(\cdot)$ could have been implemented.

Let there be K possible outcomes in a lottery; in our lottery choice task $K \leq 4$. Under EUT the probabilities for each outcome k in the lottery choice task, p_k , are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i :

$$EU_i = \sum_{k=1,K} [p_k \times U_k]. \quad (2)$$

The EU for each lottery pair is calculated for a candidate estimate of r , and the following latent index is calculated:

$$\nabla EU = EU_R - EU_L \quad (3)$$

and where EU_L is the “left” lottery and EU_R is the “right” lottery, as displayed to the subject in the risk binary choice task. This latent index, based on latent EUT preferences, is then linked to observed choices using a function $\Phi(\nabla EU)$. We assume this to be a “logit” function that takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the probit link function,

$$\text{prob (choose right lottery)} = \Phi(\nabla EU) \quad (4)$$

Thus the likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of r given the above statistical specification and the observed choices. If we ignore responses that reflect indifference³¹ the log-likelihood is then

$$\ln L(r; y, \mathbf{X}) = \sum_i [(\ln (\nabla EU) \times \mathbf{I}(y_i = 1)) + (\ln (1 - \nabla EU) \times \mathbf{I}(y_i = -1))] \quad (5)$$

where $\mathbf{I}(\cdot)$ is the indicator function, $y_i = 1(-1)$ denotes the choice of the Option R (L) lottery in risk aversion task i , and \mathbf{X} is a vector of individual characteristics.

Even though this “link function” is common in econometrics texts, it forms the critical statistical link between observed binary choices, the latent structure generating the index ∇EU , and the probability of that index being observed. The index defined by (3) is linked to the observed choices by specifying that the right lottery is chosen when $\Phi(\nabla EU) > 1/2$, which is implied by (4). Therefore, the purpose of this link function is to model the possibility that the subject might commit errors when comparing the expected utility of any two given route choices. If there were no errors from the perspective of EUT, this function would be a step function equal to zero when $\nabla EU < 0$ and equal to one when $\nabla EU > 0$. Thus, if there were no errors, for any infinitesimal difference between the subject’s expected utility evaluations of two given choices, the subject would be able to discern which of the two alternatives is better for him with complete certainty.

An important extension of the core model is to allow for subjects to make some errors. The notion of error is one that has already been encountered in the form of the statistical assumption (4) that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery.³² By varying the shape of the link function implicit in (4), one can informally imagine subjects that are more sensitive to a given difference in the index ∇EU and subjects that are not so sensitive. We use the contextual error specification proposed by Wilcox (2011). It posits the latent index

$$\nabla EU = ((EU_R - EU_L) / \nu) / \mu, \quad (3')$$

instead of (3), where ν is a normalizing term for each lottery pair L and R, and $\mu > 0$ is a structural “noise parameter” used to allow some errors from the perspective of the deterministic EUT model. The normalizing term ν is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair, and ensures that the *normalized* EU difference $[(EU_R - EU_L) / \nu]$ remains in the unit interval. As $\mu \rightarrow \infty$ this specification collapses ∇EU to 0 for any values of EU_R and EU_L , so the probability of either choice converges to $1/2$. So a larger μ means that the difference in the EU of the two lotteries, conditional on the estimate of r , has less predictive

³¹ In our lottery experiments the subjects are told at the outset that any expression of indifference would mean that the experimenter would toss a fair coin to make the decision for them if that choice was selected to be played out. Hence one can modify the likelihood to take these responses into account either by recognizing this is a third option, the compound lottery of the two lotteries, or alternatively that such choices imply a 50:50 mixture of the likelihood of choosing either lottery, as illustrated by Harrison and Rutström (2008; p.71). We do not consider indifference here because it was an extremely rare event.

³² This assumption is clear in the use of a link function from the difference between the EU of each option to the probability of picking one or other lottery; in the case of the logistic CDF that is implied by our approach in (3), this link function is $\Lambda(EU_R - EU_L)$. If the subject exhibited no errors from the perspective of EUT, this link function would instead be a step function: zero for all values of $(EU_R - EU_L) < 0$, anywhere between 0 and 1 for $(EU_R - EU_L) = 0$, and 1 for all values of $(EU_R - EU_L) > 0$. Harrison (2008; p.326) illustrates the implied CDF, referring to it as the CDF of a “Hardnose Theorist.”

effect on choices. Thus μ can be viewed as a parameter that flattens out, or “sharpens,” the link functions implicit in (4). This is just one of several different types of error story that could be used, and Wilcox (2008) provides a masterful review of the implications of the strengths and weaknesses of the major alternatives.

