Eliciting Subjective Probabilities with Binary Lotteries

by

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

September 2012

ABSTRACT.

We evaluate the binary lottery procedure for inducing risk neutral behavior in a subjective belief elicitation task. Harrison, Martínez-Correa and Swarthout [2013] found that the binary lottery procedure works robustly to induce risk neutrality when subjects are given one risk task defined over objective probabilities. Drawing a sample from the same subject population, we find evidence that the binary lottery procedure 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 certain Non-Expected Utility preference representations that satisfy weak conditions that we identify.

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

Table of Contents

The notion of subjective probabilities as prices at which one is willing to trade is due to de Finetti [1937][1970] and Savage [1970], who propose bets as one operational procedure to both define and elicit subjective probabilities. Their central point is that subjective probabilities of events are marginal rates of substitution between contingent claims that obey certain rules of coherence.

This literature relies on the assumption that subjects behave as if they are *risk neutral*, which is considered plausible *a priori* if the stakes used in the elicitation procedure are sufficiently small. 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. Therefore, it is intuitively obvious, and also well known in the literature (e.g., Winkler and Murphy [1970] and Kadane and Winkler [1988]), that risk attitudes will affect the incentive to directly report one's subjective probability.

Consider a scoring rule to elicit subjective probabilities. A sufficiently risk averse agent is going to be drawn to a report of ½, since this equalizes earnings under each state of nature, at least for the most popular scoring rules. Varying degrees of risk aversion will cause varying distortions in reports from true latent subjective probabilities. If we knew the form of the utility function of the subjects, and their degree of risk aversion, we could infer back from any report what subjective probability they must have had. 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 observable choices. Andersen, Fountain, Harrison and Rutström [2010] illustrate how joint estimation of risk attitudes and subjective beliefs, using structural maximum likelihood, can be used to make the necessary calibration to recover the latent subjective probability.

An alternative and operationally meaningful approach is to use proper scoring rules combined with the Binary Lottery Procedure (BLP) to induce linear utility in subjects. The theoretical prediction is that, under certain conditions, this approach allows the researcher to directly elicit the subjective probability without further statistical corrections for risk attitudes.

Procedures to induce linear utility functions have a long history, with the major contributions being Smith [1961], Roth and Malouf [1979] and Berg, Daley, Dickhaut and O'Brien [1986]. The consensus appears to be that these "binary lottery procedures" may be fine in theory, but behaviorally they just do not work as advertized. However, Harrison, Martínez-Correa and Swarthout [2013] show that the BLP works when contaminating factors such as strategic equilibrium concepts and traditionally used payment protocols are avoided. They find that the BLP works robustly to induce risk neutrality when subjects are given one binary lottery choice, and that it also works well when subjects are given more than one binary choice. Of course, this does not imply that BLP for all samples from different populations or that it can be systematically applied to any setting, since it is often the "contaminating factors" of interest in some designs that can dilute the power of BLP to induce risk neutrality.

Our central focus in this paper is to determine whether the result that the BLP induces risk neutrality in simple binary choices defined over objective probabilities found by Harrison, Martínez-Correa and Swarthout [2012] also extends to subjective probability elicitation tasks. In particular, we study the ability of the Quadratic Scoring Rule (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]. Schlag and van der Weel [2009] and Hossain and Okui [2012] examine the same extension of the QSR, along with certain generalizations, calling it a "randomized QSR" and "probabilistic scoring rule" respectively.

There 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, Sonnemans, van de Kuilen and Wakker [2009], Karni [2009], and Holt and Smith [2009].

Thus there is no shortage of theoretical procedures to elicit subjective probabilities, and the issue becomes which 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 differences between the procedures. That elicitation context, while important, is complex, as stressed by Rutström and Wilcox [2009].

Two aspects of elicitation concern us in operationalizing these procedures. The first is the use of procedures that assume the validity of the Becker, De Groot and Marschak [1964] procedure for eliciting certainty-equivalents; Harrison [1992] explains the concerns. The second is the use of payment protocols over multiple choices that assume the validity of the (compound or mixture) independence axiom; Harrison and Swarthout [2012] explain the concerns. We avoid both of these in our tests.

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 assumption that subjects are risk averse and that the BLP reduces the contaminating factor of risk aversion. Structural econometric estimations support these findings since the *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]

In section 1 we review the theory of scoring rules and explain the benefits of using the BLP in belief elicitation tasks. In section 2 we present our experimental design, in section 3 we evaluate the results, and in section 4 we conclude.

[1] In this respect, tests of the BLP with lotteries defined over objective probabilities are much easier. Hossain and Okui [2012; §3.1] use the QSR with and without the BLP to test if elicited beliefs are the same as *objective* probabilities when one uses the BLP, and find that they are indeed closer when one uses the BLP.

1. Theoretical Issues

A. Binary Scoring Rules for Subjective Probabilities

A binary scoring rule is defined over some binary event, which is either true or false. A binary scoring rule asks the subject to make some report θ, and then defines how an elicitor pays a subject depending on the report and the outcome of the event. This framework for eliciting subjective probabilities can be formally viewed from the perspective of a trading game between two agents: you give me a report, and I agree to pay you $X if one outcome occurs and $Y if the other outcome occurs. The scoring rule defines the terms of the exchange quantitatively, explaining how the elicitor converts the report from the subject into a lottery defined over the subjective probability of the subject.[2] We use the terminology "report" because we want to view this formally as a mechanism, and want to emphasize the idea that the report may or may not be the subjective probability π of the subject. When the report is equal to subjective probability of the individual, we say that the scoring rule is a direct mechanism, following standard methodology.

The popular QSR for binary events is defined in terms of two positive parameters, α and β, that determine a fixed reward the subject gets and a penalty for error. Assume that the possible outcomes are A or B, where B is the complement of A, that θ is the reported probability for A, and that Θ is the true binary-valued outcome for A. Hence Θ=1 if A occurs, and Θ=0 if B occurs. The subject is paid $S(\theta|A) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(1-\theta)^2$ if event A occurs and $S(\theta|B) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(0-\theta)^2$ if B occurs. The score or payment penalizes the subject by the squared deviation of the report from the true binary-valued outcome, Θ. An omniscient seer would obviously set $\theta = \Theta$. The

[2] 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.

fixed reward is a convenience[3] to ensure that subjects are willing to play this trading game, and the penalty function accentuates the penalty from not responding with what the subject thinks an omniscient seer would respond. It can be shown that a risk neutral decision maker will report his subjective probability truthfully. For example, assume $\alpha = 1$ and $\beta = 1$ so that the subject could earn up to 1 or as little as 0. If they reported 1 they earned 1 if event A occurred or 0 if event B occurred; if they reported ¾ they earned 0.9375 or 0.4375; and if they reported ½ they earned 0.75 no matter the realized event.

B. Subjective Belief Elicitation with Scoring Rules and the Binary Lottery Procedure

Our strategy is to rely on the BLP to induce linear utility functions in subjects, which implies that the QSR should provide incentives to subjects to report their subjective probabilities truthfully. The central insight 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 binary lottery. We explain below the conditions under which this combination of BLP and the QSR provides incentives to individuals to directly report "truthfully" their unobserved subjective probabilities.

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 and, under some weak conditions, independently of the shape of the probability weighting function. For exposition purposes, take first the case of a Subjective Expected Utility (SEU) maximizer that is given a QSR task defined over points using the BLP to convert those points into money.

Assume that there are two events: a ball drawn from a Bingo Cage is either Red (R) or

[3] In the language of mechanism design, it can be chosen to satisfy the participation constraint. This requires that $\alpha \geq \beta$.

White (W). A subject betting on event R might estimate that it happens with subjective probability π_R, and that W will happen with subjective probability $\pi_W = (1-\pi_R)$. Additionally, set the parameters of the QSR to be $\alpha = \beta = 100$.

If event R is realized and a subject reports θ, he wins an amount of points defined by $S(\theta|R) = 100 - 100(1-\theta/100)^2$. For simplicity, the report can be any number between 1 and 100, although for practical purposes we can think of reports in increments of single percentage points. Similarly, if event W is realized and a subject reports θ, he wins an amount of points given by $S(\theta|W) = 100 - 100(0-\theta/100)^2$. 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 51% and \$0 with 49%. 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 above, $p_R(30) = 51\%$ and $p_W(30)= 91\%$. Figure 1 represents graphically 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. The structure of these SCL are explicitly described below:

> SCL(θ): pays simple lottery ($p_R(\theta)$, \$50; 1-$p_R(\theta)$, \$0) with subjective probability π_R, and pays simple lottery ($p_W(\theta)$, \$50; 1-$p_W(\theta)$, \$0) with subjective probability $\pi_W = (1-\pi_R)$

where θ is any number in the interval [0, 100].[4]

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

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

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

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

We rename SEU(θ) as Q(θ) to emphasize that the subject's valuation of the SCL induced by θ can be interpreted as the subjective average probability Q(θ) of winning the high $50 amount in the binary lottery. Given that the state space of the report is a continuum such that $\theta \in [0, 100]$, a SEU maximizer would make a report θ^* that maximizes the subjective expected probability Q(.) of winning the binary lottery. Taking the first order condition with respect to report θ we obtain

$$\begin{aligned} & \mathrm{SEU}'(\theta) = \mathrm{Q}'(\theta) = \pi_R \times p_R'(\theta) + (1\text{-}\ \pi_R) \times p_W'(\theta) = 0 \\ & = \pi_R \times [2 \times (1\text{-}\theta/100)] + (1\text{-}\ \pi_R) \times [2 \times (0\text{-}\theta/100)] = 0 \end{aligned} \tag{2}$$

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(.) is strictly concave in θ given that it is a linear combination of two strictly concave functions, $p_R(\theta)$ and $p_W(\theta)$. Note that the strict concavity of these functions is determined by the QSR because these objective probabilities are a function of the

[4] For simplicity, in our experiments subjects are only allowed to choose integer numbers in this interval which implies that they can choose between 101 SCL.

scoring rule.[5]

For comparison purposes suppose a simple QSR with payouts defined directly in dollars. Refer to the score in this case as \$S(.). The subject would choose a report θ that maximizes the following valuation

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

Assume a simple power function with risk aversion parameter equal to 0.57 and a subjective probability of $\pi_R = 0.3$. In this case the optimal report would be $\theta^* = 34$. Conversely, if $\pi_R =0 .7$, the optimal report would be $\theta^* = 66$. Notice that 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.

