11-1-2012

Scoring Rules for Subjective Probability Distributions

Glenn Harrison
Georgia State University

Jimmy Martínez-Correa
Copenhagen Business School

Todd Swarthout
Georgia State University

Scoring Rules for Subjective Probability Distributions

by

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

April 2013

ABSTRACT.

The theoretical literature has a rich characterization of scoring rules for eliciting the subjective beliefs that an individual has for continuous events, but under the restrictive assumption of risk neutrality. It is well known that risk aversion can dramatically affect the incentives to correctly report the true subjective probability of a binary event, even under Subjective Expected Utility. To address this one can "calibrate" inferences about true subjective probabilities from elicited subjective probabilities over binary events, recognizing the incentives that risk averse agents have to distort reports. We characterize the comparable implications of the general case of a risk averse agent when facing a popular scoring rule over continuous events, and find that these concerns do not apply with anything like the same force. For empirically plausible levels of risk aversion, one can reliably elicit most important features of the latent subjective belief distribution without undertaking calibration for risk attitudes providing one is willing to assume Subjective Expected Utility.

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, Copenhagen Business School, Denmark (Martínez-Correa); Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout); and Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, USA (Ulm). E-mail contacts: gharrison@gsu.edu, jima.eco@cbs.dk, swarthout@gsu.edu and eulm@gsu.edu. We are grateful to David Gonzalez Larrahondo for research assistance, and to the Society of Actuaries and the Center for Actuarial Excellence Research Fund for financial support.

The theoretical literature has a rich characterization of scoring rules for eliciting the subjective beliefs that an individual has for continuous events, but under the restrictive assumption of risk neutrality. It is well known that, even when behavior is consistent with Subjective Expected Utility (SEU), risk aversion can dramatically affect the incentives to correctly report the true subjective *probability* of a *binary* event.[1] To address this inferential problem, one can "calibrate" inferences about true subjective probabilities from elicited subjective probabilities over binary events, recognizing the incentives that risk averse agents have to report the same probability for the two outcomes and reduce the variability of payoffs from the scoring rule.[2] Or one must use relatively complicated scoring rules that "risk-neutralize" the agent.[3] Or one must eschew the use of any incentives for truthful elicitation.[4]

We characterize the comparable implications of the general SEU case of a risk averse agent when facing a popular scoring rule over *continuous* events, and find that these concerns do not apply with anything like the same force when one is eliciting the subjective belief *distribution*. For empirically plausible levels of risk aversion, which we quantify below, our theoretical results imply that one can reliably elicit the most important features of the latent subjective belief distribution without undertaking calibration for risk attitudes. In particular, under relative weak conditions we can recover the mean of the subjective distribution. This statistic is of particular interest to economists, since it corresponds to the manner in which subjective belief distributions are reduced to a subjective probability under SEU by application of the Reduction of Compound Lotteries

[1] See Winkler and Murphy [1970], Savage [1971; p. 785] and Kadane and Winkler [1988].

[2] See Offerman, Sonnemans, van de Kuilen and Wakker [2009] or Andersen, Fountain, Harrison and Rutström [2010].

[3] See Smith [1961], Grether [1992], Köszegi and Rabin [2008; p.199], Karni [2009] and Holt and Smith [2009] for examples.

[4] Delavande, Gineé and McKenzie [2001; p. 156] make the case for not bothering about incentives. Referring to studies in developing countries that have all been hypothetical, they argue that "even without payment, the answers received from such questions appear reasonable, and as such, there seems to have been a *de facto* decision that payments are not needed." We do not know what "reasonable" might possibly mean when it comes to subjective beliefs.

axiom.

Specifically, we can draw several conclusions. Some of these are very intuitive, but it is important to collect them all formally so that we can identify what characteristics of the inferred subjective distribution derives from what assumptions:[5]

1. The individual *reports* a positive probability for an event only if the individual has a positive subjective probability for the event. So if the individual believes that inflation will never fall below 1.5% per annum, we would never see the individual reporting that it would.
2. If an individual has the same subjective probability for two events, then the reported probabilities for the two events will also be the same if the individual is weakly risk averse. So if the individual attaches a probability of 0.25 to the chance that inflation will be between 1% and 2%, and a probability of 0.25 to the chance that inflation will be between 4% and 5%, the reported probabilities for these two intervals will be the same as well (although typically not 0.25).
3. The converse is true for risk averse subjects, as well as for risk lovers. That is, if we observe two events receiving the same reported probability, we know that the true probabilities are also equal, although not necessarily the same as the reported probabilities.
4. If the individual has a *symmetric* subjective distribution, then the reported mean will be *exactly* the same as the true subjective mean, whether or not the subjective distribution is unimodal. This follows from the previous two results, and is of great significance for tests of the

[5] An excellent example of this approach in experimental economics is the theoretical characterization of bidding behavior in first-price sealed bid auction for independent, private values provided by Cox, Roberson and Smith [1982]. They identified what features of bidding behavior derived from just assuming non-satiation, what features derived from then further assuming expected utility theory, what features derived from then assuming a symmetric Bayesian Nash Equilibrium. This approach allowed a rich experimental design for the purpose of identifying which parts of the final theory might be suspect when behavioral deviations were observed.

Reduction of Compound Lotteries axiom over subjective belief distributions.[6] A testable implication of that axiom is that the individual behaves as if holding a subjective probability equal to the average of some subjective belief distribution. Hence if we simply assume symmetry of the true distribution, a relatively weak assumption in some settings, we can elicit that mean directly.

5. The more risk averse an agent is, the more will their reported distribution resemble a uniform distribution defined on the support of their true distribution. In effect, risk aversion causes the individual to report a "flattened" version of their true distribution.

6. It is possible to bound the effect of increased risk aversion on the difference between the reported distribution and true distribution. This result provides a characterization of the empirical finding that the reported distribution is "very close" to the true distribution for a wide range of empirically plausible risk attitudes.

We illustrate these findings with numerical simulations in section 1, provide general proofs in section 2, and offer some empirical evidence in section 3.

Our findings are limited to agents that are assumed to follow SEU. One would ideally also like to have comparable procedures for eliciting whole distributions under alternative theories of decision making under risk, uncertainty or ambiguity, but this is a challenging task. Most popular alternatives to SEU allow for probability weighting behavior to occur, leading to decision weights that are non-additive (or directly assume non-additive decision weights). This adds a fundamental identification problem when trying to infer subjective beliefs. One can solve it, to some degree, by using methods developed for inferring subjective probabilities for binary events[7], and undertaking

[6] For instance, see the theory of subjective compound lotteries proposed in Nau [2006] and Ergin and Gul [2009] as one way of modeling attitudes towards ambiguous events.

[7] For example, Offerman, Sonnemans, van de Kuilen and Wakker [2009], Karni [2009] or Andersen, Fountain, Harrison and Rutström [2010], who each explicitly consider the case of probabilistically sophisticated non-SEU agents.

multiple elicitations and inferences to slice up the subjective distribution. For instance, instead of eliciting the complete distribution for the percentage return on the Standard and Poors Index in one year, one could elicit the probability that it is below -5%, between 5% and 0%, between 0% and 5%, and then that it is greater than 5%. This information could be then used directly as coarse subjective belief distribution for the individual, or as the basis for some inferred parametric distribution.[8] However, it obviously adds several "chained" elicitation tasks, and requires calibration of each inference to risk attitudes and/or probability weighting behavior, depending on what model of behavior is assumed. We see our results as complementary to this approach: at the cost of assuming SEU, we argue that one can reliably infer whole subjective distributions in one task.[9] An obvious task for future research is to compare elicited distributions with our approach and elicited "chained" distributions, where one can correct the latter for possible non-SEU behavior. Obviously, given the importance of characterizing subjective belief distributions, we need to know how to do that under SEU before examining the effect of relaxing SEU in one or other way.

1. Examples

Let the decision maker report his subjective beliefs in a discrete version of a Quadratic Scoring Rule (QSR) for continuous distributions (Matheson and Winkler [1976]).[10] Partition the domain into K intervals, and denote as r_k the report of the density in interval $k = 1, ..., K$. Assume for the moment that the decision maker is risk neutral, and that the full report consists of a series of

---

[8] An excellent example is the evaluation of the *Survey of Economic Expectations* responses on equity returns in Dominitz and Manski [2011; §2.2], although they do not consider the effects of risk aversion and probability weighting on inferences.

[9] What is "one" task and what is "ten tasks on one computer screen" is a semantic matter we do not want to be overly dogmatic about.

[10] Alternative scoring rules could be characterized, but the QSR is the most popular in practice, and virtually all of the same issues apply with varying force. For instance, Andersen, Fountain, Harrison and Rutström [2010] show that behavior under a Linear Scoring Rule and QSR are behaviorally identical when applied to elicit subjective probabilities for binary events *and* one undertakes calibration for the different effects of risk aversion and probability weighting on the two types of scoring rules.

reports for each interval, $\{ r_1, r_2, ..., r_k ,..., r_K \}$ such that $r_k \geq 0\ \forall k$ and $\sum_{i = 1...K} (r_i) = 1$.

If k is the interval in which the actual value lies, then the payoff score is from Matheson and Winkler [1976; p.1088, equation (6)]:

$$S = (2 \times r_k) - \sum_{i = 1...K} (r_i)^2$$

So the reward in the score is a doubling of the report allocated to the true interval, and the penalty depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval the punishment is severe. The punishment includes all possible reports, including the correct one.

Take some examples, assuming $K = 4$. What if the subject has very tight subjective beliefs and puts all of the tokens in the correct interval? Then the score is

$$S = (2 \times 1) - (1^2 + 0^2 + 0^2 + 0^2) = 2 - 1 = 1,$$

and this is positive. But if the subject has a tight subjective belief that is wrong, the score is

$$S = (2 \times 0) - (1^2 + 0^2 + 0^2 + 0^2) = 0 - 1 = -1,$$

and the score is negative. So we see that this score would have to include some additional "endowment" to ensure that the earnings are positive.[11] Assuming that the subject has a very diffuse subjective belief and allocates 25% of the tokens to each interval, the score is less than 1:

$$S = (2 \times ¼) - (¼^2 + ¼^2 + ¼^2 + ¼^2) = ½ - ¼ = ¼ < 1.$$

So the tradeoff from the last case is that one can always ensure a score of ¼, but there is an incentive to provide less diffuse reports, and that incentive is the possibility of a score of 1.

To ensure complete generality, and avoid any decision maker facing losses, allow some endowment, α, and scaling of the score, β. We then get the generalized scoring rule

$$\alpha + \beta\, [\, (2 \times r_k) - \sum_{i=1...K} (r_i)^2\,]$$

[11] This is a point of practical behavioral significance, but is not important for the immediate theoretical point.

where we initially assumed α=0 and β=1. We can assume α>0 and β>0 to get the payoffs to any level and units we want. Let p_k represent the underlying, true, latent subjective probability of an individual for an outcome that falls into interval *k*.

Figures 1 through 6 illustrate the behavior of this scoring rule for the case in which $K = 10$, $\alpha = \beta = 25$, and we assume a subjective expected utility maximizer with a CRRA utility function $u(w) = w^{1-\rho}/(1-\rho)$ such that $\rho = 0$ denotes risk neutrality and $\rho > 0$ risk aversion.[12] Figure 1 shows the simplest case in which the true subjective distribution is symmetric. The histogram always shows the true distribution, and the black "droplines" always show the optimal report. Under risk-neutrality, Figure 1 shows that the individual truthfully reports the true subjective distribution.

Figure 2 considers the more realistic case in which the agent is risk averse, and in fact at parameter values typical of those found in the experimental laboratory (see Harrison and Rutström [2008] for a review). Here we observe that relative risk aversion of $\rho = 0.65$ causes the individual to under-report the true subjective probability for outcome 4 ($r_4 = 0.356 < p_4 = 0.4$). Although barely noticeable to the naked eye, the individual over-reports the true subjective probability for outcomes 3 and 5 ($r_3 = r_5 = 0.207 > p_3 = p_5 = 0.2$). The over-reporting for outcomes 2 and 6 is noticeable ($r_2 = r_6 = 0.115 > p_2 = p_6 = 0.1$). Since the extent of the reporting deviations are the same either side of the mode, and the true distribution is symmetric, the average of the reported distribution would

[12] Utility is defined solely over the income generated by the scoring rule. If utility is event-dependent then one must assume away any effects of the subjective outcome on initial wealth (Kadane and Winkler [1988], Karni and Safra [1995]). In our experiments this is natural, since subjects are betting on the outcome of a draw from an urn that has no connection to events outside the lab, other than the income these bets might generate. In field applications of these scoring rules this assumption might not be so natural. For instance, one might be eliciting beliefs about housing prices from somebody that already owns a house, so that the possible events affect the value of the initial endowment the individual has before any income from the scoring rule. Or preferences themselves might be state-dependent, quite apart from any effect on the arguments of the utility function: different health outcomes, over which one might naturally have subjective beliefs, might affect the utility associated with given endowments. Finally, scoring rules might be embedded in a competitive environment in which performance relative to others becomes a factor. This can lead to an additional distortion of reports (Lichtendahl and Winkler [2007]).

always equal the average of the true distribution.

Figure 3 considers the same symmetric subjective distribution and a wider range of risk attitudes. The trend toward the reported distribution being a flattened version of the subjective distribution, as ρ increases from 0 up to 3 is apparent. Also apparent is the complete absence of any reports for outcomes 1, 7, 8, 9 and 10, which have no subjective density.

Figure 4 considers the case of an asymmetric, unimodal subjective distribution, and varying levels of risk aversion. For relative risk aversion level $\rho > 0$, the true probabilities for outcomes 6 and 5 are under-reported, and for outcomes 4 and 3 are over-reported. Again, there are no reports for outcomes that have no subjective density.