Thus we extend the likelihood specification to include the noise parameter μ and maximize $\ln L(r, \mu; y, \mathbf{X})$ by estimating r and μ , given observations on y and \mathbf{X} .³³ Additional details of the estimation methods used, including corrections for “clustered” errors when we pool choices over subjects and tasks, is provided by Harrison and Rutström (2008, p.69ff).

Estimation of Subjective Probabilities

To estimate the subjective probability π that each subject holds from the responses in the belief elicitation task we have to assume something about how subjects make decisions under risk. We assume for simplicity that risk attitudes are characterized by EUT. In this model objective probabilities and subjective probabilities are treated equally. This means that by observing choices over lotteries with objective probabilities, we can identify the utility function that a subject would use in lotteries with subjective probabilities, the domain we are interested in for inferring subjective probabilities. We then jointly estimate the subjective probability and the parameters of the EUT model.

The subject that selects a report θ for the number of red balls in the Bingo Cage receives the following SEU

$$SEU_{\theta} = \pi_R \times U(\$S(\theta | R)) + (1 - \pi_R) \times U(\$S(\theta | W)) \quad (6)$$

This report can take 101 different integer values from 0 to 100. Then we can calculate the likelihood of that choice given values of r , π_R , ν and μ , where the likelihood is the multinomial analogue of the logit specification for the link function used for lottery choices in the risk task. We define

$$eu_{\theta} = \exp[(SEU_{\theta}/\nu)/\mu] \quad (7)$$

for any report θ , and then

$$\nabla EU = eu_{\theta} / (eu_{\theta} + eu_1 + \dots + eu_{100}) \quad (8)$$

for the specific report θ observed, analogously to (3).

We need r to evaluate the utility function in (6), we need π_A to calculate the SEU_{θ} in (6) for each *possible* report θ in $\{0, 1, 2, \dots, 100\}$ once we know the utility values, and we need μ to calculate the latent indices (7) and (8) that generate the subjective probability of observing the choice of specific report θ when we allow for some noise in that process. The *joint* maximum likelihood problem is to find the values of these parameters that best explain observed choices in the belief elicitation tasks as well as observed choices in the lottery tasks.

For numerical reasons we constrain the estimates for session 1 to lie in the interval (0, 0.5), for session 4 to lie in the interval (0.5, 1). These intervals span the vast majority of responses. Results are essentially unchanged if we delete the few outliers from these intervals.

³³ The normalizing term ν is given by the value of r and the lottery parameters, which are part of \mathbf{X} .

We can easily extend the joint estimation to allow risk attitudes to be characterized by the RDU model. As before the subject selects one report θ for the number of red balls in the Bingo Cage out of 101 possible ones. The subject chooses a report that maximizes the following ranked-dependent expected utility

$$RDU_{\theta} = w(\pi_R) \times U(\$S(\theta | R)) + (1-w(\pi_R)) \times U(\$S(\theta | W))$$

where $\$S(\theta | R) > \$S(\theta | W)$ since the weight depends on the ranking of the prices.³⁴ Therefore the RDU model extends the EUT model by allowing for decision weights on lottery outcomes that can be different from probabilities. We adopt the simple “Prelec” probability weighting with curvature parameter γ :

$$w(p) = p^{\gamma} / (p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$$

This function allows for the inverted-S shape popularized by Tversky and Kahneman (1992). The rest of the ML specification for the RDU model is identical to the specification for the EUT model except that the valuation of each lottery in the analysis must be calculated using (6) and that one has to estimate different parameters.

Clarifications

When we then say joint estimation is used, we simply mean that we estimate the CRRA coefficient r and the subjective probability π from the two samples together. By “together” we literally just stack the data from the risk aversion choices on top of the data from the belief choices, and find the values of the two parameters that maximize the product of the likelihood of the risk aversion choices and the likelihood of the belief choices. It is joint in the sense that the maximum likelihood value of r depends on the belief choices, as well as the risk aversion choices, as theory tells us. But the maximum likelihood value of π only depends on the belief choices, conditional on some r value. So the second last sentence implies that one could *not* conceptually estimate the r value correctly just using the risk aversion choices.