A proper scoring rule provides incentives to subjects to optimally choose one distinct report. The uniqueness of the optimal report can be achieved by guaranteeing that the scoring rule induces strict concavity in the subject's valuation of choices in the belief elicitation task. In the case of the scoring rule without BLP, the subject's valuation SEU(θ, U(.)) is a concave function of θ, assuming only weak monotonicity of U(.). Under SEU this concavity is immediately guaranteed by the concavity of the utility function.[6]

---

[5] In the present example, $p_R(\theta)$ and $p_W(\theta)$ are strictly concave because $p_R''(\theta) = p_W''(\theta) = -2/100<0$.

[6] A Linear Scoring Rule (LSR) 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. Therefore a LSR also results in a subject's valuation that is concave in the report if the utility function is concave. Andersen, Fountain, Harrison and Rutström [2010] show how one can infer true subjective probabilities with the LSR if one also knows the risk attitudes of subjects. However, if subjects are risk neutral the 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 ½ or greater than ½, respectively. This result is immediately relevant if one wants to induce risk neutrality with the BLP. Consequently, if we rely on risk neutrality being induced by the BLP, any scoring rule that we use must be concave in the report that subjects can make for all *weakly* concave utility functions, and not just for strictly concave utility functions.

C. Non-Expected Utility Theory Preference Representations

The incentive-compatibility of the QSR is normally developed theoretically in the context of Expected Utility Theory (EUT), and specifically SEU, whether or not one uses the BLP. When implemented with the BLP, can the QSR be a proper scoring rule for subjects that have non-EUT preference representations? The answer to this question depends on the specific non-EUT preference representation.

A subject who follows the Rank-Dependent Utility model (RDU) will report his subjective probability under relatively weak conditions. Assume that the decision maker is a RDU maximizer with a strictly increasing probability weighting function. Then the higher prize receives decision weight w(*p*), where *p* is the probability of the higher prize, and the lower prize receives decision weight 1-w(*p*). SEU is violated in this case, but none of the axioms needed for the BLP to induce linear utility are violated. For the BLP to directly elicit the subjective probability from a RDU maximizer we need the following assumptions must hold: (1) uniqueness of U(.) up to an affine positive transformation and U(.) increasing, (2) probabilistic sophistication as defined by Machina and Schmeidler [1992][1995], (3) ROCL for binary lotteries, (4) a strictly increasing probability weighting function, and (5) a strictly concave scoring rule. We formally derive below the conditions under which a non-EUT subject would optimally report his true subjective belief.

An individual with RDU preferences will have a QSR valuation of the subjective compound lottery induced by a report θ given by

$$\mathrm{RDU}(\theta) = \mathrm{w}(\mathrm{Q}(\theta)) \times \mathrm{U}(\$50) + (1-\mathrm{w}(\mathrm{Q}(\theta))) \times \mathrm{U}(\$0) \tag{4}$$

Binary ROCL implies that the probability weighting is done on the reduced compound probability Q(θ). Since U(.) is unique up to an affine positive transformation in the RDU model, we can also normalize U($50)=1 and U($0)=0 and the valuation of the individual becomes RDU(θ) = w(q(θ)). An RDU maximizer and a SEU maximizer, each with subjective probability $\pi_R = 0.3$ for example, would have incentives to make exactly the same optimal report, $\theta^* = \pi_R \times 100 = 30$. This is easily seen by taking the first order condition on the subject's valuation of report θ,

$$\text{RDU}'(\theta) = w'(Q(\theta)) \times Q'(\theta) = 0,$$

which is satisfied when $Q'(\theta)=0$, exactly equal to the first order condition of an SEU maximizer, because the probability weighting function is assumed to be strictly increasing (i.e., $w'(Q(\theta))>0$). 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. To guarantee the uniqueness of the optimal report we rely on two assumptions: (1) the scoring rule is strictly concave because $Q(\theta)$ is a linear combination of strictly concave functions that depend on the scoring rule (i.e., $p_R(\theta)= S(\theta|R)/100$ and $p_W(\theta)= S(\theta|W)/100$); and (2) the probability weighting function is strictly increasing, guaranteeing that $w(Q(\theta))$ is strictly quasi-concave, which implies that there is a unique global maximum.

Proposition 1. If F(.) is strictly increasing and f(.) is strictly concave, then F(f(.)) is strictly quasi-concave.

We want to show this result so the uniqueness of the optimal report is guaranteed in the scoring rule task with certain non-EUT preference representations.

Proof. Suppose F is strictly increasing and f is strictly concave. We define g(.) to be strictly quasi-concave if $g(\lambda x+(1-\lambda)x^*) > \min\{g(x),g(x^*)\}$ for $\lambda \in (0,1)$.

Since f(.) is strictly concave then $f(\lambda x+(1-\lambda)x^*)>\lambda f(x)+(1-\lambda)f(x^*)$; and since F is strictly increasing then $F(f(\lambda x+(1-\lambda)x^*))>F(\lambda f(x)+(1-\lambda)f(x^*))$.

Because $\lambda \in (0,1)$, then $\lambda f(x)+(1-\lambda)f(x^*)>\min\{f(x),f(x^*)\}$ and since F is increasing we have that

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

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

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

Hossain and Okui [2012] independently prove the same result with respect to RDU, but our proof is arguably more instructive because it points to a general mechanism-design principle to show how to make the incentives for the scoring rule more powerful. By recognizing that the strict quasi-concavity of $w(Q(\theta))$ is needed to ensure that the scoring rule is proper, we can identify ways to design scoring rules that have better properties in certain regions of θ. This issue is an important one when trying to elicit subjective beliefs with respect to extremely small probabilities, as occurs almost all of the time when designing insurance contracts for low-probability, but high cost, events.

2. Experiment

A. 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. There were no other salient tasks, before or after a subject's choices, affecting the outcome. Table 1 summarizes our experimental design for each of four sessions, and the sample size of subjects in each treatment per session.

Upon arrival at the laboratory, each subject drew a number from a box which determined random seating position within the laboratory. After being seated and signing the informed consent document, subjects were given printed introductory instructions and allowed sufficient time to read these instructions.[7] Then a Verifier was selected at random among subjects solely for the purpose of

[7] Appendix A provides complete subject instructions.

verifying that the procedures of the experiment were carried out according to the instructions. The Verifier was paid a fixed amount for this task and did not participate in the decision-making task.

Each subject was assigned to one of two treatments depending on whether her seat number was even or odd. One of the treatment groups was then taken out of the lab for a few minutes, always under the supervision of an experimenter. The other group remained in the laboratory and went over the treatment-specific instructions with an experimenter. Simultaneously, subjects waiting outside were given instructions to read individually. Then the groups swapped places and the experimenter read the treatment-specific instructions designed for the other group. Once all instructions were finished, and both groups were brought together in the room again, and we proceeded with the remainder of the experiment.

We implement the two between-subjects treatments within each session so that both groups are presented with the same realized session-specific stimuli (i.e., the particular composition of the bingo cage) in any given session. 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. We did not want to explain P-specific instructions in the presence of M subjects and tell them not to pay attention, and *vice versa*. Subjects are told the reason for this step using the following language:

> 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.

An important reason to assign subjects to treatments according to their station number in the laboratory is to avoid potential confounds due to subjects in each treatment having very different vantage points from which to observe the stimuli. By mapping even or odd station numbers to treatment M or P, we ensure that if there exist any difference in subjects' vantage point, this difference was the same across treatments.

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.[8] 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.

Bingo Cage 2 was then covered and placed on a platform in the front of the room. Bingo Cage 2 was then 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. 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

---

[8] When shown in public, Bingo Cage 1 and 2, were always displayed always 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 generally equivalent view of the bingo cages.

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 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. Then subjects 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 in the task. Finally, subjects left the laboratory and were privately paid their earnings: a $7.50 participation payment in addition to the monetary outcome of the belief elicitation task. The Verifiers were paid a flat $25 fee plus the participation fee. Subjects on average earned approximately $45.60 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 confounds in belief elicitation tasks. In particular, 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 is equally likely, thus minimizing any prior beliefs that particular compositions may be more likely.

We used software we created in *Visual Basic .NET* 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 B, respectively, which are variants on the "slider interface" proposed by Andersen, Fountain, Harrison and Rutström [2010]. 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 the figures 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.

B. Evaluation of Hypothesis

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

One way of testing our hypothesis is to compare observed behavior across the two treatments. We take advantage of the implications of observed behavior in scoring rules of risk averse individuals versus behavior of risk neutral individuals. If subjects are indeed risk averse and the BLP does induce linear utility, then subjects in treatment M should be drawn on average to make reports closer to 50 than subjects in treatment P. This implies that, depending on the true underlying subjective probability, the average report in treatment M would be smaller or greater than in treatment B in such a manner that the former is always closer to 50. To make this test operational, we calculate a measure of distance between each report and the middle of the report interval. For example, if a subject made a report of 30 for red balls, then the measure of report distance is the absolute value of the difference which is 20 (= |30-50|). 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 might be very likely in situations where Bingo Cage 2 has a composition of red and white balls 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 have an identification problem. This did not happen in any of the sessions, but was a risk in this design, of course the risk was just wasted subject fees and time.