Figure 5 shows the case of a bimodal distribution which is symmetric around each mode. The behavior is qualitatively the same as the symmetric unimodal distribution, and always will be providing there is a zero subjective density outcome between the two symmetric modes.

Finally, using the parameters and beliefs from Figure 4, Figure 6 shows how the average of the reported distribution deviates from the average of the true subjective distribution in the unimodal, asymmetric case. For a wide range of risk attitudes observed in the same experimental context that we would undertake these belief elicitations ($\rho < 1$), we find the difference to be less than a percentage point. Of course, there is no point showing comparable figures for the symmetric distributions, since in that case there is no difference at all.

The preceding discussion used numerical simulations to provide visual and descriptive evidence of our results. We now formalize our results with the theory in the following section.

2. Theory

We focus on the finite case, in part for expository reasons, but also because this is the interesting case in terms of operational scoring rules. The proofs for the continuous case are similar, and collected in Appendix A. Again, the following proofs assume that the decision maker is a SEU maximizer.

Lemma 1: Let p_k represent the underlying subjective probability of an individual for outcome k and let r_k represent the reported probability for outcome k in a given scoring rule. Let $w(k) = \alpha + \beta 2r_k - \beta \sum_{i=1...K} (r_i)^2$ be the scoring rule that determines the wealth if state k occurs. If the individual has a utility function $u(w)$ that is continuous, twice differentiable, increasing and concave and maximizes expected utility over actual subjective probabilities, the actual and reported probability must obey the following system of equations:

$$p_k \times \partial u/\partial w \,|_{w=w(k)} - r_k \times E_p[\partial u/\partial w] = 0, \forall\, k = 1,..., K \tag{1}$$

Proof. Suppose a subjective discrete probability distribution $\{p_1, p_2 ,..., p_k ,..., p_K\}$ over K states of nature and utility function $u(w)$ over random wealth. If the subject is given a scoring rule determined by $w(k) = \alpha + \beta 2r_k - \beta \sum_{i=1...K} (r_i)^2$, then the optimal report $r = \{r_1, r_2,..., r_k ,..., r_K\}$ solves the following problem:

$$\mathbf{Max}_{\{r\}} \; E_p[\, u(w)\,] \text{ subject to } \sum_{i=1...K} (r_i) = 1 \tag{2}$$

where $E_p[\, u(w)\,] = \sum_{j=1...K} p_j \times u[\, \alpha + \beta 2r_j - \beta \sum_{i=1...K} (r_i)^2\,]$. This problem can be solved by maximizing the Lagrangian

$$\mathcal{L} = \sum_{j=1...K} p_j \times u[\, \alpha + \beta 2r_j - \beta \sum_{i=1...K} (r_i)^2\,] - \lambda\, [\, \sum_{i=1...K} (r_i) - 1] \tag{3}$$

The solution to the problem must satisfy K+1 conditions. The K first order conditions with respect to report r_k, $\forall\, k = 1,..., K$, are

$$\partial\mathcal{L}/\partial r_k = \sum_{j=1...K} (\, p_j \times \partial u(w(j))/\partial r_k\,) - \lambda = 0, \forall\, k = 1,..., K \tag{4}$$

where $\partial u(w(j))/\partial r_k = \partial u/\partial w \,|_{w=w(j)} \times (2\beta\delta_{jk} - 2\beta \times r_k)$ and δ_{jk} is equal to 1 if $j = k$ and equal to zero if $j \neq k$. The (K+1)-*th* condition is the first order derivative of (3) with respect to the Lagrangian constant:

$$\Sigma_{i=1\ldots K}(r_i) - 1 = 0. \tag{5}$$

We can simplify the K equations in (4) as:

$$2\beta p_k \times (\partial u/\partial w\,|_{w=w(k)}) - 2\beta r_k \Sigma_{j=1\ldots K}\ p_j \times (\partial u/\partial w\,|_{w=w(j)}) - \lambda = 0, \forall\ k = 1,\ldots, K.$$

or
$$p_k \times (\partial u/\partial w\,|_{w=w(k)}) - r_k\, E_p\,[\,\partial u/\partial w\,] = \lambda/2\beta, \forall\ k = 1,\ldots, K. \tag{4′}$$

Summing over the K first-order conditions we get

$$E_p\,[\,\partial u/\partial w\,|_{w=w(k)}\,] - \Sigma_{k=1\ldots K}\, r_k\, E_p\,[\,\partial u/\partial w\,] = K\lambda/2\beta\,. \tag{6}$$

Notice that $\Sigma_{k=1\ldots K}\, r_k\, E_p\,[\,\partial u/\partial w\,] = E_p\,[\,\partial u/\partial w\,]$ because the expectation term is a constant and because of (5). Then (6) implies that $K\lambda/2\beta = 0$, which can only be satisfied if $\lambda = 0$. This result and (4′) implies that the solution to problem (2) must satisfy the following K conditions:

$$p_k \times \partial u/\partial w\,|_{w=w(k)} - r_k \times E_p[\partial u/\partial w] = 0, \forall\ k = 1,\ldots, K. \qquad \blacksquare$$

Lemma 2: Under the condition in Lemma 1, let $\boldsymbol{\varepsilon}_k = r_k - p_k$ be the deviation between the reported and actual subjective probability for outcome k. Then

$$\boldsymbol{\varepsilon}_k = p_k \times \{\partial u/\partial w\,|_{w=w(k)} - E_p[\,\partial u/\partial w\,]\}/E_p[\,\partial u/\partial w\,], \forall\ k=1,\ldots, K. \tag{7}$$

Proof. Assume that the conditions of Lemma 1 in (1) are satisfied and the distortions between the actual and reported probabilities are given by $r_k = p_k + \boldsymbol{\varepsilon}_k$, with $\Sigma_{k=1\ldots K}\, \boldsymbol{\varepsilon}_k = 0$. Define f_k = $\partial u/\partial w\,|_{w=w(k)}$ and $f = \partial u/\partial w$). Then the K conditions in (1) become:

$$p_k \times f_k - p_k \times E_p[f] - \boldsymbol{\varepsilon}_k \times E_p[f] = 0, \forall\ k = 1,\ldots, K \tag{1′}$$

Solving for $\boldsymbol{\varepsilon}_k$ we get the K conditions stated in Lemma 2:

$$\boldsymbol{\varepsilon}_k = p_k \times \{f_k - E_p[f]\}/E_p[f], \forall\ k = 1,\ldots, K. \qquad \blacksquare$$

Lemma 3: Assume an individual that has a continuous, twice differentiable utility function $u(w)$ that is increasing in random wealth and who is also risk averse (i.e., $\partial^2 u/\partial^2 w < 0, \forall\ w$). If $p_i = p_j$ for some i and j, then $r_i = r_j$.

The intuition for the result is as follows. Suppose that $\{r_1, r_2, \ldots, r_k, \ldots, r_K\}^*$ is a solution to (2).

Then if $p_i = p_j$ for some i and j, the subject must assign the same weight to reports in states i and j, that is $r_i = r_j$. The proof is then by contradiction.

Proof. Assume that $p_i = p_j$ for some i and j. Now suppose without loss of generality that $r_i > r_j$. By definition of the deviation of subjective and reported probabilities the latter implies that $\boldsymbol{\varepsilon}_i > \boldsymbol{\varepsilon}_j$ because

$$r_i = p_i + \boldsymbol{\varepsilon}_i > r_j = p_j + \boldsymbol{\varepsilon}_j \tag{8}$$

Since $r_i > r_j$, we also know that $w(i) > w(j)$, and by the concavity of $u(.)$ the latter implies that $f_i < f_j$. Therefore

$$\{f_i - E_p[f]\}/E_p[f] < \{f_j - E_p[f]\}/E_p[f]\ . \tag{9}$$

But by Lemma 2, (9) implies that $r_i < r_j$, which is a contradiction. ■

Lemma 3 does not hold for risk loving individuals. The following counterexample proves it. Suppose that $u(w) = w^2$, $p_1 = ½$ and $p_2 = ½$, $\alpha = 0$ and $\beta = 1$.

$$E_p[\,u(w(r_1))\,] = 0.5\ (2\ r_1 - r_1^2 - (1\text{-}r_1)^2\,)^2 + 0.5\ (2\ (1\text{-}r_1) - r_1^2 - (1\text{-}r_1)^2\,)^2$$

$$= 4\ r_1^4 - 8\ r_1^3 + 8\ r_1^2 - 4\ r_1 + 1$$

$$\partial E_p[\,u\,]/\partial r_1 = \ \ 16\ r_1^3 - 24\ r_1^2 + 16\ r_1 - 4.$$

To maximize subjective EU set the first order condition equal to zero, and then check the end points $r_1 = 0$ and $r_1 = 1$. We then have

$$r_1^3 - (1½)\ r_1^2 + r_1 - ¼ = 0$$

$$(r_1 - ½)(r_1^2 - r_1 + ½) = 0.$$

Solving for the real root we get $r_1 = ½$. By reporting $r_1 = ½$, the subjective EU is equal to ¼, while if the report is $r_1 = 1$ or $r_1 = 0$ the subjective EU is equal to 1. Thus symmetry is broken, and the optimal report is $(r_1 = 1, r_2 = 0)$ or $(r_1 = 0, r_2 = 1)$: that is, $p_1 = p_2$ but $r_1 \neq r_2$.

Proposition 1: For the risk-averse individual in Lemma 3, if the subjective distribution is symmetric then the mean of the reported distribution is equal to the mean of the actual subjective distribution.

Proof. A symmetric subjective distribution for random variable y with mean μ is one of two types: odd and even. Take the case of the odd type first. Consider a subjective probability p_k and report r_k, for $k = 1, ..., n$, with n being an odd integer. Let $m = (n+1)/2$ such that the subjective probability p_m is the likelihood that the random variable takes the value of μ. Also let p_{m-i} and p_{m+i} be, respectively, the subjective probability that the random variable takes the value of $\mu-\eta_i$ and $\mu+\eta_i$, for $i = 1...m-1$ and $p_{m-i} = p_{m+i}$.

By Lemma 3, $r_{m-i} = r_{m+i}$

$$E_p[\text{ y }] = \Sigma_{i=1...m-1}\, p_{m-i}\,(\mu-\eta_i) + \Sigma_{i=1...m-1}\, p_{m+i}\,(\mu+\eta_i) + p_m\,\mu \tag{10}$$

$$= \Sigma_{j=1...n}\, p_j\,\mu + \Sigma_{i=1...m-1}\,[\,p_{m-i} - p_{m+i}\,]\,\eta_i = \mu + 0 = \mu$$

and

$$E_r[\text{ y }] = \Sigma_{i=1...m-1}\, r_{m-i}\,(\mu-\eta_i) + \Sigma_{i=1...m-1}\, r_{m+i}\,(\mu+\eta_i) + r_m\,\mu \tag{11}$$

$$= \Sigma_{j=1...n}\, r_j\,\mu + \Sigma_{i=1...m-1}\,[\,r_{m-i} - r_{m+i}\,]\,\eta_i = \mu + 0 = \mu.$$

By (10) and (11) we have that $E_p[\text{y}]-E_r[\text{y}] = 0$. The even case is similar except that $m = n/2$ and the random variable taking a value equal to μ has no weight. ■

Lemma 4: The converse of Lemma 3. Assume an individual with a continuous, differentiable utility function $u(w)$, where risk aversion is not necessary in this case. If $r_i = r_j$ for this individual, then $p_i = p_j$.

Proof. It follows from Lemma 1. If $r_i = r_j$ then

$$p_i\,\partial u/\partial w\,|_{r_i} - r_i\,\mathrm{E}_p[\partial u/\partial w]=0 \text{ and } p_j\,\partial u/\partial w\,|_{r_j} - r_j\,\mathrm{E}_p[\partial u/\partial w]=0. \tag{12}$$

Thus,

$$p_i = r_i\,\mathrm{E}_p[\partial u/\partial w]/\,\partial u/\partial w\,|_{r_i} = r_j\,\mathrm{E}_p[\partial u/\partial w]/\,\partial u/\partial w\,|_{r_j} = p_j\,. \quad ■$$

Proposition 2: For the individual in Lemma 4, if the reported distribution is symmetric then the mean of the reported distribution is equal to the mean of the actual subjective distribution.

Proof. Identical to Proposition 1, with r_k and p_k, $\forall k$, interchanged at all steps. ■

Proposition 3: Assume an individual with a continuous, twice differentiable utility function $u(w)$ that is increasing in w. The individual reports probability $r_k = 0$ if and only if the true subjective probability of the individual for state k is $p_k = 0$.

Proof. Using Lemma 2, if $p_i = 0$, then $\boldsymbol{\varepsilon}_i=0$ and $r_i = 0$. The converse claim follows from Lemma 1: since $\partial u/\partial w\,|_{w=w(k)}$ and $E_p[f]$ are both positive, if $r_k = 0$ then $p_k = 0$. ■

Proposition 4: A risk-averse individual has a reported probability distribution that approaches a uniform distribution over those states where $p_k > 0$ in the following sense: There exists a value p^* for this individual such that if $p_k > p^*$ then $p_k > r_k > p^*$ and if $p_k < p^*$ then $p_k < r_k < p^*$. A risk-loving agent reverses all the conditions.