Even if we assume that every subject in the risk aversion and belief tasks had the same risk attitudes, we would still infer a distribution of estimated subjective probabilities. To see why, assume the elicited subjective belief reports are as follows for a sample of three

$$(0.1, 0.9), (0.5, 0.5), (0.9, 0.1), \tag{9}$$

where the numbers inside parentheses are (report for red, report for white) and both are from the money treatment. Now assume that the CRRA coefficient was 0.6, a reasonable value given our estimates in Appendix C. If the reports in (9) are from the money treatment for three distinct subjects, and we assume that the *same* CRRA coefficient of 0.6 was applied to all three subjects to infer their subjective probabilities, we would end up with something like this:

$$(0.23, 1-0.23), (0.5, 0.5), (0.77, 1-0.77) \tag{10}$$

We just pick increments and decrements of 0.13 to illustrate; the exact effect could be calculated, but is not important. The first and last reports for red get moved toward 0.5 because of risk aversion,

³⁴ If $\$S(\theta | R) \leq \$S(\theta | W)$, then the valuation of the lottery induced by the report is given by $RDU_{\theta} = (1-w(\pi_W)) \times U(\$S(\theta | R)) + w(\pi_W) \times U(\$S(\theta | W))$

and the middle report stays the same. So (10) is then a *distribution* of elicited/estimated subjective probabilities, based on the *distribution* of raw reports in (9) and the *one*, common CRRA coefficient. One would only see the reports in (9) all become 0.5 if there was *really extreme* risk aversion, say $CRRA = 2$ or 3 using the utility specification we use. These high values of CRRA are implausible given our data.

Having said this, it is possible to estimate the risk aversion coefficient as a linear function of observable demographics. In this case the inferred subjective probabilities would be conditioned on the risk attitude of subjects with the same demographic characteristic. For instance, if we estimated a risk coefficient r for men and a risk coefficient r for women, and they differed, then the same reported subjective probability by a man and woman would generate different inferred subjective probabilities. To take an extreme case, if the man was risk neutral and the woman risk averse, the estimated subjective probability for the man would be the same as the reported subjective probability, and the estimated subjective probability for the woman would be closer to $\frac{1}{2}$ than the reported subjective probability.

We estimate models first assuming homogeneity in risk attitudes and beliefs. We also estimate the same models controlling for observable characteristics in risk attitudes. For the latter case, assuming that risk attitudes are characterized by Subjective Expected Utility, we estimate the models for all sessions using two demographic characteristics: whether the subject is a female and whether the subject self-reports being White or Caucasian. In the models where we assume that risk attitudes are characterized by Ranked-Dependent Expected Utility, where there is one additional parameter to estimate, we control for gender characteristics.

Additional References

- Harrison, G. W., Rutström, E. E., 2008. Risk Aversion in the Laboratory. In: Cox, J.C., Harrison, G.W. (Eds.). Risk Aversion in Experiments. Bingley, UK: Emerald, Research in Experimental Economics, Volume 12.
- Quiggin, J., 1982. A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization* 3, 323-343.
- Tversky, A., Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5, 297-323.
- Wilcox, N. T., 2008. Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison. In: Cox, J.C., Harrison, G.W. (Eds.). Risk Aversion in Experiments. Bingley, UK: Emerald, Research in Experimental Economics, Volume 12.
- Wilcox, N. T., 2011. Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk. *Journal of Econometrics* 162, 89-104.

Appendix C: Estimates of Subjective Probabilities

Estimates for Sessions 1 and 4: Subjective Expected Utility

```

Log likelihood = -254.66619
Number of obs   =      194
Wald chi2(0)    =          .
Prob > chi2     =          .
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
_cons	.6068706	.1084811	5.59	0.000	.3942516	.8194896
muRA						
_cons	.2260159	.06829	3.31	0.001	.09217	.3598618
sprob_1						
_cons	-.4242404	.9982502	-0.42	0.671	-2.380775	1.532294
sprob_4						
_cons	.3462428	.9736456	0.36	0.722	-1.562067	2.254553