An ideal test of our hypothesis would involve comparing the reports in treatment P with the underlying subjective probabilities of subjects in treatment M, and we do this in the structural econometric test. However, since subjective probability is an unobserved latent variable, we can use for now the true proportion of red balls in the Bingo Cage 2 that subjects actually faced in each session as a proxy for the underlying subjective probability.[9] If the BLP does induce linear utility (and subjects are risk averse), subjects in treatment P should make reports on average that are closer to the true number of red balls in Bingo Cage 2 than subjects in treatment M. To operationalize this test, we use also 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 also refer to this measure of distance as a report distance. For example, if a subject made a report of 30 and the correct number of red balls in the Bingo Cage 2 he faced was 25, then the measure of report distance is the absolute value of the difference which is 5 (= |30-25|). The comparison across treatments of this measure of 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 perceptual accuracy of subjects.[10]

We pool data across sessions, but also analyze data from each individual session. We apply non-parametric tests to the distance measures to test if there is support for our main hypothesis that the BLP induces linear utility in the belief elicitation task.

Finally, we estimate a structural model that *jointly* estimates risk attitudes and the underlying

[9] This assumption implies that on average subjects have the right idea about the number of red balls in Bingo Cage 2. In the absence of an estimate, this is a natural proxy for the underlying subjective probability.

[10] 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.

subjective probabilities of subjects in treatment M, and then we compare them with the reports in treatment P. If the BLP does induce linear utility then the risk-attitudes-adjusted subjective probabilities in the M treatment should be equal to the average report of subjects in treatment P. The reason is that if BLP does induce risk neutrality in the P subjects, they should directly report their true underlying subjective probability which should be equal to the underlying estimated probabilities for M subjects.

3. Results

A. Does the BLP Mitigate 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. Our sample size is almost evenly distributed among treatments: 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 session. Figure 5 displays, again by session, 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 a subject who gave an idiosyncratic report of 100 red balls, the average report for treatment P in session 1 is decreased to 26.8.[11] 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 of the 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

[11] Although this subject's reporting behavior was certainly puzzling and idiosyncratic, it can still be rationalized by non-EU preferences. In particular, an appropriate combination of probability weighting function that violates the assumptions outline in section 1 and subjective beliefs can provide incentives to subjects to make a report close to 100. A more simple explanation is that the subject had a strong preference for color red that was not related to the actual configuration of Bingo Cage 2.

treatment P. This feature is not readily seen in the case of sessions 3 and 4, but one can use non-parametric statistics to test the statistical significance of this treatment effect across sessions.

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 absolute value of the difference between reports and 50 is 14.2 and 18.7 for treatments M and P, respectively. A one-sided Fisher-Pitman permutation test for difference in means results in a *p*-value of 0.02, which suggests that on average subjects in treatment M tend to make reports closer to 50. Session 3 had a Bingo Cage 2 with a composition of red and white balls close to 50/50, precisely where we predict this treatment effect test would have low power. Thus we present non-parametric test results on distributions with and without this session included. Figure 6 shows the empirical cumulative distribution of our measure of distance of reports from 50 for sessions 1, 2 and 4 and for all sessions. There is a perceptible difference between the distributions of treatments M and P, especially when only sessions 1,2 and 4 are considered. When we exclude session 3, the one-sided Kolmogorov-Smirnov test results in a *p*-value of 0.02, which supports the hypothesis that subjects in treatment M tend to provide reports closer to 50. 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.

B. Does the BLP Improve Accuracy?

We find evidence that subjects from treatment P tend to make reports closer to the correct number of red balls in Bingo Cage 2, supporting the hypothesis that the BLP induces linear utility in our belief elicitation task. 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. Figure 7 shows the empirical distribution of the absolute value of differences between reports and the correct number of red balls, pooling over all sessions. We see how the cumulative distribution of treatment P is dominated by the distribution of treatment M, which implies that distances are smaller in treatment P. The one-sided Kolmogorov-Smirnov test for two samples results in a *p*-value of 0.04, which supports the hypothesis that distances of reports from the correct number of red balls in treatment P are smaller than in treatment M. Under the assumption that the correct number of red balls in Bingo Cage 2 is a good proxy for the average underlying subjective probability, we find that subjects tend to make reports closer to the correct number of red balls. This could be interpreted as better accuracy from the part of subjects in treatment P. However we interpret this observed behavior as a result of the BLP: this procedure induces linear utility in subjects and provides incentive to reveal the true latent subjective probability, thus mitigating the distortion in reports introduced by risk attitudes.

C. Do Subjects Report their Underlying Subjective Probabilities with the BLP?

We follow Andersen, Fountain, Harrison and Rutström [2010] and develop a structural econometric model to estimate the underlying subjective probabilities of subjects in the M treatment. We then compare these estimates with the raw reports of subjects in the P treatment. If BLP does induce linear utility, the estimated (risk-attitudes adjusted) subjective probabilities for M subjects should be equal to the average reports of P subjects. We find that once we control for risk attitudes, the underlying subjective probabilities of M subjects are statistically 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. We use choices from a binary lottery choice task under risk [12] to identify risk attitudes, and the reports of subjects in the M treatment in the belief elicitation task to then identify the subjective probabilities. Conditional on EUT and the assumption of a CRRA utility function being the model that characterizes individual decision under risk in our sample,[13] we maximize the *joint* likelihood of observed choices in the risk task and the belief elicitation task. The solution to this maximization yields the values of the risk attitudes parameter and subjective probabilities that best explain observed choices in the belief elicitation task as well as observed choices in the lottery tasks.[14]

In sessions 1 and 4 the estimated subjective probabilities were, respectively, equal to 30.2% with a *p*-value of 0.01 and 70.8% with *p*-value less than 0.001.[15] Each of these estimates is statistically 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 M subjects of 26.8% (75.8%) results in a *p*-value equal to 0.77 (0.67). Similarly, in session 2 a test for the hypotheses that the estimated subjective probability of 62% for M subjects is equal to the P subjects' average report of 65.8 P results in a *p*-value equal to 0.82. Finally, the *p*-value for the equivalent null hypothesis for session 3 is equal to 0.82; however, although consistent with our overall conclusions, the choices from this session are

[12] We use choices from other two experiments (Harrison and Swarthout [2012] and Harrison, Martínez-Correa Swarthout [2012]) that collect responses to binary choices between lotteries with objective probabilities. As in our belief elicitation task, subjects in these two studies made one, and only one, choice and was paid for it. Subjects in all tasks were sample from the same population. Payoffs were roughly the same.

[13] 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.

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

[15] 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.

not informative because subjects in both treatments had strong incentives to make a report of 50 given that the stimuli was very close to this number.

4. Conclusions

Harrison, Martínez-Correa and Swarthout [2013] found that the binary lottery procedure works robustly to induce risk neutrality when subjects are given one risk task defined over objective probabilities and the evaluation of the hypothesis does not depend on the assumed validity of any strategic equilibrium behavior, or even the customary independence axiom. Using individuals sampled from the same pool of subjects, we find evidence that the Binary Lottery Procedure induces linear utility in a belief elicitation task when using the Quadratic Scoring Rule and presenting only one question.

First, we observe that subjects who do not use the BLP tend on average to make reports closer to the middle of the report interval, which reduces the uncertainty involved in the belief elicitation task. Second, we also see that subjects who do use the BLP tend to make reports closer to the correct number of red balls in Bingo Cage 2. Finally, we econometrically control for risk attitudes and recover the underlying subjective probabilities of subjects who were not exposed to the BLP and find that they equal to the raw average reports of subjects exposed to the BLP. We interpret these findings as evidence that the BLP induces linear utility in subjects, thus mitigating the distortion in reports introduced by risk attitudes.

An important feature of the BLP is that it theoretically provides incentives for subjects to directly report their underlying latent subjective probability. This applies for individuals with subjective expected utility representations and, under certain weak conditions, for individuals with certain non-EU preference representations as well. In particular, the BLP theoretically works for individuals who follows Rank-Dependent Utility theory.

There are two important extensions of our approach left for future work. First, a natural extension of our approach would be to test if the BLP works with more general scoring rules designed to elicit full distributions for continuous events, such as the generalization of the QSR proposed by Mathieson and Winkler [1976]. Second, the procedures we develop here can be used to test the validity of the reduction of compound lotteries axiom defined over subjective beliefs. This important direction entails some subtle extensions in the experimental design, building on the design employed here.