Proof. We will show that $\exists\, p^*$ such that if $p_k > p^*$ then $p_k > r_k > p^*$ and p^* is the value such that $\partial u/\partial w\,|_{w=w(p^*)} = E_p[\,\partial u/\partial w\,]$. From Lemma 2 we know that

$$\boldsymbol{\varepsilon}_k = p_k \times \{\partial u/\partial w\,|_{w=w(k)} - E_p[\,\partial u/\partial w\,]\}/E_p[\,\partial u/\partial w\,],\ \forall\, k=1,\ldots, K.$$

We also know that $w(k)$ is monotonically increasing in r_k, and therefore $\partial u/\partial w$ is monotonically decreasing in r_k. If $r_k >(<)\, p^*$, $\boldsymbol{\varepsilon}_k<(>)0$, $r_k < (>)\, p_k$ by definition. Since $p^* <(>)\, r_k$, then $p^* <(>)\, r_k <(>)\, p_k$. ■

Proposition 5: An individual with sufficiently high risk aversion will have a reported probability arbitrarily close to p^*.

Proof. By Lemma 2 we know that $r_k = p_k \times \{\partial u/\partial w\,|_{w=w(k)}\}/E_p[\,\partial u/\partial w\,]$. Let p^* be selected such that $\partial u/\partial w\,|_{w=w(p^*)} = E_p[\,\partial u/\partial w\,]$. Then let $u(w) = w - c\,[w - w(p^*)]^2$ without loss of generality. Therefore

$$E_p[\,\partial u/\partial w\,]=\partial u/\partial w\,|_{w=w(p^*)} =1.$$

Let $r_k = p^* + \delta_k$ be the deviations in reports with respect to p^* due to risk aversion. Additionally,

$$w(k) = \alpha + \beta 2 r_k - \beta \sum_{i=1...K} (r_i)^2$$

$$w(k) = \alpha + \beta 2(p^* + \delta_k) - \beta \sum_{i \neq k} [(p^* + \delta_i)^2] - \beta\, (p^* + \delta_k)^2 \tag{13}$$

and
$$w(p^*) = \alpha + \beta 2 p^* - \beta \sum_{i=1...K} (r_i)^2$$

$$w(p^*) = \alpha + \beta 2(p^*) - \beta \sum_{i \neq k} [(p^* + \delta_i)^2] - \beta\, (p^* + \delta_k)^2. \tag{14}$$

Both (13) and (14) imply that $w(k) - w(p^*) = \beta 2 \delta_k$. Taking the derivative of the utility function with respect to w and evaluating at $w(k)$, we obtain

$$\partial u/\partial w\, |_{w=w(k)} = 1 - 2c\,[w(k) - w(p^*)] = 1 - 2c\,[\,\beta 2 \delta_k] = 1 - 4c\,\beta \delta_k. \tag{15}$$

By the definition of r_k, $\partial u/\partial w\, |_{w=w(k)}$ and $E_p[\,\partial u/\partial w\,]$ we have

$$r_k = p^* + \delta_k = p_k \times \{\partial u/\partial w\, |_{w=w(k)}\}/E_p[\,\partial u/\partial w\,],$$

which implies that

$$p^* + \delta_k = p_k \times \{1 - 2c\,[\,\beta 2 \delta_k]\,\}/\{1\}.$$

Solving for δ_k we obtain $\delta_k = \{p_k - p^*\}/\{1 + 4c\,\beta\, p_k]\,\}$. If $p_k \neq 0$, then $\lim_{c \to \infty} \delta_k = 0$ and the deviations become vanishingly small for sufficiently risk-averse individuals.

Now prove that $p^* \approx 1/K$, where K is the number of states for which $p_k \neq 0$. By definition

$$\sum_{i=1...K} (r_i) = \sum_{i=1...K} (p^* + \delta_k).$$

If $p_k = 0$, then $\lim_{c \to 0} \delta_k = p_k - p^*$ and $\lim_{c \to \infty} r_k = p^* + \delta_k = p_k = 0$. If $p_k \neq 0$, then $\sum_{p^k \neq 0} (\delta_k)$ tends to zero and $\sum_{p^k \neq 0} (p^*) = 1 = K p^* = 1$, so $p^* = 1/K$ in the limit. These two facts combine to prove that if $p_k \neq 0$ then $\lim_{c \to \infty} r_k = \lim_{c \to \infty} p^* + \delta_k = 1/K$. That is, the reported probabilities approach a uniform distribution over the outcomes where the subjective probability is non-zero. ■

Proposition 6: The following relationship exists between means of the reported and actual subjective distributions: If $u(w) = w + \delta \times u^*(w)$ with δ small, then for any random variable y, $E_r[\text{y}] - E_p[\text{y}] = \delta \times Cov_p\,[\partial u/\partial w, \text{y}]$.

Proof. If a subject exhibits utility function $u(w) = w + \delta \times u^*(w)$, we know from (1′) that the following K conditions must be satisfied:

$$p_k \times [1+\delta \times \partial u^*/\partial w\,|_{w=w(k)}] - p_k \times \{1+\delta\, E_p[\,\partial u^*/\partial w\,]\} - \varepsilon_k \times \{1+\delta\, E_p[\,\partial u^*/\partial w\,]\} = 0,\ \forall\ k = 1,\ldots, K,$$

where ε_k is defined in (7) for Lemma 2. Solving for ε_k we obtain

$$\varepsilon_k = \delta\, p_k \times \{\partial u^*/\partial w\,|_{w=w(k)} - E_p[\,\partial u^*/\partial w\,]\} / \{1+\delta\, E_p[\,\partial u^*/\partial w\,]\},\ \forall\ k = 1,\ldots, K. \tag{16}$$

Assume a random variable y with K possible states of nature. Define $E_r[\,y\,] = \sum_{k=1\ldots n} r_k\, y_k$ and $E_p[\,y\,] = \sum_{k=1\ldots n} p_k\, y_k$. Then the difference of the expected value of y under measures $\{r_1, r_2, \ldots, r_k, \ldots, r_K\}$ and $\{p_1, p_2, \ldots, p_k, \ldots, p_K\}$ is equal to $E_r[\,y\,] - E_p[\,y\,] = \sum_{k=1\ldots K} \varepsilon_k\, y_k$. Substituting for ε_k using (16), it can be shown that the denominator $\{1+\delta\, E_p[\,\partial u^*/\partial w\,]\}$ drops out (take a Taylor Series expansion of the reciprocal, multiply terms with the numerator, and drop higher-order terms). Then we have

$$\begin{aligned} E_r[\,y\,] - E_p[\,y\,] &= \delta \times \textstyle\sum_{K=1\ldots K} p_k\{\partial u^*/\partial w\,|_{w=w(k)} - E_p[\,\partial u^*/\partial w\,]\}\ y_k \\ &= \delta \times \{E_p\,[\,\partial u^*/\partial w \times y\,] - E_p\,[\,\partial u^*/\partial w\,]\ E_p\,[\,y\,]\} \\ &= \delta \times Cov_p[\partial u^*/\partial w, y] = Cov_p[\partial u/\partial w, y]. \qquad \blacksquare \end{aligned}$$

3. Some Evidence

The above theoretical results are meant to help apply and interpret empirical efforts to elicit subjective belief distributions. Many of the properties of the scoring rule cannot be directly tested, given that they refer to unknown subjective beliefs: *de opinio non est disputandum*. For instance, Proposition 3 is a valuable property, but to test it we would need to know that some individual attached zero subjective weight to some specific interval of events.

However, it is possible with a controlled laboratory experiment to offer some evidence in support of the claim that, for a risk-averse individual with symmetric subjective beliefs, the mean of the reported distribution is equal to the mean of the actual subjective distribution (Proposition 1).

This result is important for tests of the Reduction of Compound Lotteries axiom of SEU theory. In the laboratory we can present a stimulus for beliefs which provides a basis for symmetric beliefs, and for which we know, by design, the true objective probability distribution implied by the stimulus. It is then a simple matter to compare that true stimulus with the average elicited belief, under the maintained assumption that there is no basis for subjective beliefs to be biased in comparison to the true stimulus. This assumption cannot be easily made outside of this laboratory environment.

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. The subjects were told that there were no other salient, rewarded choices for them to make before or after they made their choices, avoiding possible confounds by having to assume the "isolation effect" if one were making many choices.[13]

Table 1 summarizes our experimental design for each of the 8 laboratory sessions we ran, as well as the sample size of subjects in each treatment per session. A total of 245 participants were recruited from a general subject pool of undergraduates at Georgia State University.

We implement two between-subjects treatments within each session so that both groups are presented with the same randomly chosen and session-specific stimulus, thus we are able to compare treatment effects while conditioning on a specific realized stimulus. There are three treatments in total. In **treatment 10m**, which was used in all 8 sessions, we elicit subjective belief *distributions* about the true fraction of red balls in the bingo cage by using a generalized QSR with monetary outcomes.

[13] The "random lottery" payment protocol in which one asks the subject to make K>1 choices, and pick 1 of the K at random for payment at the end, requires that the Independence axiom applies. But then one cannot use those data to estimate models of decision-making behavior that assumes the invalidity of that axiom. The only reliable payment protocol in this case is to ask subjects to only make one choice, and pay them for it. See Harrison and Swarthout [2012] for discussion, including the literature evaluating the behavioral validity of the isolation effect.

In **treatment 2m**, which was used in the first 4 sessions, we elicit subjective *probabilities* that a single red ball would be drawn from bingo cage by using the QSR with monetary outcomes. This probability elicitation task is known to be one in which risk averse subjects would rationally and significantly distort their reports towards ½. By comparing elicited reports across treatments in these first 4 sessions, we can assess the practical significance of our claims about the weak effects of risk aversion on optimal reports under treatment 10m. Finally, in **treatment 10p**, which was used in sessions 5 through 8, we elicited subjective belief distributions with a generalized QSR that paid out with binary lottery procedure "points" instead of money.[14] These points subsequently determined the objective probability of winning a binary lottery.

Each session was conducted in the manner described below. 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.[15] Then a Verifier was selected at random among the subjects solely for the purpose of 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.

In the introductory instructions subjects were informed that part of the experiment was to test different computer screens and that they will be divided into two groups. Subjects were told that each of them was assigned to one of the two groups depending on whether their 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

[14] The binary lottery procedure, explained below, is due to Smith [1961], who actually introduced it for the purpose of eliciting a subjective probability.

[15] Appendix B provides complete subject instructions.

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 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.[16] 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 on the ball 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 with a black blanket and placed on a platform in the front of the room. After subjects were alerted to pay attention, Bingo Cage 2 was then uncovered for subjects to see, spun for 10 turns, and covered again. This visual display was the information that each subject received. Subjects then made their decisions based on this information about the

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

number of red and white balls in Bingo Cage 2. After decisions were made, subjects completed a non-salient demographic survey. Immediately after, earnings were determined. To resolve payments in treatment 2m, the experimenter drew a ball from Bingo Cage 2. The sealed envelope was then 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.

A computer interface was used to present to subjects the belief elicitation tasks and to record their choices, allowing them to allocate tokens to reflect their subjective beliefs. Figure 7 presents the interface used for the distribution elicitation treatment common to all 8 sessions (treatment 10m).[17] The interface implements the QSR discussed earlier, with $\alpha=\beta=25$. Subjects could move the sliders at the bottom of the screen interface to re-allocate the 100 tokens as they wished, ending up with some distribution as shown in Figure 8. The instructions explained that they could earn up to \$50, but only by allocating all 100 tokens to one interval *and* that interval containing the true percent: if the true percent was just outside the selected interval, they would in that case receive \$0.

The stimulus, the number of red balls in Bingo cage 2, was different in each session since we wanted the true number of red balls to be generated in a credible manner, to avoid subjects second-guessing the procedure. This credibility comes at the risk that the stimulus is extreme and uninformative: if there had been only 1 red ball, or 99 red balls, we would not have generated informative data. As it happens, we had a good variety of realizations over the 8 sessions.

A. Eliciting Belief Distributions

We first consider the subjective belief distributions elicited with the generalized QSR in all 8 sessions (treatment 10m). To illustrate the data, Figure 9 presents the elicited beliefs pooled over

[17] The exact interface varied by treatment, and the additional versions of the interface will be presented later in this section.

the 15 subjects in the first session, in which the true percent was 69%. Of course there is some dispersion in beliefs, since the stimulus was deliberately designed not to provide exact information (unless, by unfortunate chance, the number of red balls was extreme). As it happens, the average of this elicited distribution is 72.3%, very close to the true proportion of red balls.[18]

Figure 10 reports the results across all eight sessions. With one exception, the elicited averages closely track the true averages. Again, the maintained, joint hypothesis that allows us to view this as evidence for the truthful elicitation of subjective belief distributions is that subjects behave consistently with SEU and that their subjective belief distribution are distributed around the true population average that provides the common stimulus they all observe.

The exception is session 7, in which the true number of red balls was 11% and the elicited average was 23.0%. This disparity was due to three outliers, subjects who we believe *a priori* not to have understood the task. One subject allocated 36 tokens to the interval for 81% to 90%, and 64 tokens to the interval for 91% to 100%; it is possible this subject was confused as to whether he was betting on red or white. If this subject is removed, the average becomes 18.4%. Then there were two subjects that exhibited some degree of confusion, although less extreme than the first outlier.[19] If these are also removed, the average becomes 13.5%, close to the true number of red balls. Of course one is always wary claiming that someone is an outlier, although every behavioral economist knows that such subjects exist, and occasionally even in clusters like this.[20]

---

[18] The average is estimated using an interval regression model with no covariates. Hence the dependent variable is literally the interval selected by the subject, and the weight on that interval is the number of tokens allocated to the interval.

[19] One of these subjects allocated roughly 10 tokens to each and every interval, and the other allocated roughly 10 tokens to each interval below 50%, 28 tokens to the interval for 71% to 80%, and small numbers of tokens for other intervals greater than 50%.