```

sprob1: 0.5/(1+exp([sprob_1]_cons))
sprob4: 0.5+(0.5/(1+exp([sprob_4]_cons)))
sprob1d: 0.5/(1+exp([sprob_1]_cons))- .2676470588235294
sprob4d: 0.5+(0.5/(1+exp([sprob_4]_cons)))- .758421052631579
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sprob1	.3022487	.119331	2.53	0.011	.0683644	.5361331
sprob4	.7071469	.1181297	5.99	0.000	.4756169	.9386769
sprob1d	.0346017	.119331	0.29	0.772	-.1992827	.2684861
sprob4d	-.0512741	.1181297	-0.43	0.664	-.2828041	.1802559

```

Log likelihood = -251.92554
Number of obs   =      194
LR chi2(2)      =       5.48
Prob > chi2     =     0.0645
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
female	.1470683	.226732	0.65	0.517	-.2973183	.5914548
white	-.7754617	.4993239	-1.55	0.120	-1.754119	.2031952
_cons	.6077832	.1890013	3.22	0.001	.2373474	.978219
muRA						
_cons	.2270518	.0665378	3.41	0.001	.0966401	.3574634
sprob_1						
_cons	-.4217744	.9909054	-0.43	0.670	-2.363913	1.520364
sprob_4						
_cons	.3401037	.9620159	0.35	0.724	-1.545413	2.22562

```

sprob1: 0.5/(1+exp([sprob_1]_cons))
sprob4: 0.5+(0.5/(1+exp([sprob_4]_cons)))
sprob1d: 0.5/(1+exp([sprob_1]_cons))- .2676470588235294
sprob4d: 0.5+(0.5/(1+exp([sprob_4]_cons)))- .758421052631579
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sprob1	.3019539	.1185139	2.55	0.011	.069671	.5342368
sprob4	.7078921	.1168405	6.06	0.000	.4788889	.9368954
sprob1d	.0343068	.1185139	0.29	0.772	-.1979761	.2665897
sprob4d	-.0505289	.1168405	-0.43	0.665	-.2795322	.1784743

Estimates for Sessions 1 and 2: Ranked-Dependent Utility

Log likelihood = -239.71879 Number of obs = 192
 Wald chi2(0) = .
 Prob > chi2 = .

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
r					
_cons	.2641713	.171394	1.54	0.123	-.0717548 .6000974
gamma					
_cons	.4883822	.0592821	8.24	0.000	.3721914 .6045729
muRA					
_cons	.1478071	.0302434	4.89	0.000	.0885311 .2070831
sprob_1					
_cons	1.249737	.7273325	1.72	0.086	-.1758087 2.675282
sprob_2					
_cons	-.6137131	.6045293	-1.02	0.310	-1.798569 .5711426

sprob1: 1/(1+exp([sprob_1]_cons))
 sprob2: 1/(1+exp([sprob_2]_cons))
 sprob1d: 1/(1+exp([sprob_1]_cons))- .2676470588235294
 sprob2d: 1/(1+exp([sprob_2]_cons))- .6576470588235294

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
sprob1	.2227457	.1259231	1.77	0.077	-.0240591 .4695505
sprob2	.6487873	.1377495	4.71	0.000	.3788034 .9187713
sprob1d	-.0449014	.1259231	-0.36	0.721	-.2917061 .2019034
sprob2d	-.0088597	.1377495	-0.06	0.949	-.2788437 .2611242

Log likelihood = -239.313 Number of obs = 192
 LR chi2(2) = 0.81
 Prob > chi2 = 0.6664

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
r					
female	.2497302	.3641912	0.69	0.493	-.4640715 .9635319
_cons	.1142376	.2968732	0.38	0.700	-.467623 .6960983
gamma					
female	.0067382	.1232844	0.05	0.956	-.2348949 .2483712
_cons	.4866615	.084526	5.76	0.000	.3209936 .6523295
muRA					
_cons	.1468466	.0310443	4.73	0.000	.0860008 .2076923
sprob_1					
_cons	1.228522	.7280314	1.69	0.092	-.198393 2.655438
sprob_2					
_cons	-.5988049	.6149065	-0.97	0.330	-1.803999 .6063897

sprob1: 1/(1+exp([sprob_1]_cons))
 sprob2: 1/(1+exp([sprob_2]_cons))
 sprob1d: 1/(1+exp([sprob_1]_cons))- .2676470588235294
 sprob2d: 1/(1+exp([sprob_2]_cons))- .6576470588235294

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
sprob1	.2264402	.1275256	1.78	0.076	-.0235055 .4763858
sprob2	.6453828	.1407299	4.59	0.000	.3695574 .9212083
sprob1d	-.0412069	.1275256	-0.32	0.747	-.2911525 .2087387
sprob2d	-.0122642	.1407299	-0.09	0.931	-.2880897 .2635612

Estimates for Sessions 3 and 4: Ranked-Dependent Utility

Log likelihood = -254.6024

Number of obs = 195
 Wald chi2(0) = .
 Prob > chi2 = .