Figure 1: Binary Scoring Rule Using the Binary Lottery Procedure

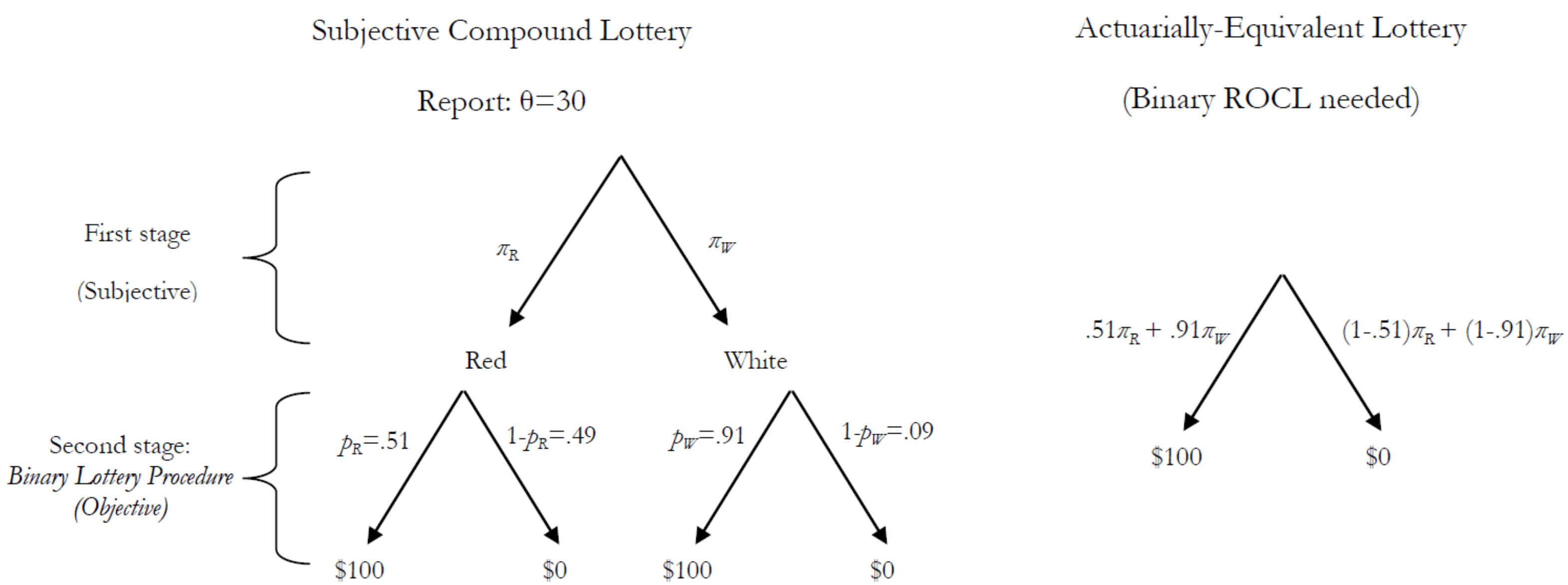


Table 1: Experimental Design

	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

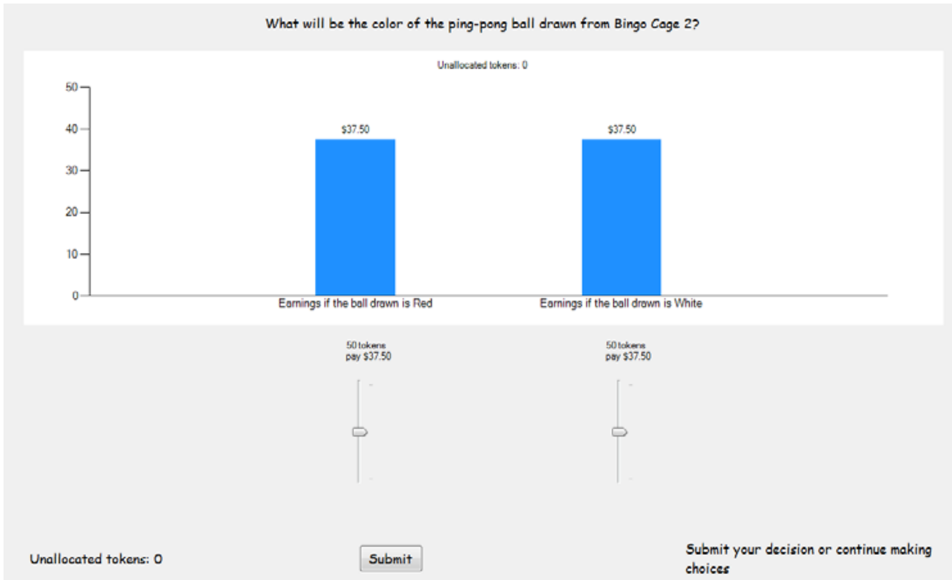


Figure 3: Subject Display for Treatment P

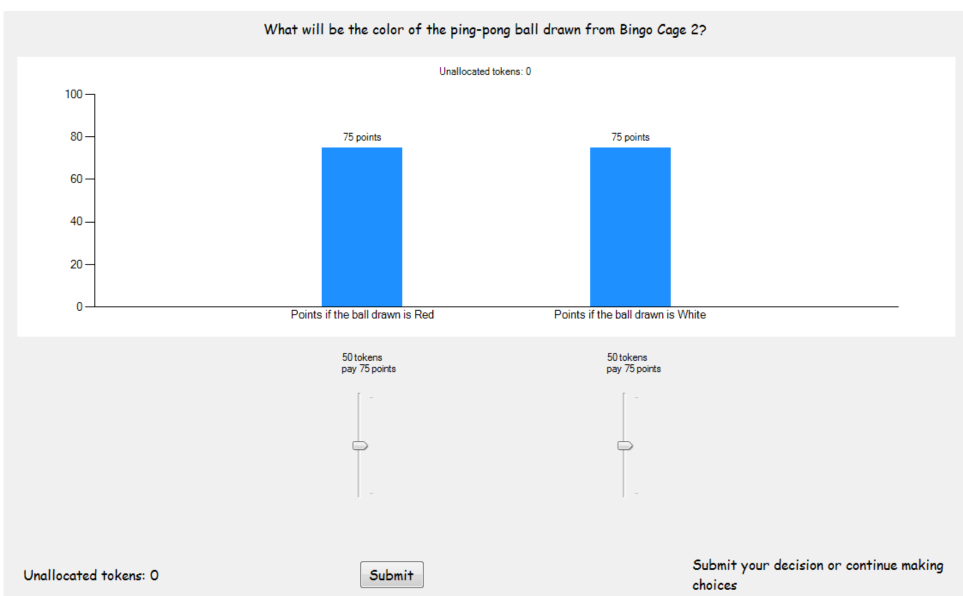


Figure 4: Frequency of Reports by Session

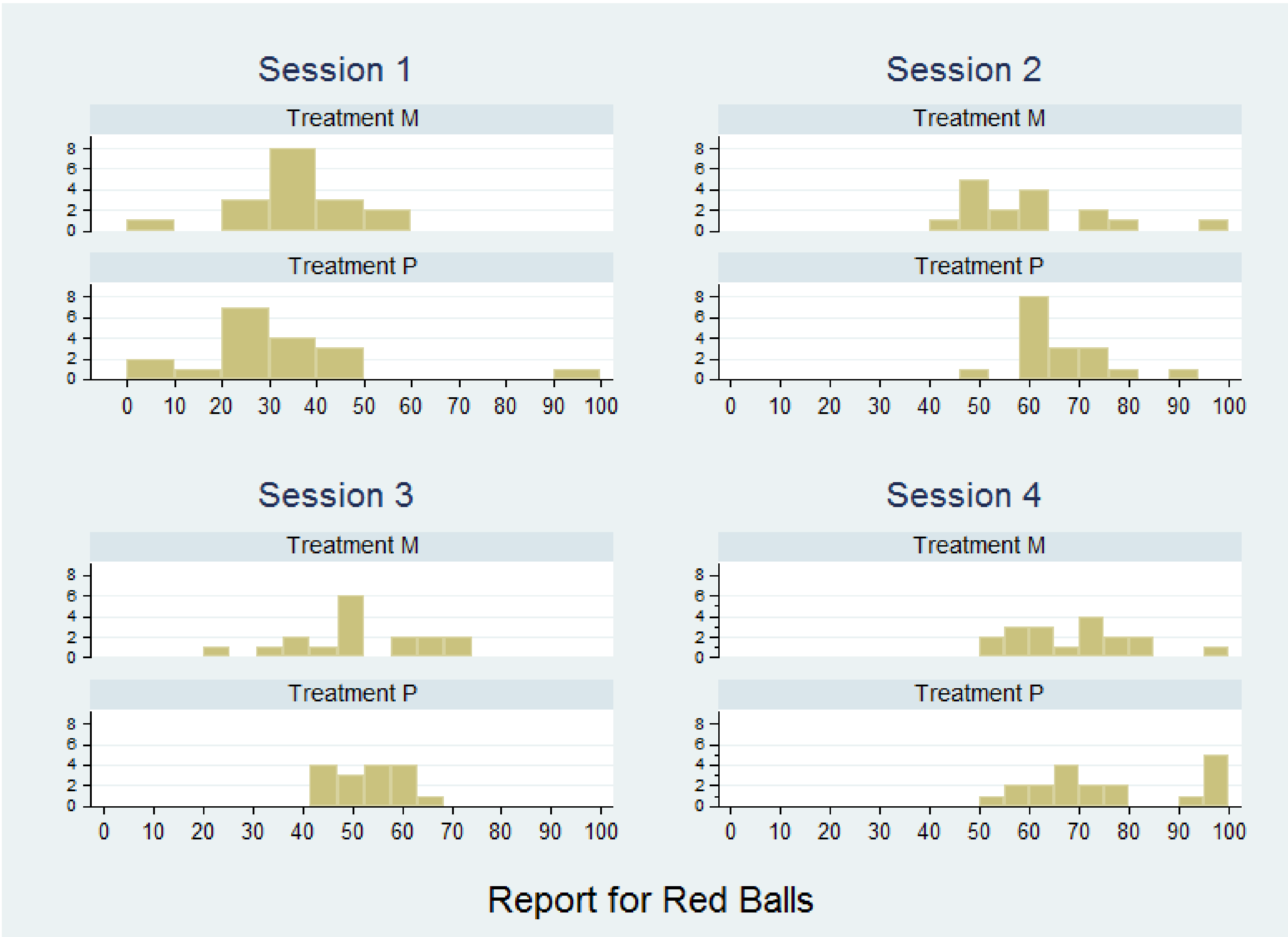


Figure 5: Estimated Densities of Reports by Session