[20] We can statistically test the hypothesis that the elicited averages in Figure 10 are equal to the true percent by estimating an interval regression model in which the intervals are the bin "labels" in Figures 7 and 8, and the tokens allocated to each bin are frequency weights for each subject. We also cluster the standard errors on each subject. If we estimate this model with no constant term, but a variable with the true percent,

We have independent evidence that the subjects from this population do "robustly" exhibit risk aversion over stakes comparable to those used in the present experiment: see Holt and Laury [2002][2005] and Harrison and Swarthout [2012], for instance. Thus the close correspondence with the predictions of Proposition 1 is not due to the risk neutrality of the subjects over these stakes.

B. Eliciting Distributions Versus Probabilities

We conduct a direct test of the effect of risk aversion by comparing elicited beliefs *for the same physical stimulus* but using different scoring rules. As is well known, the QSR for *binary* events will elicit biased responses if the subject is risk averse: intuitively, the subject is drawn to report 50% so as to equalize earnings under each possible outcome, providing subjective beliefs are not degenerate. In sessions 1 through 4, we compare the elicited belief distributions discussed above with elicited probabilities based on the QSR for binary events, as shown in Figure 11. In this case the binary event was a single draw from Bingo Cage 2 containing the red and white balls. Although all subjects within a given session are presented with the same physical stimulus, the two groups of subjects face different tasks: in the binary case the individual is betting over their subjective perception of an order statistic, and in the 10-event case the individual is betting over their subjective perception of a sufficient statistic of the population. Nonetheless, for the sessions in which the stimuli were not close to 50% already, sessions 1 through 3, we observe in Figure 12 a striking tendency for the reports using the binary scoring rule to be closer to 50% than to the true proportion. This is perfectly consistent with our predictions, since risk aversion has a significantly biasing effect for the

we can directly test the hypothesis that the coefficient on that variable is 1. We find a point estimate of 1.015 with a 95% between 0.966 and 1.064, and a *p*-value of 0.55 on the null hypothesis that it is equal to 1. Hence we conclude that average elicited beliefs are not statistically significantly different from the true percent.

binary scoring rule, and virtually none for the 10-event scoring rule.[21]

C. Scoring Rules and the Binary Lottery Procedure

Finally, sessions 5 through 8 allow us to conduct a different test of our predictions by comparing the distributional scoring rule defined directly over monetary payoffs with the same scoring rule defined over a binary lottery. This binary lottery rewards subjects with points, which convert into increased probability of winning a high monetary prize or a low monetary prize. Figure 13 illustrates the interface. The information is similar to Figure 8, except that earnings are in points.

The manner in which points were converted into money was explained as follows:

> 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. [...]
>
> Your earnings depend on your reported beliefs, the true answer and the outcome of a dice roll. For instance, suppose you allocated your tokens as in the figure shown above. In this case the true composition of the bingo cage might have been 25 red balls and 75 white balls. Then you would earn 39 points. Now suppose that you rolled a 35 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 39, 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.

The report distribution chosen by the subject only ever generates a probability of receiving $50 instead of receiving $0. We can normalize the utility of the subject from $50 to 1, and the utility

[21] Again, an interval regression can be used to provide statistical support for this conclusion. The choices from the binary scoring rule are specific probabilities that a red ball will be drawn, hence treated naturally as uncensored observations in which the upper interval equals the lower interval. The model assumes clustered standard errors on the individual, and no constant. In this instance we include a dummy variable taking on the value 0 for the distributional scoring rule (Figures 7 and 8), +1 for the binary scoring rule (Figure 7) when the true percent is less than ½, and -1 for the binary scoring rule when the true percent is greater than ½. This allows for the *a priori* prediction that risk aversion will lead the reports in the binary scoring rule setting to more towards ½ from below or above, depending on which side of ½ the true percent is. This dummy for the binary scoring rule has a point estimate of 14.02 percentage points, and is statistically significantly different from zero with a *p*-value of less than 0.001. Hence the binary scoring rule and the distributional scoring rule elicit different averages.

from \$0 to 0, and it is immediate that the subject has had a linear utility function induced (Smith [1961]). Given our theoretical results, we therefore predict that subjects facing this elicitation mechanism will behave identically to those facing direct monetary payoffs.

It might seem theoretically redundant to check if these variants elicit different distributions, but there is a strong behavioral reason for undertaking this check. The reason is that the binary lottery procedure is not regarded widely by experimental economists as behaviorally reliable, whatever the theoretical claims. For example, Cox and Oaxaca [1995] and Selten, Sadrieh and Abbink [1999] provide the strongest criticisms of the procedure, even arguing that it makes matters worse, rather than just being useless at inducing linear utility.[22] We are interested in evaluating if indeed the procedure merely adds an additional layer of protection from the effects of non-linear utility. If so, it might be attractive in applications for the binary lottery procedure to be added to practical implementations.

Figure 14 displays the averages of the elicited distributions, shown separately by the session in which these were elicited, since subjects within each session faced exactly the same physical stimulus. The data from Figure 10 are reproduced here, and we add the averages from the elicited distributions over points. It is apparent how close the averages are, supporting our theoretical predictions.[23]

---

[22] We see the procedure as having more behavioral validity than the received wisdom, especially in the context of indiviudal decision-making, and critically review these concerns and the literature in Harrison, Martínez-Correa and Swarthout [2012a][2012b] for objective probabilities and subjective probabilities, respectively. We conclude there that the procedures works behaviorally as theoretically advertized in the simplest settings examined here, in which subjects make individually just one choice. We also show that it works well in more complex settings, but we have not examined all settings of interest (e.g., in an N-person game, such as a first-price auction). In any event, our immediate purpose here is not to resurrect the binary lottery procedure as being behaviorally valid, if that is needed, but to provide additional evidence with respect to our theoretical predictions.

[23] Again we can use an interval regression model to evaluate this claim statistically. A binary dummy for the points procedure, no pun intended, has a point estimate of 1.97 percentage points, a 95% confidence interval between -1.5 and +5.4, and a p-value of 0.26.

4. Conclusions

These results provide strong support for the use of practical methods for eliciting subjective belief *distributions* over *continuous* events. Contrary to the case in which one elicits subjective *probabilities* over *binary* events, there is *a priori* and empirical support for not needing to adjust or de-bias the reports for continuous events on account of risk aversion.

Figure 1: Optimal Reports Assuming Risk Neutrality and Subjective Expected Utility

CRRA utility function with ρ = 0 (risk neutral)

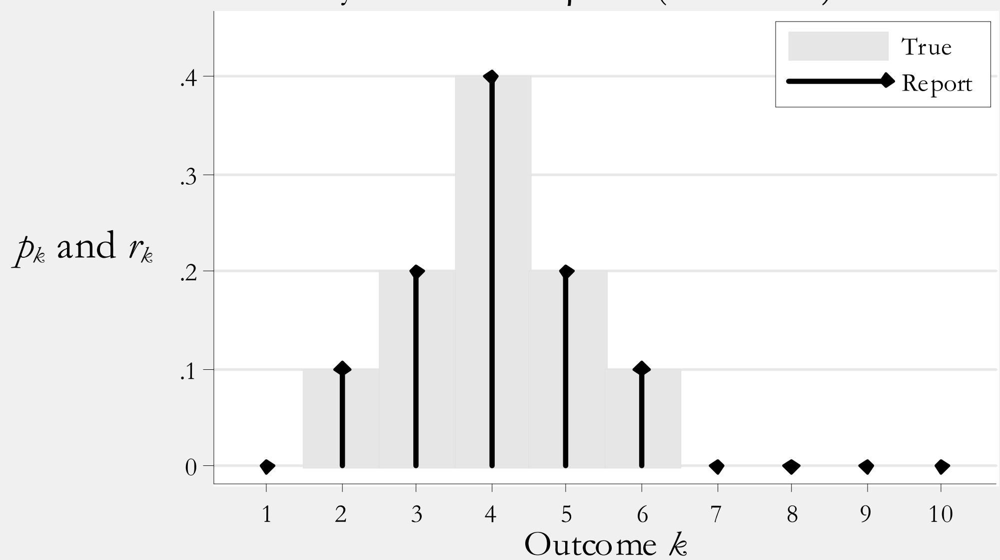


Figure 2: Optimal Reports Assuming Risk Aversion and Subjective Expected Utility

CRRA utility function with ρ = 0.65 (risk averse)

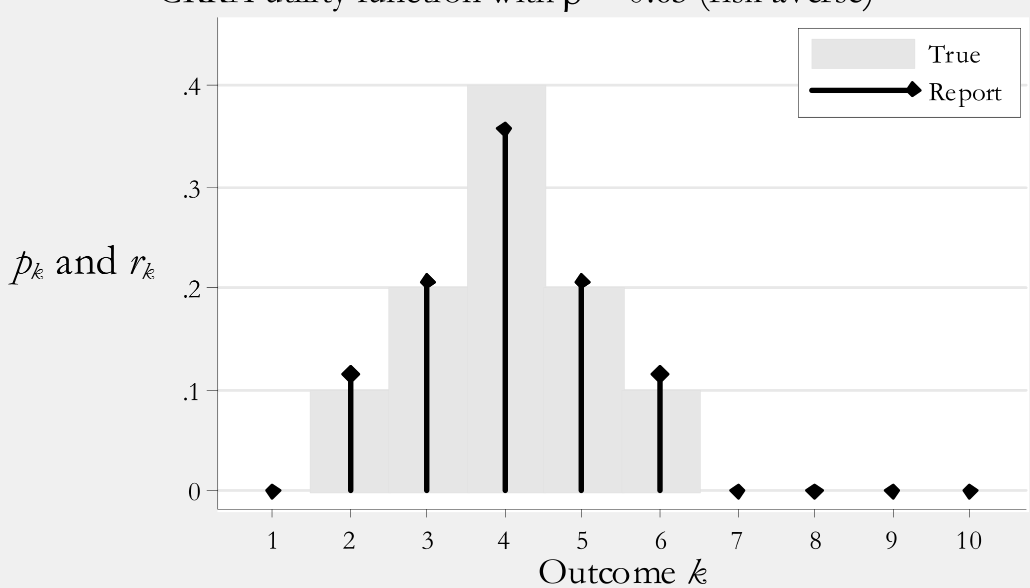


Figure 3: Optimal Reports Assuming Unimodal, Symmetric Beliefs and Subjective Expected Utility

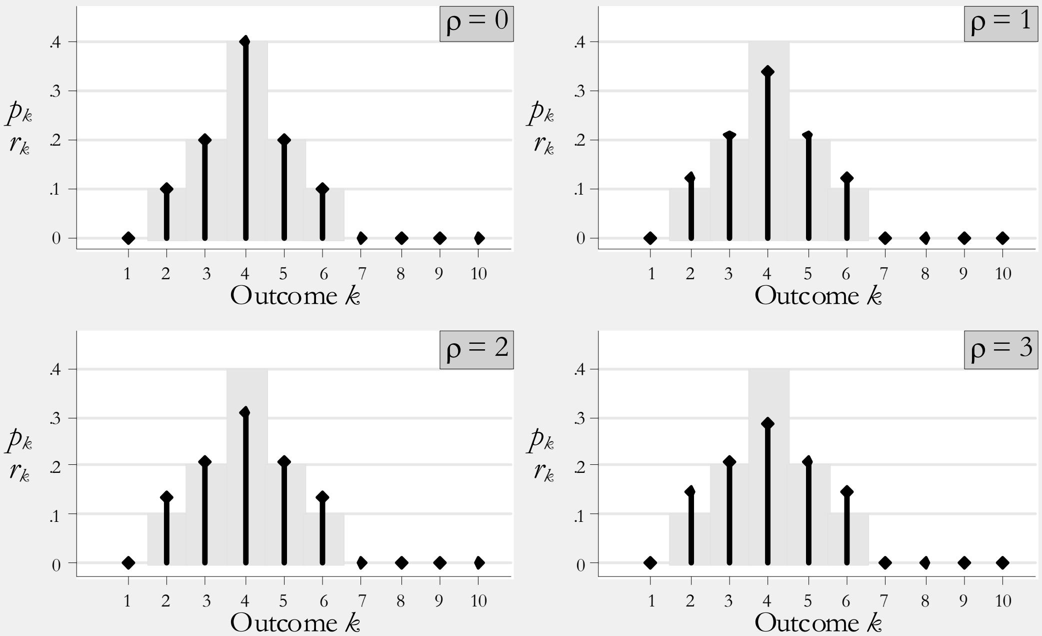


Figure 4: Optimal Reports Assuming Unimodal, Asymmetric Beliefs and Subjective Expected Utility

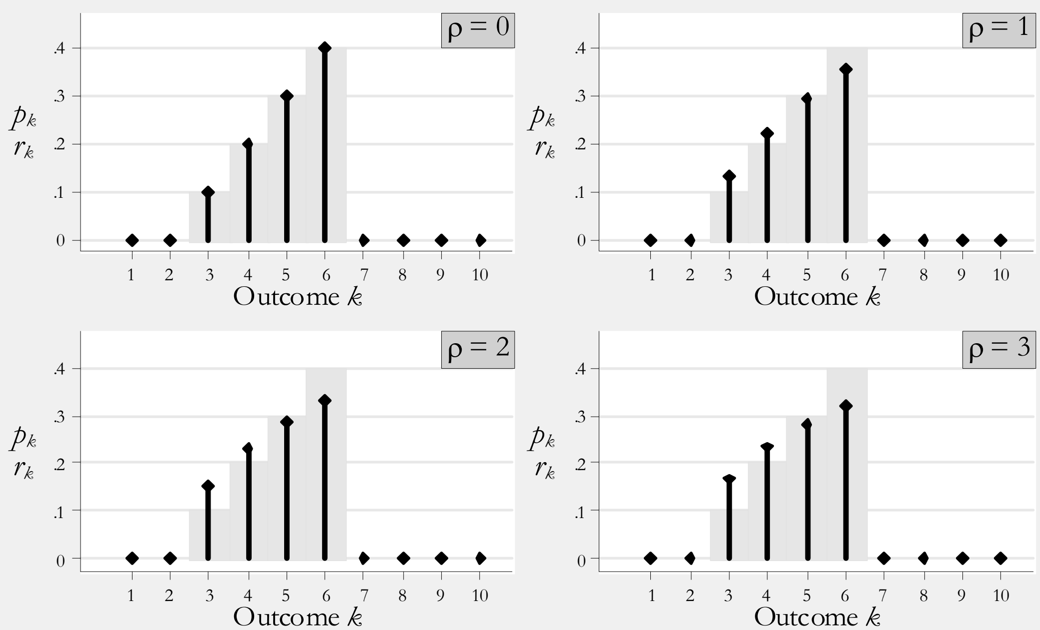


Figure 5: Optimal Reports Assuming Bimodal, Symmetric Beliefs and Subjective Expected Utility