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
_cons	.247263	.1894128	1.31	0.192	-.1239793	.6185053
gamma						
_cons	.4708093	.0626644	7.51	0.000	.3479894	.5936291
muRA						
_cons	.1572991	.0336146	4.68	0.000	.0914157	.2231825
sprob_3						
_cons	-.0943526	.6308636	-0.15	0.881	-1.330823	1.142117
sprob_4						
_cons	-1.483217	.8131453	-1.82	0.068	-3.076952	.1105186

sprob3: 1/(1+exp([sprob_3]_cons))
 sprob4: 1/(1+exp([sprob_4]_cons))
 sprob3d: 1/(1+exp([sprob_3]_cons))- .53625
 sprob4d: 1/(1+exp([sprob_4]_cons))- .758421052631579

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sprob3	.5235707	.1573654	3.33	0.001	.2151401	.8320012
sprob4	.815058	.1225723	6.65	0.000	.5748207	1.055295
sprob3d	-.0126793	.1573654	-0.08	0.936	-.3211099	.2957512
sprob4d	.0566369	.1225723	0.46	0.644	-.1836003	.2968742

Log likelihood = -253.78716

Number of obs = 195
 LR chi2(2) = 1.63
 Prob > chi2 = 0.4425

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
female	.2200597	.3872123	0.57	0.570	-.5388625	.9789819
_cons	.1113641	.3099169	0.36	0.719	-.4960618	.71879
gamma						
female	-.0558036	.1221344	-0.46	0.648	-.2951826	.1835753
_cons	.5055247	.1005693	5.03	0.000	.3084126	.7026368
muRA						
_cons	.1554652	.0333069	4.67	0.000	.0901848	.2207456
sprob_3						
_cons	-.1020267	.6484954	-0.16	0.875	-1.373054	1.169001
sprob_4						
_cons	-1.531031	.8525038	-1.80	0.073	-3.201908	.1398458

sprob3: 1/(1+exp([sprob_3]_cons))
 sprob4: 1/(1+exp([sprob_4]_cons))
 sprob3d: 1/(1+exp([sprob_3]_cons))- .53625
 sprob4d: 1/(1+exp([sprob_4]_cons))- .758421052631579

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sprob3	.5254846	.1617027	3.25	0.001	.2085532	.842416
sprob4	.8221571	.1246487	6.60	0.000	.5778502	1.066464
sprob3d	-.0107654	.1617027	-0.07	0.947	-.3276968	.306166
sprob4d	.0637361	.1246487	0.51	0.609	-.1805709	.308043

Appendix D: Proof of Proposition 1

Proposition 1. If $w(\cdot)$ is strictly increasing and $Q(\cdot)$ is strictly concave, then $w(Q(\cdot))$ is strictly quasi-concave.

This is a simple mathematical fact of composite functions for which we provide a simple proof below.

Proof. Suppose $w(\cdot)$ is strictly increasing and $Q(\cdot)$ is strictly concave. Any function $g(\cdot)$ is defined as strictly quasi-concave if $g(\lambda x + (1-\lambda)x^*) > \min\{g(x), g(x^*)\}$ for $\lambda \in (0,1)$. Since $Q(\cdot)$ is strictly concave then $Q(\lambda x + (1-\lambda)x^*) > \lambda Q(x) + (1-\lambda)Q(x^*)$; and since $w(\cdot)$ is strictly increasing then $w(Q(\lambda x + (1-\lambda)x^*)) > w(\lambda Q(x) + (1-\lambda)Q(x^*))$. Because $\lambda \in (0,1)$ we have $\lambda Q(x) + (1-\lambda)Q(x^*) > \min\{Q(x), Q(x^*)\}$. Since $w(\cdot)$ is strictly increasing we have

$$w(\lambda Q(x) + (1-\lambda)Q(x^*)) > \min\{w(Q(x)), w(Q(x^*))\}.$$

These results imply that $w(Q(\lambda x + (1-\lambda)x^*)) > w(\lambda Q(x) + (1-\lambda)Q(x^*)) > \min\{w(Q(x)), w(Q(x^*))\}$.

Therefore $w(Q(\lambda x + (1-\lambda)x^*)) > \min\{w(Q(x)), w(Q(x^*))\}$ for any $\lambda \in (0,1)$. This is the definition of a strictly quasi-concave function, so $w(Q(\cdot))$ is a strictly quasi-concave function. ■