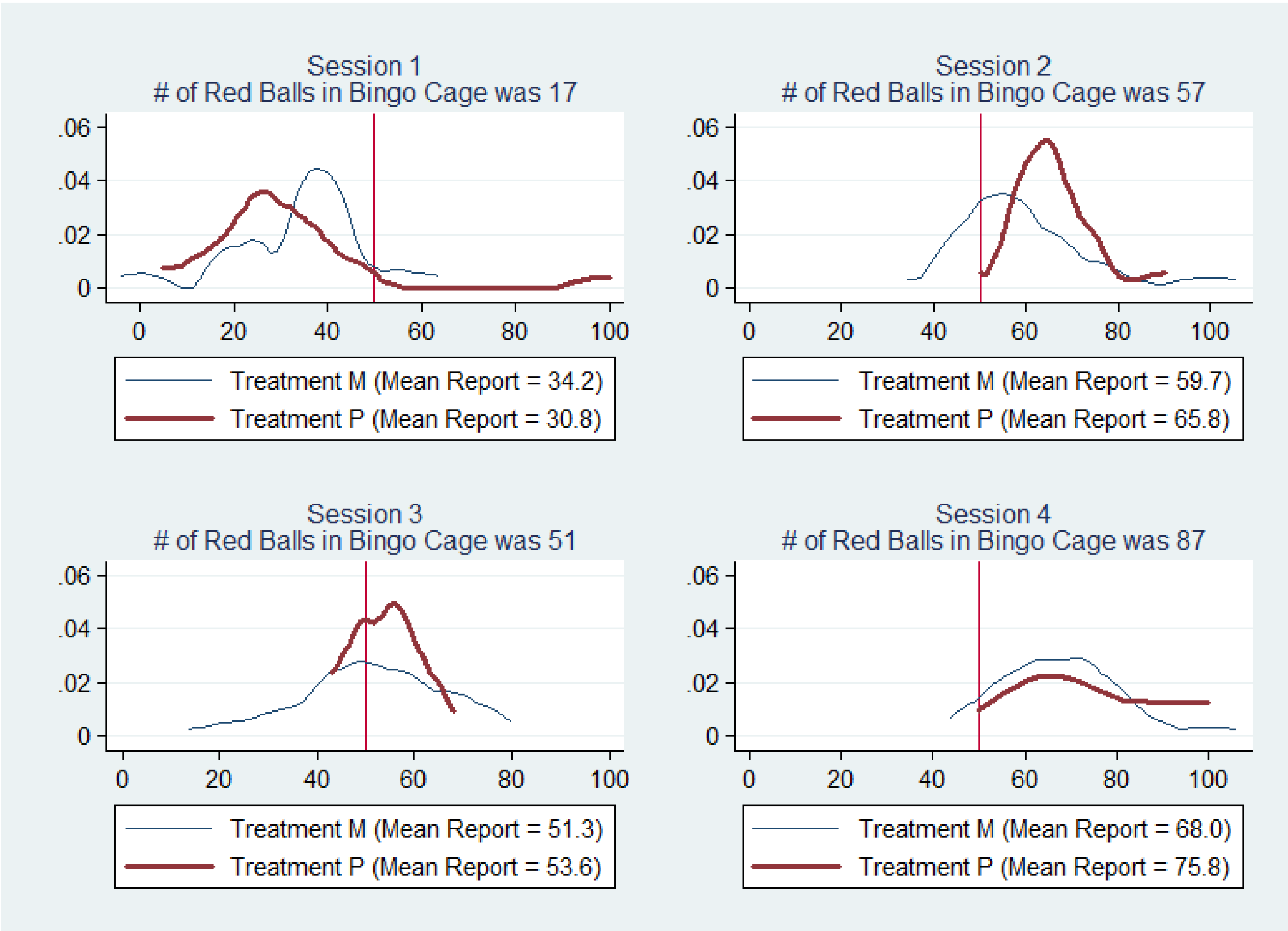


Figure 6: Empirical Cumulative Distribution of Distance of Reports from 5

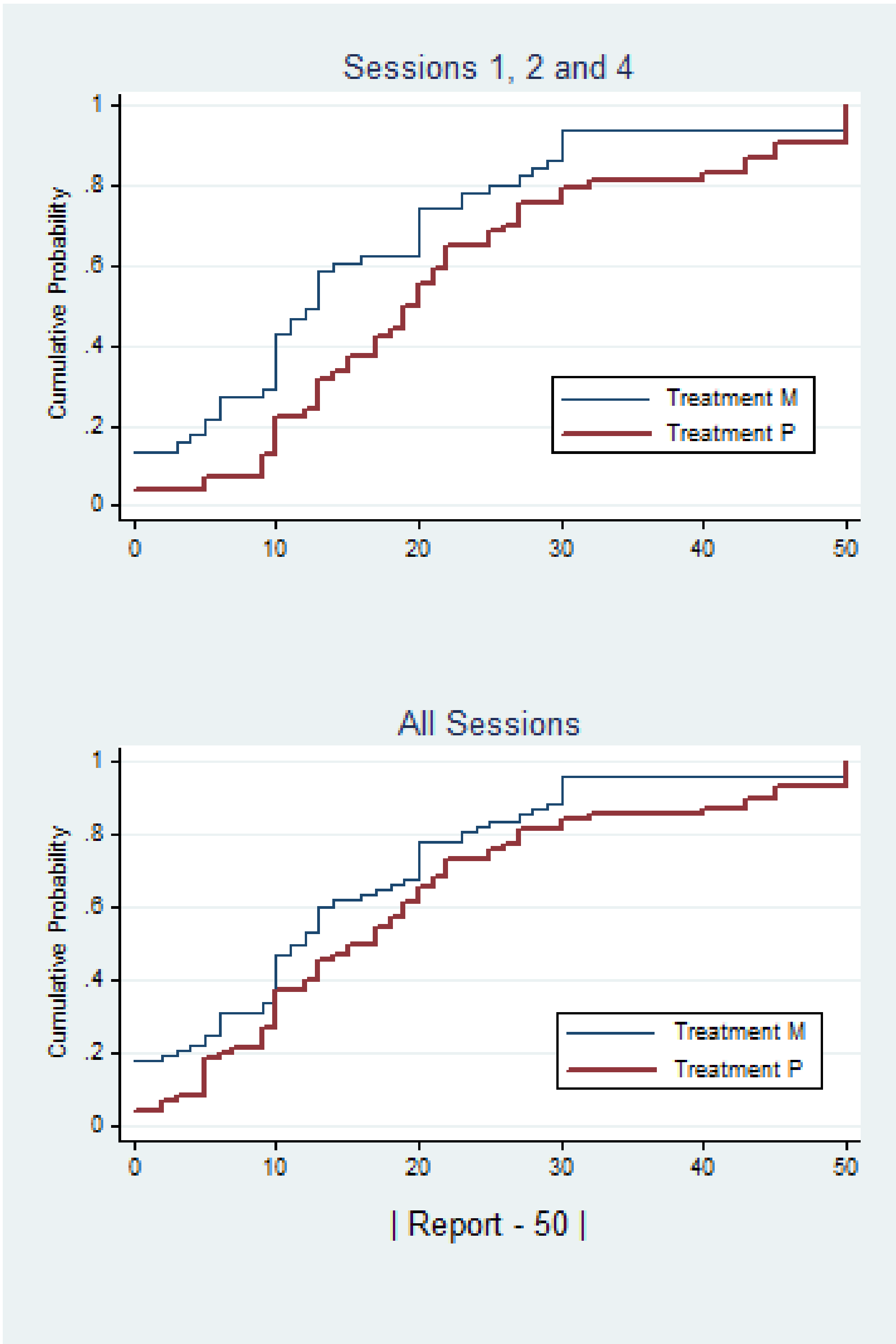
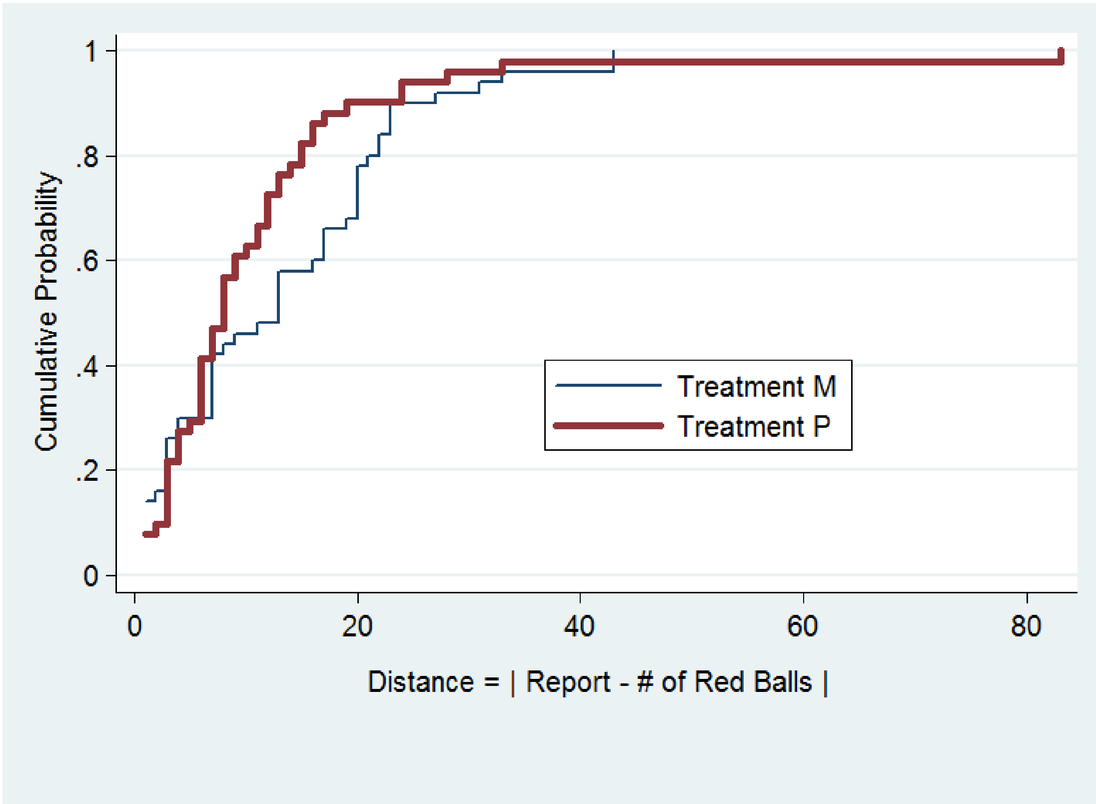


Figure 7: Empirical Cumulative Distribution of Distance Pooling Data From All Sessions



References

Allen, Franklin, "Discovering Personal Probabilities When Utility Functions Are Unknown," *Management Science*, 33(4), 1987, 452-454.