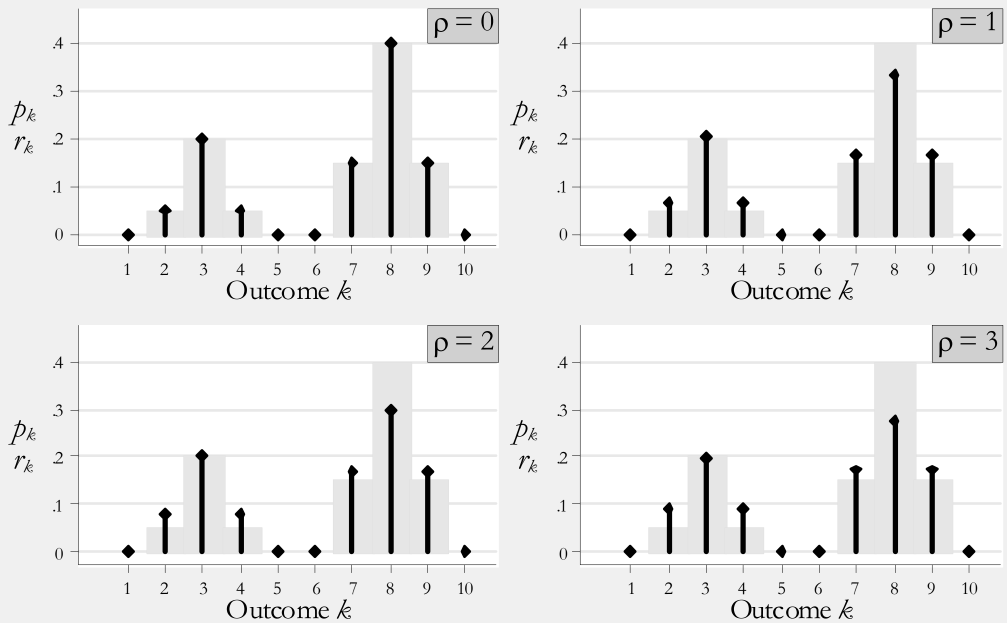


Figure 6: Difference Between True Average and Reported Average with Asymmetric Beliefs and Varying Relative Risk Aversion

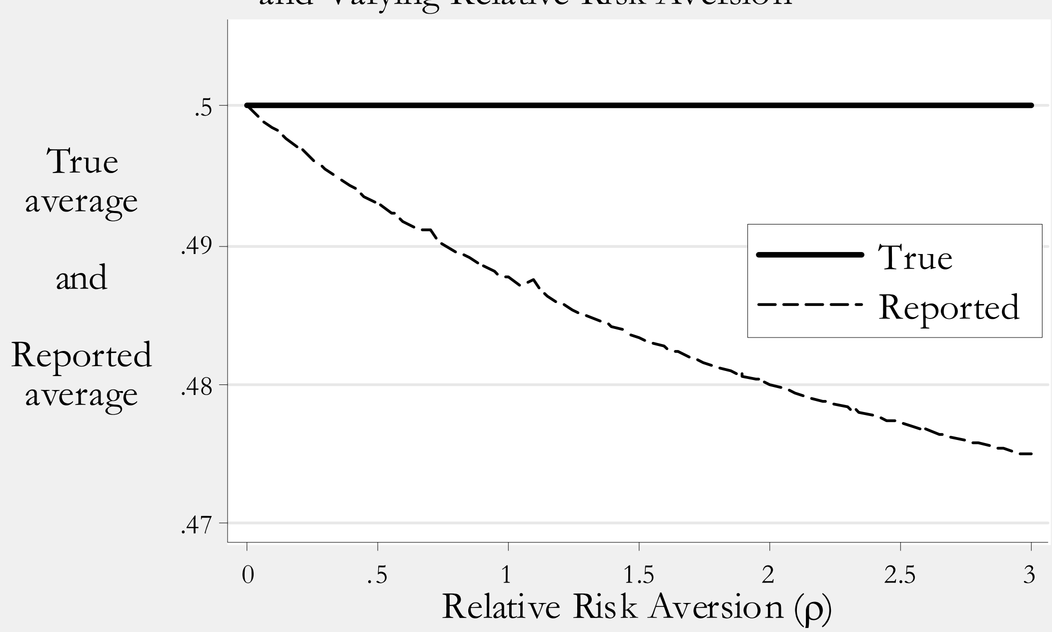


Table 1: Experiment Design and Sample Sizes

Session	Treatments			Total
	10m	2m	10p	
1	15	14	0	29
2	15	16	0	31
3	15	17	0	32
4	13	14	0	27
5	15	0	12	27
6	18	0	17	35
7	18	0	18	36
8	14	0	14	28
Total	123	61	61	245

Notes: treatment 10m is elicitation of a distribution with the QSR defined directly over money, treatment 2m is elicitation of a probability with the QSR over money, and treatment 10p is elicitation of a distribution with the QSR over binary lottery procedure "points" which convert into the probability of winning a high monetary prize.

Figure 7: Initial Belief Elicitation Interface

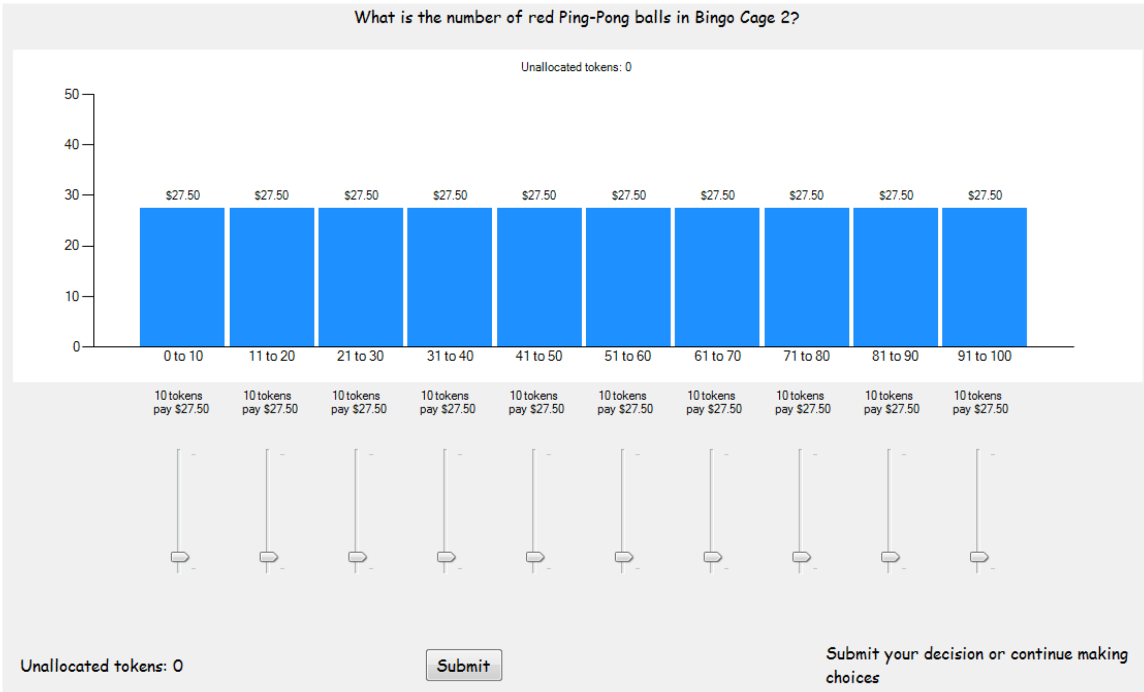


Figure 8: Typical Belief Elicitation Response

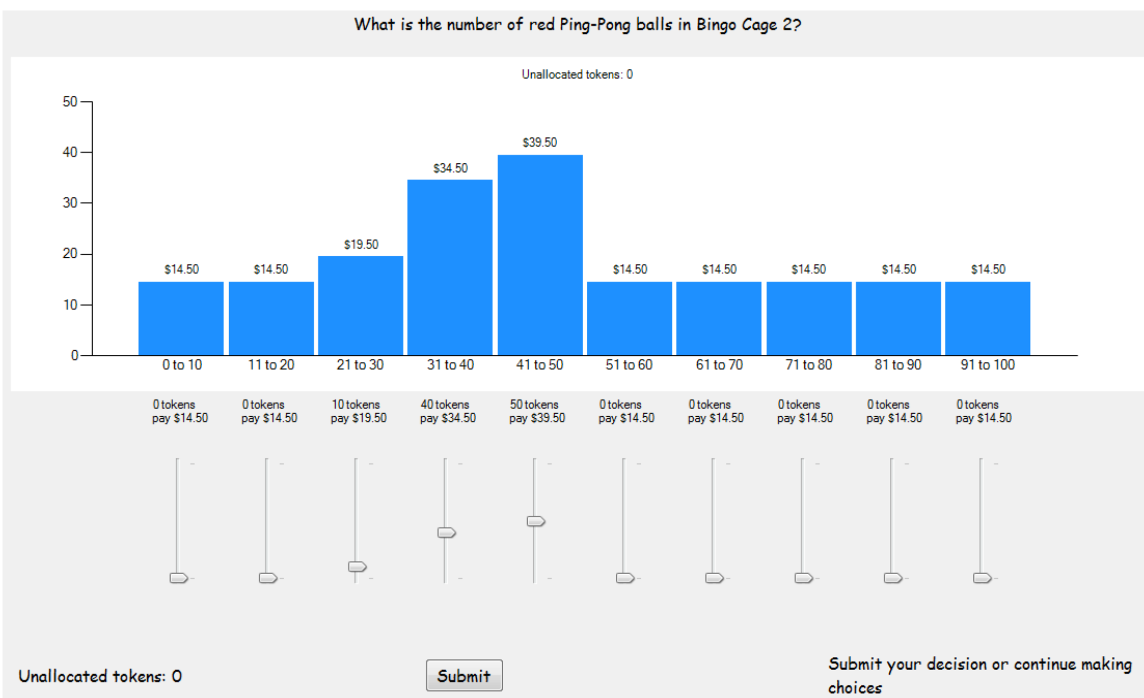


Figure 9: Elicited Subjective Distribution Pooled Over 15 Subjects and With True Percent of 69%

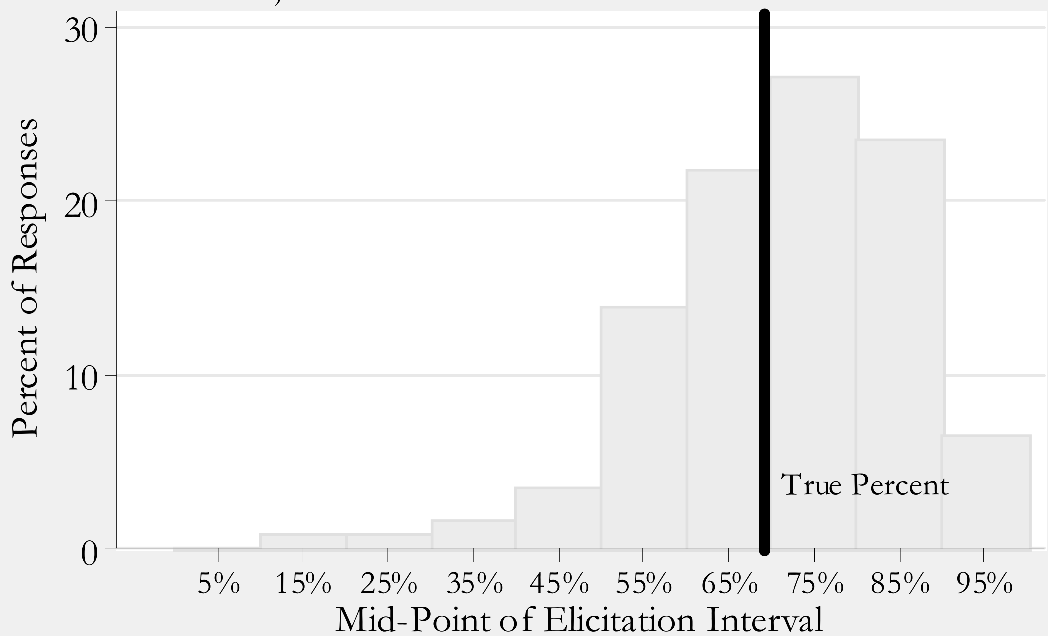


Figure 10: Average Elicited Subjective Belief Distribution

Pooled average for each of 8 sessions
Sample sizes: 15, 15, 15, 13, 15, 18, 18 and 14