Andersen, Steffen; Fountain, John; Harrison, Glenn W., and Rutström, E. Elisabet, "Estimating Subjective Probabilities," *Working Paper 2010-06*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2010.

Becker, Gordon M.; DeGroot, Morris H., and Marschak, Jacob., "Measuring Utility By A Single-Response Sequential Method," *Behavioral Science*, 9, July 1964, 226-232.

Berg, Joyce E.; Daley, Lane A.; Dickhaut, John W.; and O'Brien, John R., "Controlling Preferences for Lotteries on Units of Experimental Exchange," *Quarterly Journal of Economics*, 101, May 1986, 281-306.

de Finetti, Bruno, "La Prévision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de l'Institut Henri Poincairé*, 7, 1937, 1-68; English translation as "Foresight: Its Logical Laws, Its Subjective Sources," in H.E. Kyburg and H.E. Smokler (eds.), *Studies in Subjective Probability* (New York: John Wiley & Sons, 1964)

de Finetti, Bruno, "Logical Foundations and Measurement of Subjective Probability," *Acta Psychologica*, 34, December 1970, 129-145

Grether, David M., "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence," *Journal of Economic Behavior & Organization*, 17, 1992, 31-57.

Harrison, Glenn W., "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 82, December 1992, 1426-1443.

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

Harrison, Glenn W, Martínez-Correa, Jimmy and Swarthout, J. Todd, "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, 2012.

Harrison, Glenn W, Martínez-Correa, Jimmy and Swarthout, J. Todd, "Inducing Risk Neutral Preferences with Binary Lotteries: A Reconsideration," *Journal of Economic Behavior & Organization*, 2013, forthcoming; available as *Working Paper 2012-05*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012.

Harrison, Glenn, and Swarthout, J. Todd, "The Independence Axiom and the Bipolar Behaviorist," *Working Paper 2012-01*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012.

Holt, Charles A., and Smith, Angela M., "An Update on Bayesian Updating," *Journal of Economic Behavior & Organization*, 69, 2009, 125-134.

Hossain, Tanjim and Okui, Ryo, "The Binarized Scoring Rule," *Working Paper*, University of Toronto, Rotman School of Management, 2012.

Kadane, J. B. and Winkler, Robert L., "Separating Probability Elicitation from Utilities," *Journal of the American Statistical Association*, 83(402), 1988, 357-363.

Karni, Edi, "A Mechanism for Eliciting Probabilities," *Econometrica*, 77(2), March 2009, 603-606.

Köszegi, Botond, and Rabin, Matthew, "Revealed Mistakes and Revealed Preferences," in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook* (New York: Oxford University Press, 2008).

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

McKelvey, Richard D., and Page, Talbot, "Public and Private Information: An Experimental Study of Information Pooling," *Econometrica*, 58(6), 1990, 1321-1339.

Offerman, Theo; Sonnemans, Joep; van de Kuilen, Gijs, and Wakker, Peter P., "A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *Review of Economic Studies*, 76(4), 2009, 1461-1489.

Roth, Alvin E., and Malouf, Michael W. K., "Game-Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 86, 1979, 574-594.

Rutström, E. Elisabet, and Wilcox, Nathaniel T., "Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test," *Games and Economic Behavior*, 67, 616-632.

Savage, Leonard J., *The Foundations of Statistics* (New York: John Wiley and Sons, 1954)

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

Schlag, Karl H., and Joël van der Weele, "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk-neutrality," *Working Paper*, Universitat Pompeu Fabra, Barcelona, 2009.

Smith, Cedric A.B., "Consistency in Statistical Inference and Decision,." *Journal of the Royal Statistical Society*, 23, 1961, 1-25.

Trautmann, Stefan T., and Gijs van de Kuilen, "Belief Elicitation: A Horse Race among Truth Serums," *CentER Working Paper 2011-117*, Tilburg University, November 2011.

Winkler, Robert L., "Probabilistic Prediction: Some Experimental Results," *Journal of the American Statistical Association*, 66(336), December 1971, 675-685.

Winkler, Robert L. And Murphy, Allan H., "Nonlinear Utility and the Probability Score," *Journal of Applied Meteorology*, 9, February 1970, 143-148.

Appendix A: Instructions (NOT FOR PUBLICATION)

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 <u>Odd group</u>. If you are sitting in a computer station that has an even number on it, you are part of the <u>Even group</u>.

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.

B. Instructions for Treatment M

2m

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.

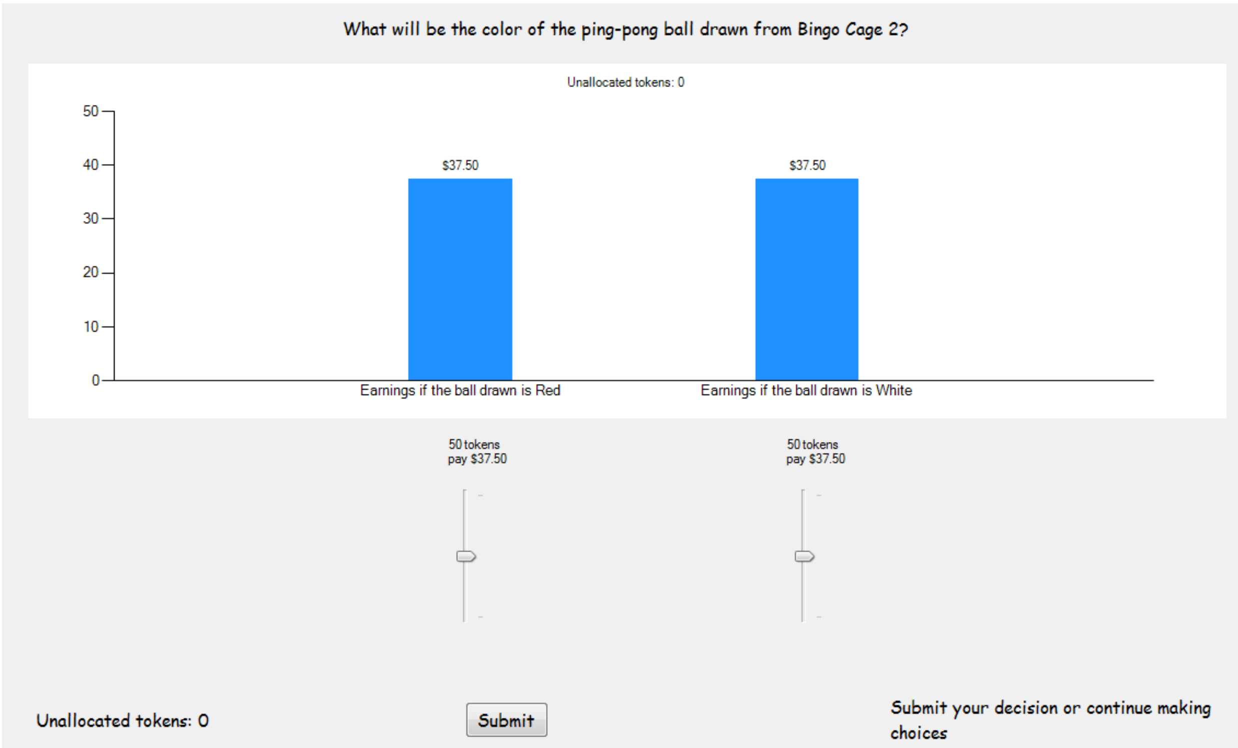
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab 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.

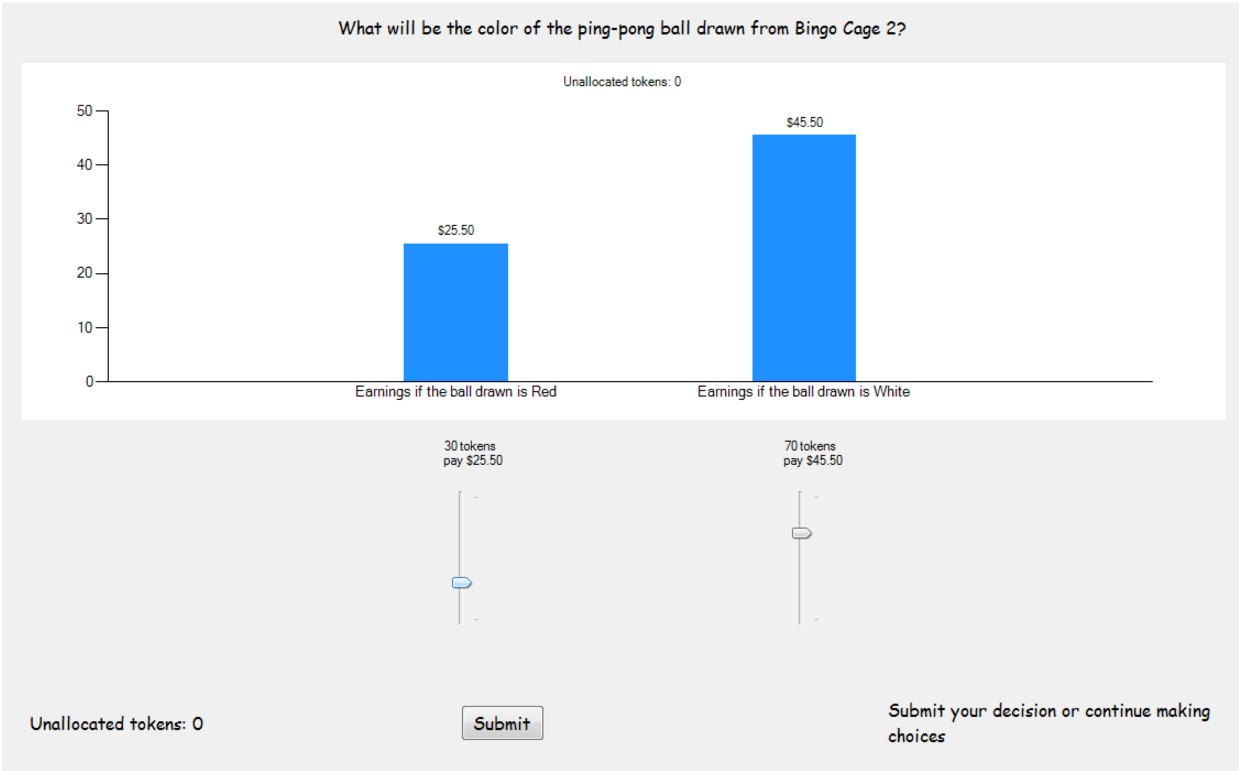


The display on your computer will be larger and easier to read. You have 2 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we 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.

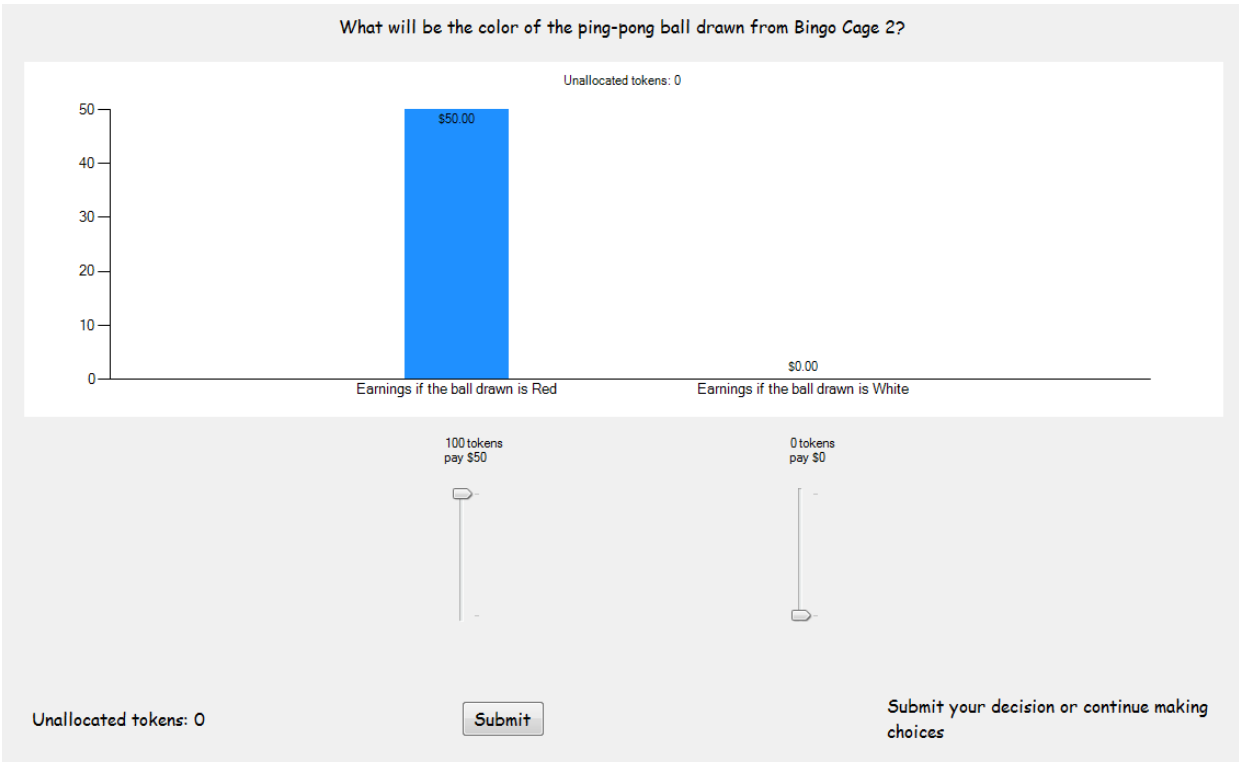
To illustrate how you use these sliders, suppose you think there is a fair chance that there are less red balls than white balls in Bingo Cage 2. Then you might allocate 30 tokens to the first bar, as shown below. Notice that the second bar will be automatically adjusted depending on the number of tokens 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 pickone 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?

C. Instructions for Treatment P

2p

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 <u>one and only one</u> 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.

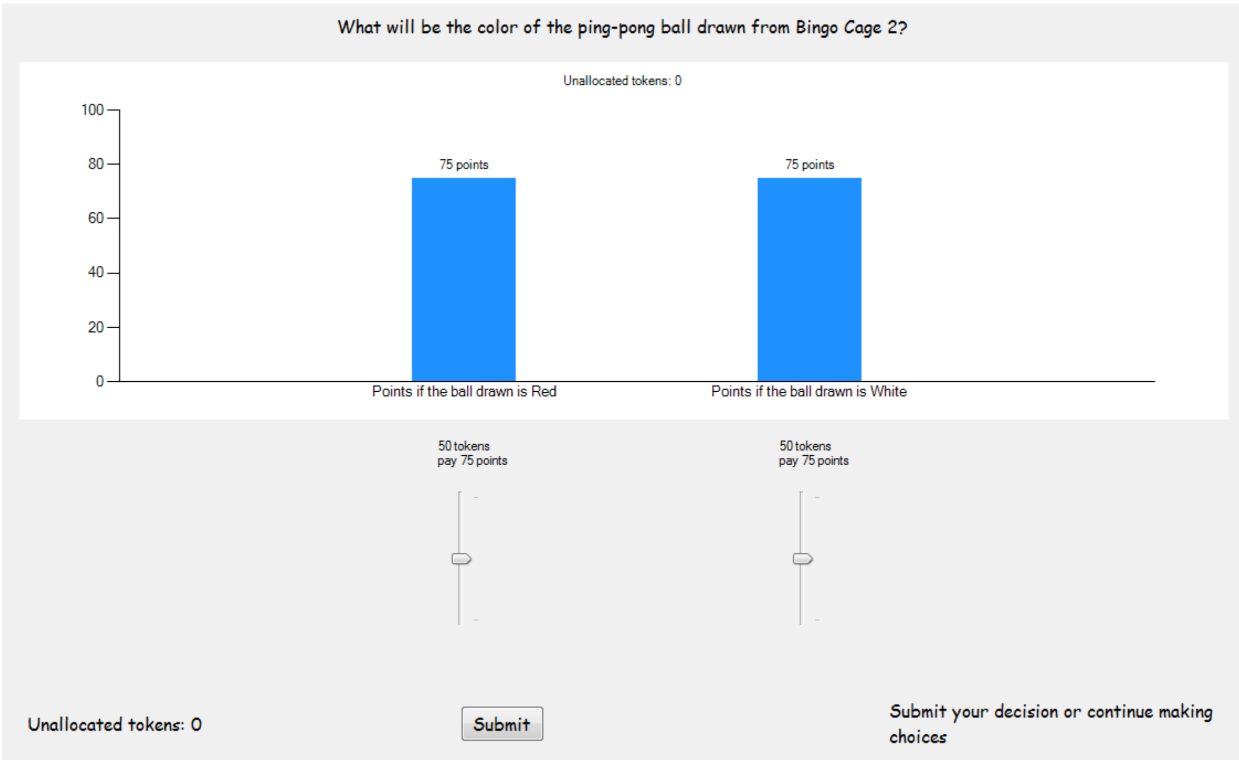
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab 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 <u>total</u> 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.

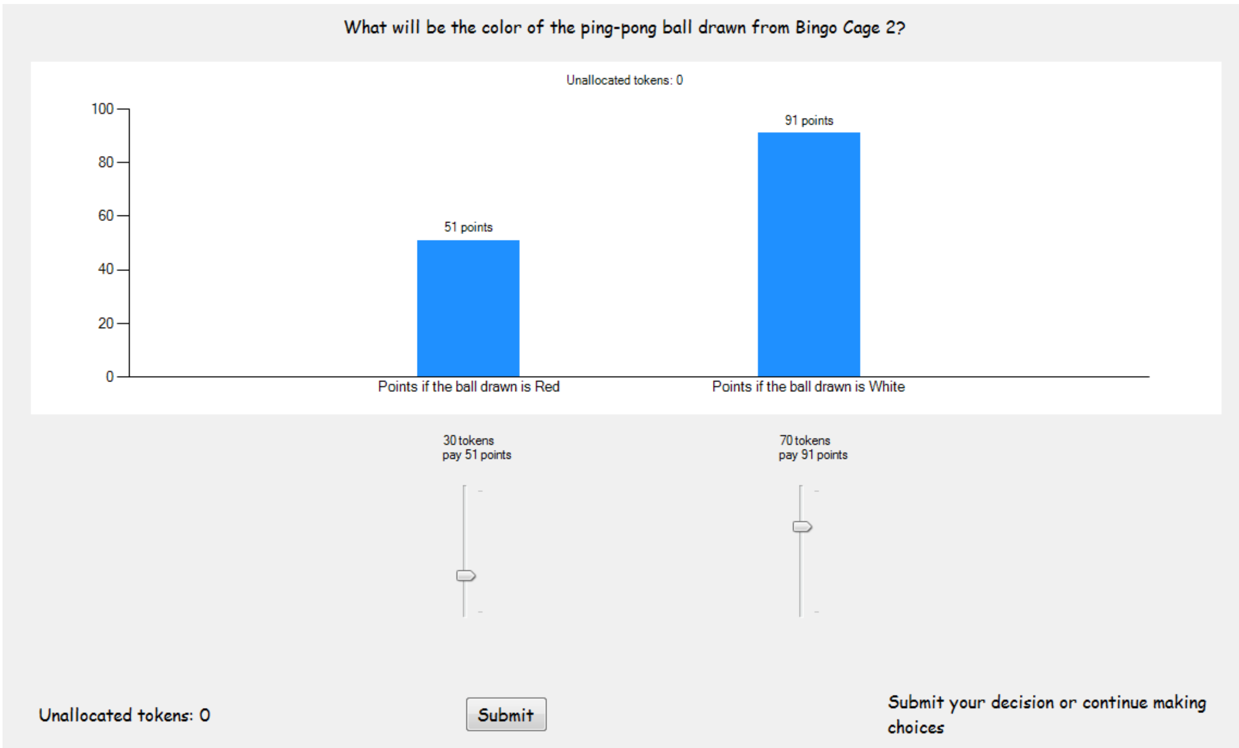


The display on your computer will be larger and easier to read. You have 2 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we 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.