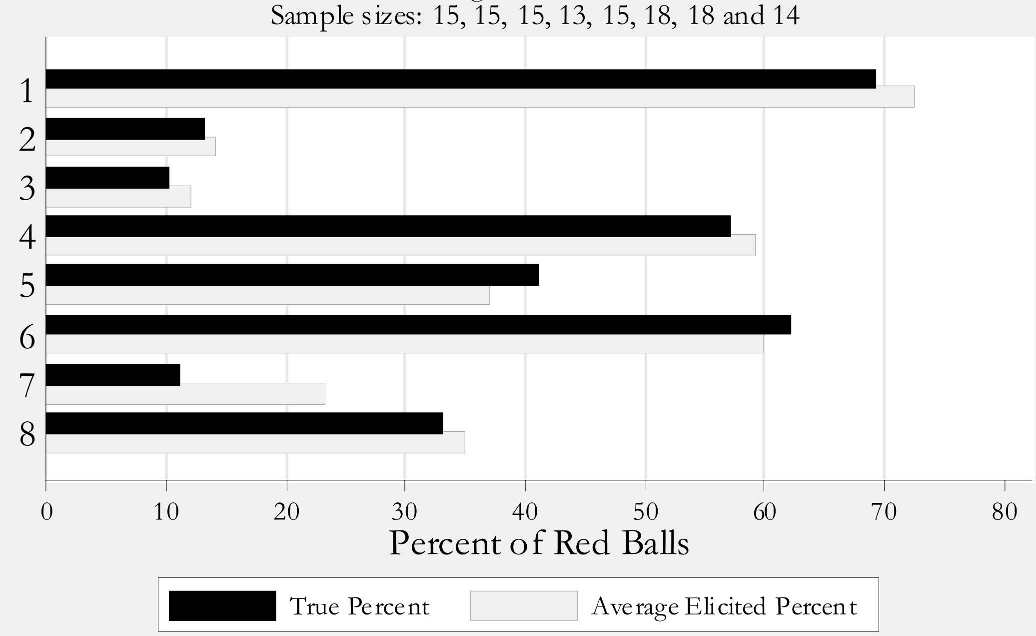


Figure 11: Subjective Probability Elicitation Interface

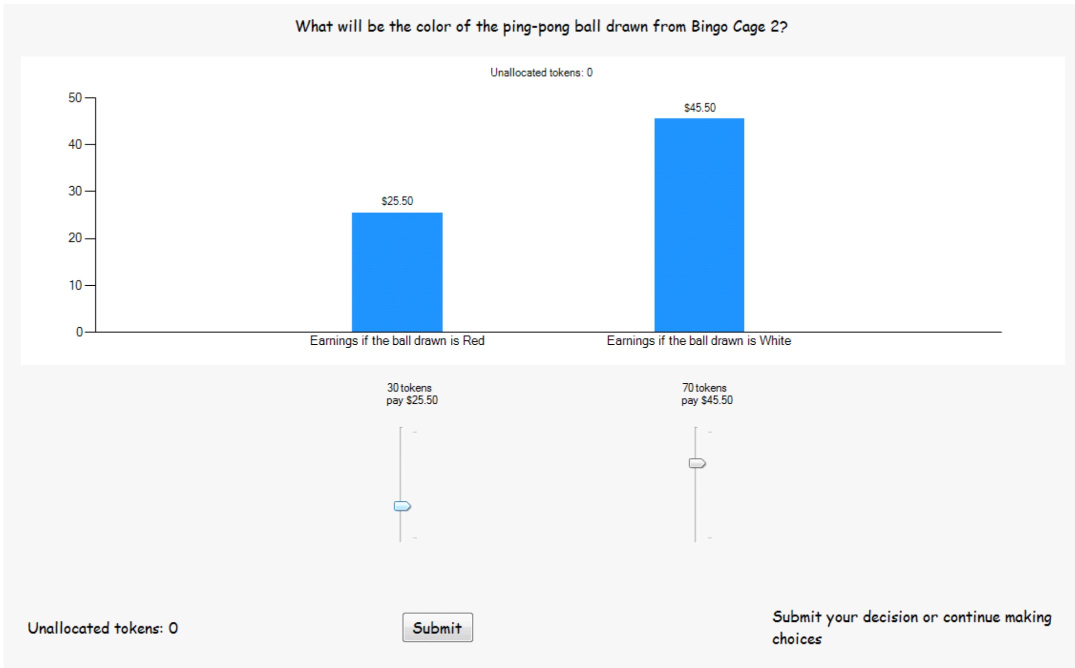


Figure 12: Average of Elicited Subjective Belief Distributions and Average Elicited Subjective Probabilities

Pooled averages for each of 4 sessions, with treatments *within* each session.
Each session used the same random stimulus.
One treatment elicited beliefs about the true distribution of red balls,
and another treatment elicited the probability of a red ball being drawn.
Sample sizes for distribution (probability): 15(14), 15(16), 15(17) and 13(14).

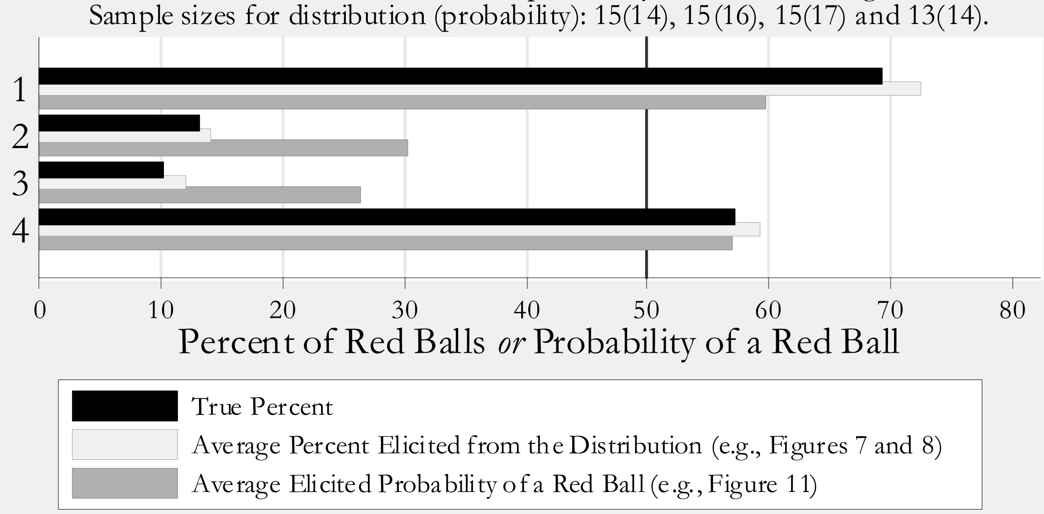


Figure 13: Belief Elicitation Interface Over Points

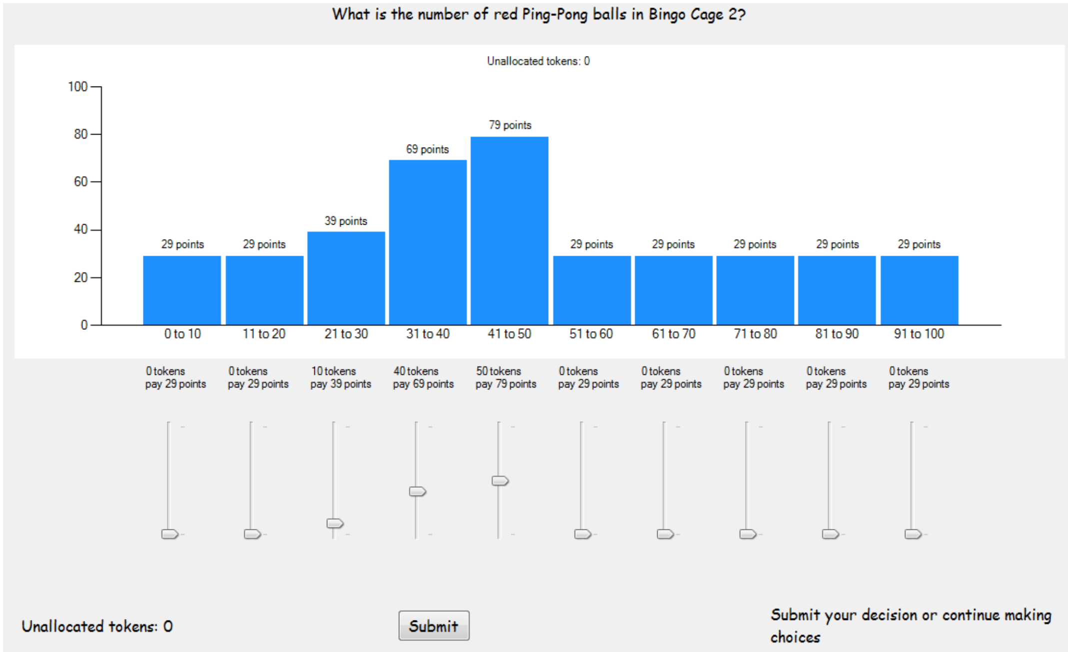
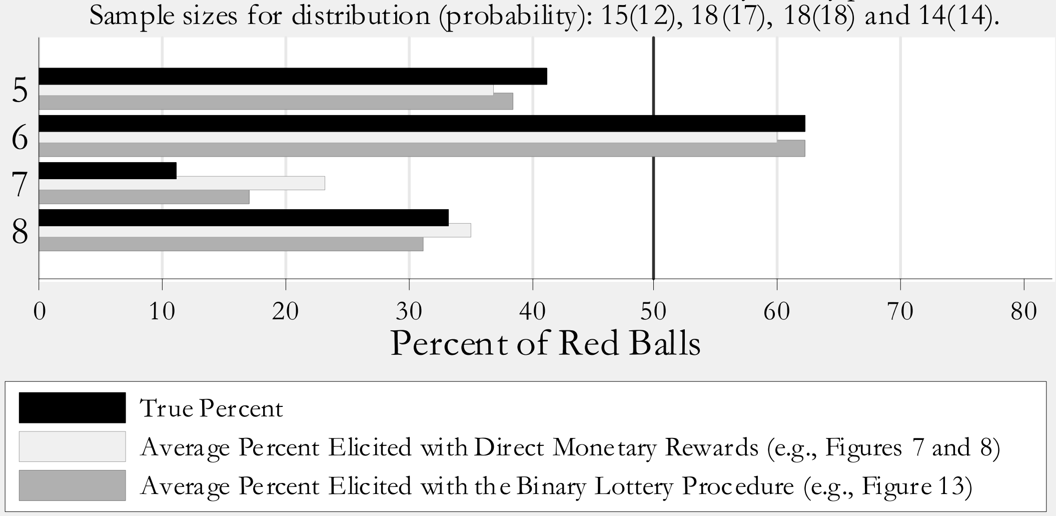


Figure 14: Average of Elicited Subjective Belief Distributions using Money and the Binary Lottery Procedure

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



References

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

Cox, James C., and Oaxaca, Ronald L., "Inducing Risk-Neutral Preferences: Further Analysis of the Data," *Journal of Risk and Uncertainty*, 11, 1995, 65-79.

Cox, James C.; Roberson, Bruce; and Smith, Vernon L., "Theory and Behavior of Single Object Auctions," in V.L. Smith (ed.), *Research in Experimental Economics* (Greenwich: JAI Press, volume 2, 1982).

Delavande, Adeline; Giné, Xavier, and McKenzie, David, "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence," *Journal of Development Economics*, 94, 2011, 151-163.

Dominitz, Jeff, and Manski, Charles F., "Measuring and Interpreting Expectations of Equity Returns," *Journal of Applied Econometrics*, 26, 2011, 352-370.

Ergin, Haluk, and Gul, Faruk, "A Theory of Subjective Compound Lotteries," *Journal of Economic Theory*, 144, 2009, 899-929.

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, Martínez-Correa, Jimmy and Swarthout, J. Todd, "Inducing Risk Neutral Preferences with Binary Lotteries: A Reconsideration," *Working Paper 2012-02*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012a; *Journal of Economic Behavior & Organization*, forthcoming.

Harrison, Glenn W, Martínez-Correa, Jimmy and Swarthout, J. Todd, "Eliciting Subjective Probabilities With Binary Lotteries," *Working Paper 2012-09*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012b.

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

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

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

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, 902-912.

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

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.

Karni, Edi, and Safra, Zvi, "The Impossibility of Experimental Elicitation of Subjective Probabilities," *Theory and Decision*, 38, 1995, 313-320.

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).

Lichtendahl, Jr., Kenneth C., and Winkler, Robert L., "Probability Elicitation, Scoring Rules, and Competition Among Forecasters," *Management Science*, 53(11), November 2007, 1745-1755.

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

Nau, Robert F., "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, 52, 2006, 136-156.

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.

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

Selten, Reinhard; Sadrieh, Abdolkarim, and Abbink, Klaus, "Money Does Not Induce Risk Neutral Behavior, but Binary Lotteries Do even Worse," *Theory and Decision*, 46(3), June 1999, 211-249.