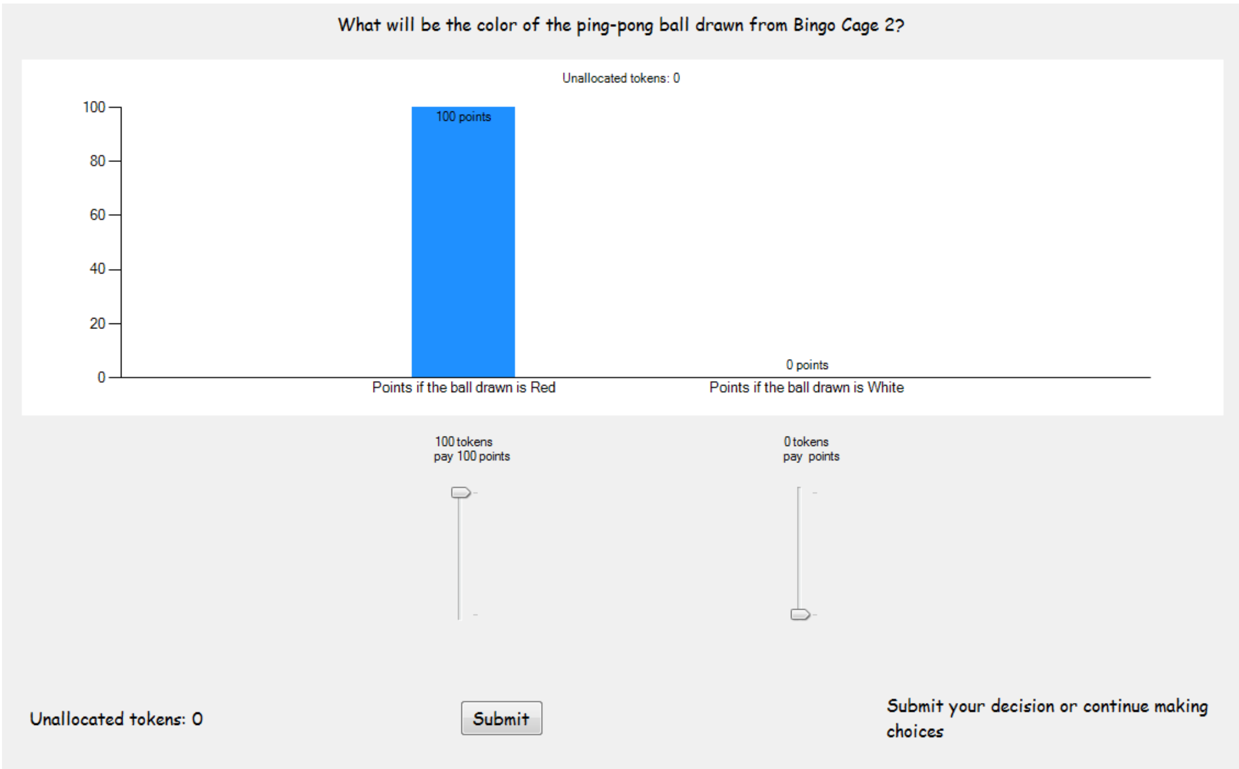


The above pictures show someone who allocated 30 tokens to red Ping-Pong balls and 70 tokens to white Ping-Pong balls. You can adjust this as much as you want to best reflect your personal beliefs about the composition of the bingo cage.

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 (NOT FOR PUBLICATION)

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[16] 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.[17]

Estimation of Risk Attitudes

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

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

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(.) 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:

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

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

$$\nabla EU = EU_R - EU_L \tag{3}$$

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

[16] We use choices from other two experiments (Harrison and Swarthout [2012] and Harrison, Martínez-Correa Swarthout [2012]) that collect responses to binary choices between lotteries with objective probabilities. As in our belief elicitation task, subjects in these two studies made one, and only one, choice and was paid for it. Subjects in all tasks were sample from the same population. Payoffs were roughly the same.

[17] 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.

argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the probit link function,

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

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 indifference[18] the log-likelihood is then

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

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 $\Phi(\nabla EU) > \frac{1}{2}$, 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.[19] 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

$$\nabla EU = ((EU_R - EU_{L,i})/\nu)/\mu, \tag{4'}$$

[18] 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.

[19] 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."

instead of (4), 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 ½. So a larger μ means that the difference in the EU of the two lotteries, conditional on the estimate of r, has less predictive 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 $\mathbf{X}$.[20] 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

$$SEU_\theta = \pi_R \times U(\ \$S(\theta|R)\ ) + (1- \pi_R) \times U(\ \$S(\theta|W)\ ) \quad (5)$$

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

$$eu_\Theta = \exp[(SEU_\theta/\nu)/\mu] \quad (6)$$

for any report θ, and then

$$\nabla EU = eu_\theta/(eu_0 + eu_1 + ... + eu_{100}) \quad (7)$$

for the specific report θ observed, analogously to (4').

We need r to evaluate the utility function in (5), we need π_A to calculate the EU_θ in (5) for each *possible* report θ in {0, 1, 2, ..., 100} once we know the utility values, and we need μ to calculate

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

the latent indices (6) and (7) 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 5 to lie in the interval (0.5, 1), and for sessions 2 and 3 to lie in the interval (0.25, 0.75). These intervals span the vast majority of responses. Results are essentially unchanged if we delete the few outliers from these intervals.

Appendix C: Estimates of Subjective Probabilities (NOT FOR PUBLICATION)

Estimates for Sessions 1 and 4

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

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
r            |
       _cons |   .6068863   .1085074     5.59   0.000     .3942157    .8195568
-------------+----------------------------------------------------------------
muRA         |
       _cons |   .2260858   .0683509     3.31   0.001     .0921204    .3600512
-------------+----------------------------------------------------------------
sprob_1      |
       _cons |  -.4235414    .998688    -0.42   0.671    -2.380934    1.533851
-------------+----------------------------------------------------------------
sprob_4      |
       _cons |   .3460463   .9740016     0.36   0.722    -1.562962    2.255054
------------------------------------------------------------------------------

. nlcom (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))-`sp_points_1')
        (sprob4d: 0.5+(0.5/(1+exp([sprob_4]_cons)))-`sp_points_4')

      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 |   .3021652   .1194007     2.53   0.011      .068144    .5361863
      sprob4 |   .7071708   .1181769     5.98   0.000     .4755483    .9387932
     sprob1d |   .0345181   .1194007     0.29   0.773     -.199503    .2685393
     sprob4d |  -.0512503   .1181769    -0.43   0.665    -.2828727    .1803722
------------------------------------------------------------------------------
```

Estimates for Sessions 2 and 3

```
                                                  Number of obs   =        193
                                                  Wald chi2(0)    =          .
Log likelihood = -257.17834                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
r            |
       _cons |   .6097245   .1204072     5.06   0.000     .3737307    .8457183
-------------+----------------------------------------------------------------
muRA         |
       _cons |   .2571174   .0852141     3.02   0.003     .0901008     .424134
-------------+----------------------------------------------------------------
sprob_2      |
       _cons |  -.4922544   .4694209    -1.05   0.294    -1.412302    .4277937
-------------+----------------------------------------------------------------
sprob_3      |
       _cons |  -.0687132   .3751789    -0.18   0.855    -.8040502    .6666238
------------------------------------------------------------------------------

. nlcom (sprob2: 1/(1+exp([sprob_2]_cons)))
        (sprob3: 1/(1+exp([sprob_3]_cons)))
        (sprob2d: 1/(1+exp([sprob_2]_cons))-`sp_points_2')
        (sprob3: 1/(1+exp([sprob_3]_cons))-`sp_points_3')

      sprob2:  1/(1+exp([sprob_2]_cons))
      sprob3:  1/(1+exp([sprob_3]_cons))
     sprob2d:  1/(1+exp([sprob_2]_cons))-.6576470588235294
      sprob3:  1/(1+exp([sprob_3]_cons))-.53625

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sprob2 |   .6206374   .1105236     5.62   0.000     .4040152    .8372596
      sprob3 |   .5171715   .0936841     5.52   0.000     .3335541     .700789
     sprob2d |  -.0370097   .1105236    -0.33   0.738    -.2536319    .1796125
      sprob3 |  -.0190785   .0936841    -0.20   0.839    -.2026959     .164539
------------------------------------------------------------------------------
```