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

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

Appendix A: Proofs for the Continuous Case

Lemma 1: Let $p(s)$ represent the underlying subjective probability density function of an individual and let $r(s)$ represent the reported probability density function in a given scoring rule. Let $w(s) = \alpha + \beta 2\, r(s) - \beta \int r^2(s)ds$ be the scoring rule that determines the wealth if state k occurs. If the individual has a utility function $u(w)$ that is continuous and differentiable and maximizes expected utility over actual subjective probabilities, the actual and reported probability must obey the following equation:

$$p(s) \times \partial u/\partial w\,|_{r(s)} - r(s) \times E_p[\partial u/\partial w] = 0 \tag{A.1}$$

Proof. Suppose an expected utility maximizer has a continuous subjective probability density function $p(s)$ over the possible states of nature and utility function $u(w)$ over random wealth. If the subject is given a scoring rule determined by $w(s) = \alpha + \beta 2\, r(s) - \beta \int r^2(s)ds$, then the optimal reported probability density function $r(s)$ solves the following problem:

$$\mathbf{Max}_{\{r\}}\ E_p[\, u(w)\,] \text{ subject to } \int r^2(s)ds\ = 1 \tag{A.2}$$

where $E_p[\, u(w)\,] = \int u[\, \alpha + \beta 2 r(s) - \beta \int r^2(s)ds\,]\, p(s)\, ds$. This problem can be solved by maximizing the Lagrangian

$$\mathcal{L} = \int u[\, \alpha + \beta 2 r(s) - \beta \int r^2(s)ds\,]\, p(s)\, ds - \lambda \int r(s)ds \tag{A.3}$$

Again assume that $r(s)$ is the solution and $\delta m(s)$ is a small deviation from the solution

$$\mathcal{L} = \int u[\, \alpha + \beta 2 r(s) - \beta \int r^2(t)dt + \beta 2\, \delta m(s) - \beta 2 \delta \int r(t)m(t)dt$$

$$- \beta\, \delta^2 \int m^2(t)dt\,]\, p(s)\, ds - \lambda \int r(s)ds - \lambda\, \delta \int m(s)ds \tag{A.4}$$

The first order condition with respect to report δ is

$$\partial\mathcal{L}/\partial\delta = \int p(s)\, \partial u/\partial w|_{r(s)}\, (\beta 2 m(s) - \beta 2 \int r(t)m(t)dt - \beta 2 \delta^2 \int m^2(t)dt\,)ds - \lambda \int m(s)ds\ = 0, \tag{A.5}$$

$$\beta 2 \int p(s)\, \partial u/\partial w|_{r(s)}\, [m(s) - \int r(t)m(t)dt]ds\ = 0, \forall\ m(s)$$

$$\int p(s)\, \partial u/\partial w|_{r(s)} m(s) - \int r(t)m(t)dt\, E_p[\partial u/\partial w] = 0, \forall\ m(s)$$

$$\int [p(s)\, \partial u/\partial w|_{r(s)} - r(s) \mathrm{E_p}[\partial u/\partial w]] m(s)ds = 0, \forall\ m(s)$$

Therefore

$$p(s)\, \partial u/\partial w|_{r(s)} - r(s) \mathrm{E}_p[\partial u/\partial w] = 0$$

■

Lemma 2: Under the condition in Lemma 1, let $\boldsymbol{\varepsilon}(s) = r(s) - p(s)$ be the deviation between the reported and actual subjective probability for state of nature s. Then

$$\boldsymbol{\varepsilon}(s)\ =\ p(s) \times \{\partial u/\partial w|_{r(s)} - E_p[\, \partial u/\partial w\,]\}/E_p[\, \partial u/\partial w\,] \tag{A.6}$$

Proof. Assume that the condition of Lemma 1 in (A.1) is satisfied and the distortions between the actual and reported probabilities are given by $r(s) = p(s) + \varepsilon(s)$, with $\int \varepsilon(s)ds = 0$. Then

$$p(s)[\,\partial u/\partial w\,|_{r(s)} + E_p[\partial u/\partial w]] = \varepsilon(s)\, E_p[\partial u/\partial w]$$

Solving for $\varepsilon(s)$ we obtain

$$\varepsilon(s) \;=\; p(s) \times \{\partial u/\partial w\,|_{r(s)} - E_p[\,\partial u/\partial w\,]\}/E_p[\,\partial u/\partial w\,] \qquad \blacksquare$$

Lemma 3: Assume an individual that has a continuous, differentiable utility function $u(w)$ over random wealth and who is also risk averse (i.e., $\partial^2 u/\partial^2 w < 0$, $\forall\, w$). If $p(s) = p(t)$ for some values s and t, then $r(s) = r(t)$.

Proof. By contradiction. Assume that $p(s) = p(t)$ for some values s and t. Now suppose without loss of generality that $r(s) > r(t)$. By definition of the deviation of subjective and reported probabilities the latter implies that $\varepsilon(s) > \varepsilon(t)$ because

$$r(s) \;=\; p(s) + \varepsilon(s) \;>\; p(t) + \varepsilon(t) = r(t) \tag{A.7}$$

Since $r(s) > r(t)$, we also know that $w(\,s\,) > w(\,t\,)$, and by the concavity of $u(.)$ the latter implies that $\partial u/\partial w\,|_{w(s)} < \partial u/\partial w\,|_{w(t)}$. Therefore

$$\{\,\partial u/\partial w\,|_{w(s)} - E_p[\partial u/\partial w\,]\}/E_p[\,\partial u/\partial w] < \{\,\partial u/\partial w\,|_{w(t)} - E_p[\partial u/\partial w\,]\}/E_p[\,\partial u/\partial w] \tag{A.8}$$

But by Lemma 2, (A.8) implies that $r(s) < r(t)$, which is a contradiction. $\blacksquare$

Proposition 1: For the risk-averse individual in Lemma 3, if the subjective distribution is symmetric then the mean of the reported distribution is equal to the mean of the actual subjective distribution.

Proof. A symmetric and continuous subjective distribution p for random variable y has mean $E_p[\mathrm{y}]$. Let $p(\mathrm{y_m}-\Delta\mathrm{y}) = p(\mathrm{y_m}+\Delta\mathrm{y})$. Define the mean of the distribution as follows

$$E_p[\mathrm{y}] = \int_{-\infty}^{\mathrm{y_m}} \mathrm{y}p(\mathrm{y})d\mathrm{y} + \int_{\mathrm{y_m}}^{\infty} \mathrm{y}p(\mathrm{y})d\mathrm{y} \tag{A.9}$$

In the first term in (A.9), let $\mathrm{y} = \mathrm{y_m}-\Delta\mathrm{y}$ and in the second term let $\mathrm{y} = \mathrm{y_m}+\Delta\mathrm{y}$. We obtain

$$E_p[\mathrm{y}] = \int_{-\infty}^{0} (\mathrm{y_m}-\Delta\mathrm{y}\,)p(\mathrm{y_m}-\Delta\mathrm{y}\,)[d\Delta\mathrm{y}] + \int_{0}^{\infty} (\mathrm{y_m}+\Delta\mathrm{y})p(\mathrm{y_m}+\Delta\mathrm{y})[d\Delta\mathrm{y}]$$

and since $p(\mathrm{y_m}-\Delta\mathrm{y}) = p(\mathrm{y_m}+\Delta\mathrm{y})$, the mean of the distribution is equal to

$$E_p[\mathrm{y}] = 2\int_{0}^{\infty} \mathrm{y_m}\, p(\mathrm{y_m}+\Delta\mathrm{y})[d\Delta\mathrm{y}]$$

$$+ \int_{0}^{\infty} \Delta\mathrm{y}\; p(\mathrm{y_m}+\Delta\mathrm{y})[d\Delta\mathrm{y}] - \int_{0}^{\infty} \Delta\mathrm{y}\; p(\mathrm{y_m}+\Delta\mathrm{y})[d\Delta\mathrm{y}]$$

$$= \; 2\int_{0}^{\infty} \mathrm{y_m}\, p(\mathrm{y_m}+\Delta\mathrm{y})[d\Delta\mathrm{y}] = \mathrm{y_m} = E_p[\mathrm{y}]$$

E_r[y] follows similarly since $r(y_m - \Delta y) = r(y_m + \Delta y)$ by Lemma 3. ■

Lemma 4: The converse of Lemma 3. Assume an individual with a continuous, differentiable utility function $u(w)$, where risk aversion is not necessary in this case. If $r(s) = r(t)$ for this individual, then $p(s) = p(t)$.

Proof. It follows from Lemma. If $r(s) = r(t)$ then

$$p(s)\ \partial u/\partial w|_{r(s)} - r(s)\mathrm{E}_p[\partial u/\partial w]=0 \text{ and } p(t)\ \partial u/\partial w|_{r(t)} - r(t)\mathrm{E}_p[\partial u/\partial w]=0 \quad \text{(A.10)}$$

Thus,

$$p(s) = r(s)\mathrm{E}_p[\partial u/\partial w]/\ \partial u/\partial w|_{r(s)} = r(t)\mathrm{E}_p[\partial u/\partial w]/\ \partial u/\partial w|_{r(t)} = p(t)$$ ■

Proposition 2: For the individual in Lemma 4, if the reported distribution is symmetric then the mean of the reported distribution is equal to the mean of the actual subjective distribution.

Proof. The proof is identical to Proposition 1 with $p(.)$ and $r(.)$ reversed. ■

Proposition 3: Assume an individual with a continuous, differentiable utility function $u(w)$. If the actual subjective probability of the individual for a value s is $p(s) = 0$, then the reported probability is $r(s) = 0$. That is, the individual does not report weight where none is believed to exist. Conversely, if $r_k = 0$ then $p_k = 0$.

Proof. Follows directly from Lemmas 1 and 2 in the same fashion as the discrete case. ■

Proposition 4: A risk-averse individual has a reported probability distribution that approaches a uniform distribution over those states where $p(s) > 0$ in the following sense: There exists a constant value p^* for this individual such that if $p(s) > p^*$ then $p(s) > r(s) > p^*$ and if $p(s) < p^*$ then $p(s) < r(s) < p^*$. A risk-loving agent reverses all the conditions.

Proof. Follows directly from Lemma 2 in the same fashion as the discrete case. ■

Proposition 5: An individual with sufficiently high risk aversion will have a reported probability arbitrarily close to p^*.

Proof. By Lemma 2

$$r(s) = p(s)\ \partial u/\partial w|_{r(s)}/\mathrm{E}_p[\partial u/\partial w] \quad \text{(A.11)}$$

Also define $u(w) = w - c\,[w - w(p^*)]^2$ without loss of generality. Therefore

$$E_p[\ \partial u/\partial w\ ]=1.$$

Let $r(s) = p^* + \delta_m(s)$ be the deviations in reports with respect to p^* due to risk aversion. Additionally,

$$w(s) = \alpha + \beta 2\, r(s) - \beta \int_{-\infty}^{\infty} r(t)^2 dt$$
$$= \alpha + \beta 2(p^* + \delta_m(s)) - \beta \int_{-\infty}^{\infty} (p^* + \delta_m(s))^2 ds \qquad \text{(A.12)}$$

and
$$w(p^*) = \alpha + \beta 2 p^* - \beta \int_{-\infty}^{\infty} (p^* + \delta_m(s))^2 ds \qquad \text{(A.13)}$$

Both (A.12) and (A.13) imply that $w(s) - w(p^*) = \beta 2\delta_m(s)$. Taking the derivative of the utility function with respect to w and evaluating at $w(k)$, we obtain

$$\partial u/\partial w\,|_{w=w(s)} = 1 - 2c\,[w(s) - w(p^*)] = 1 - 2c\,[\beta 2\, \delta_m(s)] = 1 - 4c\beta\delta_m(s) \qquad \text{(A.14)}$$

By the definition of $r(s)$, $\partial u/\partial w\,|_{w=w(s)}$ and $E_p[\,\partial u/\partial w\,]$ we have

$$r(s) = p^* + \delta_m(s) = p(s)\times \{\partial u/\partial w\,|_{w=w(s)}\}/E_p[\,\partial u/\partial w\,],$$

which implies that

$$p^* + \delta_m(s) = p(s) \times \{1 - 2c\,[\,\beta 2\, \delta_m(s)]\}/\{1\}$$

Solving for $\delta_m(s)$ we obtain $\delta_m(s) = \{p(s) - p^*\}/\{1 + 4c\,\beta\, p(s)]\,\}$. If $p(s) \neq 0$, then $\lim_{c\to\infty} \delta_m(s) = 0$ and the deviations become vanishingly small for sufficiently risk-averse individuals.

Now prove that $p^* = 1/K$, where K is the length of the support of the distribution with $p(s) \neq 0$. By definition

$$\int_{-\infty}^{\infty} r(s)ds = \int_{-\infty}^{\infty} (p^* + \delta_m(s))ds.$$

If $p(s) = 0$, then $\lim_{c\to 0} \delta_m(s) = p(s) - p^*$ and $\lim_{c\to\infty} r(s) = p^* + \delta_m(s) = p(s) = 0$. If $p(s) \neq 0$, then

$$\int_{p(s)\neq 0} \delta_m(s) \text{ tends to zero and } \int_{p(s)\neq 0} p^* ds = \int_{-\infty}^{\infty} r(s)ds = 1 \text{ and } p^* = 1/\int_{p(s)\neq 0} ds.$$

These two facts combine to prove that if $p(s) \neq 0$ then

$$\lim_{c\to\infty} r(s) = \lim_{c\to\infty} p^* + \delta_m(s) = 1/\ 1/\int_{p(s)\neq 0} ds.$$

That is, the reported probabilities approach a uniform distribution over the outcomes where the subjective probability is non-zero. ■

Proposition 6: The following relationship exists between means of the reported and actual subjective distributions: If $u(w) = w + \delta \times u^*(w)$ with δ small, then for any random variable y, $E_r[y] - E_p[y] = \delta \times Cov_p\,[\partial u/\partial w, y]$.

Proof. The following conditions are satisfied by Lemma 1

$$p(s) \times [1+\delta \times \partial u^*/\partial w\,|_{w=w(s)}] - p(s)\times \{1+\delta\, E_p[\,\partial u^*/\partial w\,]\} - \varepsilon(s) \times \{1+\delta\, E_p[\,\partial u^*/\partial w\,]\} = 0 \qquad \text{(A.15)}$$

Solving for $\varepsilon(s)$ we obtain

$$\varepsilon(s) = \delta\, p(s) \times \{\partial u^*/\partial w\,|_{w=w(s)} - E_p[\,\partial u^*/\partial w\,]\}/\{1+\delta\, E_p[\,\partial u^*/\partial w\,] \qquad \text{(A.16)}$$

Assume a continuous random variable y. Define $E_r[\, y \,] = \int y \; r(s) \; ds$ and $E_p[\, y \,] = \int y \; p(s) \; ds$. Then the difference of the expected value of y under measures $r(s)$ and $p(s)$ is equal to

$$E_r[\, y \,] - E_p[\, y \,] = \int y \; \boldsymbol{\varepsilon}(y) \; dy \tag{A.17}$$

Substituting $\boldsymbol{\varepsilon}(y)$ defined by (A.16) into (A.17), it can be shown that the denominator

$$\{1 + \delta \, E_p[\, \partial u^* / \partial w]\}$$

drops out (take a Taylor Series expansion of the reciprocal, multiply terms with the numerator, and drop higher-order terms). Then we have

$$E_r[\, y \,] - E_p[\, y \,] = \delta \int_{-\infty}^{\infty} \{\partial u^* / \partial w \,|_{w=w(s)} - E_p[\, \partial u^* / \partial w \,]\} y \; p(s) dy$$

$$= \delta \times \{E_P [\, \partial u^* / \partial w \times y \,] - E_P [\, \partial u^* / \partial w \,] \; E_P [\, y \,]\}$$

$$= \delta \times Cov_p[\partial u^* / \partial w, y] = Cov_p[\partial u / \partial w, y]. \qquad \blacksquare$$

Appendix B: Experimental Instructions (NOT FOR PUBLICATION)

We provide the following information about the experiments:

- the cover sheet that all subjects saw, explaining the setup for the session;
- the specific instructions for the 10-bin scoring rule with payoffs directly over money (Figures 7 and 8);
- the specific instructions for the 2-bin binary scoring rule with payoffs directly over money (Figure 11);
- the specific instructions for the 10-bin scoring rule with payoffs defined over points (Figure 13);
- the protocol for constructing the bingo cages in the experiments in which we elicited the true percent as well as the outcome of a draw from the second cage (i.e., when we had some subjects using the interface of Figures 7 and 8, and some subjects using the interface of Figure 11);
- the protocol for constructing the bingo cages in the experiments in which we only elicited the true percent (i.e., when we had some subjects using the interface of Figures 7 and 8, and some subjects using the interface of Figure 13); and
- the "verification sheet" projected on the public screens so that everyone could see what the Verifier was in fact verifying as the second cage was constructed.

Although some of these steps are mildly tedious, we believe them necessary to ensure that all subjects know that the process generating the second cage was as we had explained.

A. Cover Sheet

Introductory Instructions

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

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

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

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

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

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

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

B. Instructions for the 10-Bin Scoring Rule Defined Over Direct Monetary Payoffs

10m

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 composition of 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.

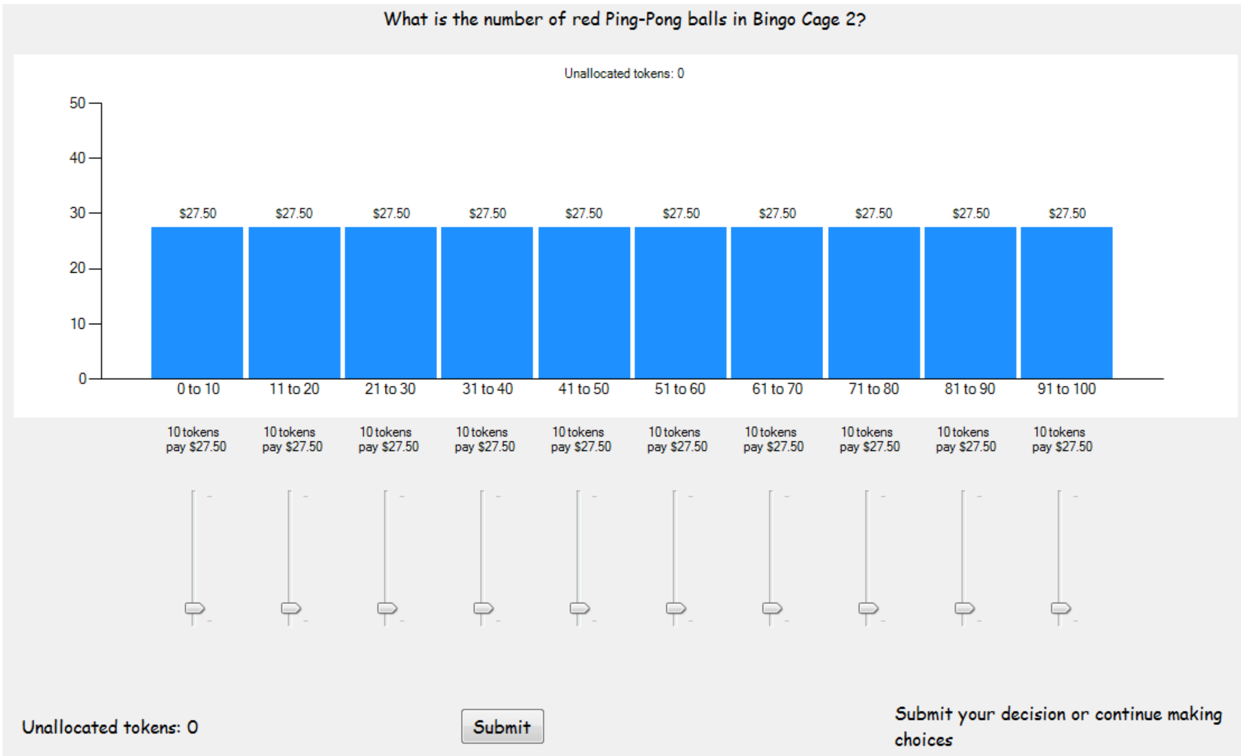
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab 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, 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. Your winnings will depend on your choices and the actual number of red balls in Bingo Cage 2.

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.

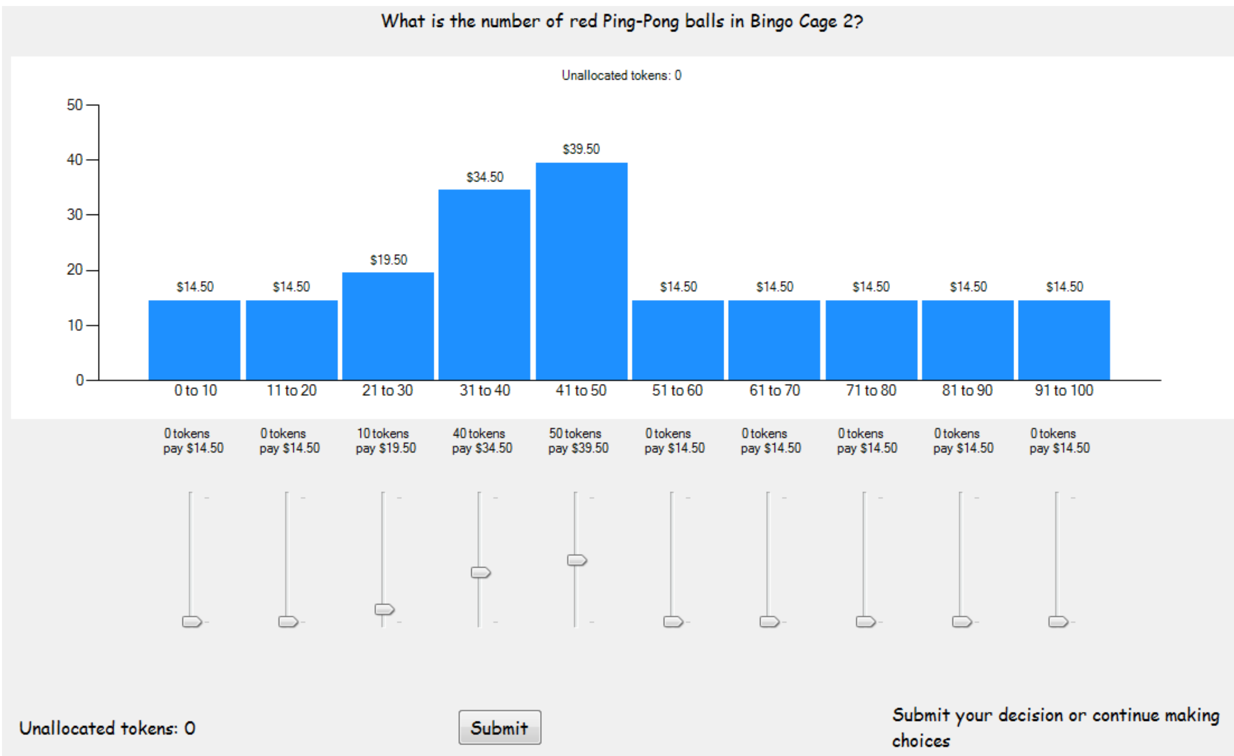


The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we always start with 10 tokens being allocated to each slider. The payoffs shown on the screen only apply when you allocate all 100 tokens. As you allocate tokens, by adjusting sliders, the 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 number of red Ping-Pong balls in Bingo Cage 2. 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 first bar here corresponds to your belief that the number of red balls in the bingo cage is between 0 and 10, the second bar corresponds to your belief that the number of red balls is between 11 and 20, and so on. Each bar here shows the amount of money you earn if the actual number of red balls in the bingo cage is in the interval shown under the bar.

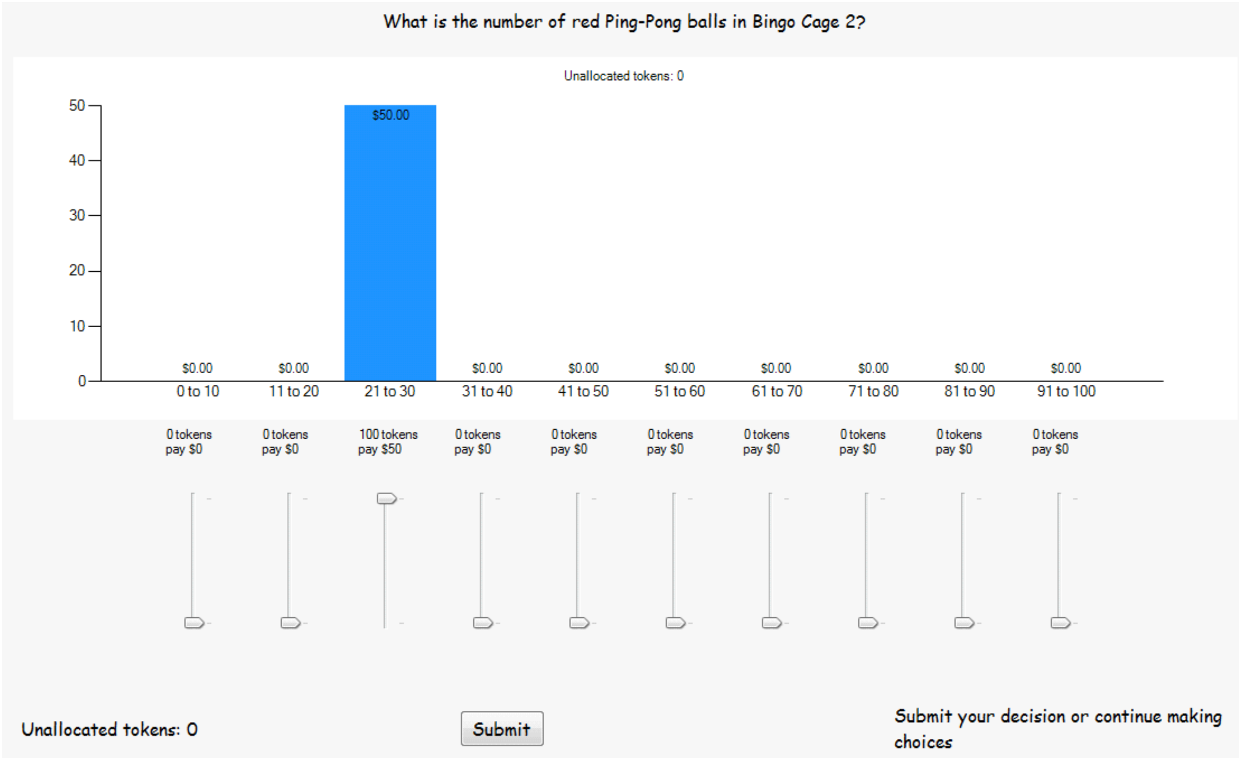
To illustrate how you use these sliders, suppose you think there is a fair chance the number of red balls is just under 50. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 41 to 50, 40 tokens to the interval 31 to 40, and 10 tokens to the interval 21 to 30. So you can see in the picture below that if indeed the number of red balls in the bingo cage is between 41 and 50 you would earn \$39.50. You would then earn less than \$39.50 for any other outcome. You would earn \$34.50 if the number of red balls is between 31 and 40, \$19.50 if it is between 21 and 30, and for any other number of red balls you would earn \$14.50.



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 true answer. For instance, suppose you allocated your tokens as in the figure shown above. In this case the true composition of the bingo cage might have been 25 red balls and 75 white balls. So if you had reported the beliefs shown above, you would have earned \$19.50.

What if you had put all of your eggs in the true basket, and allocated 100 tokens to the interval corresponding to 25 red balls? Then you would have faced the earnings outcomes shown below.



Note the "good news" and "bad news" here. If the number of red balls in Bingo Cage 2 is indeed between 21 and 30, you earn the maximum payoff, shown here as $50. But if the number of red balls had been 35 instead of 25, 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 number of red balls in 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 reveal the number of red balls in Bingo Cage 2 by showing the numbered ball in the sealed envelope and counting the red and white balls in Bingo Cage 2. Then an experimenter will come to you and record your earnings according to the correct number of red balls in Bingo Cage 2 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 the 2-Bin Binary Scoring Rule Defined Over Direct Monetary Payoffs

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 <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.

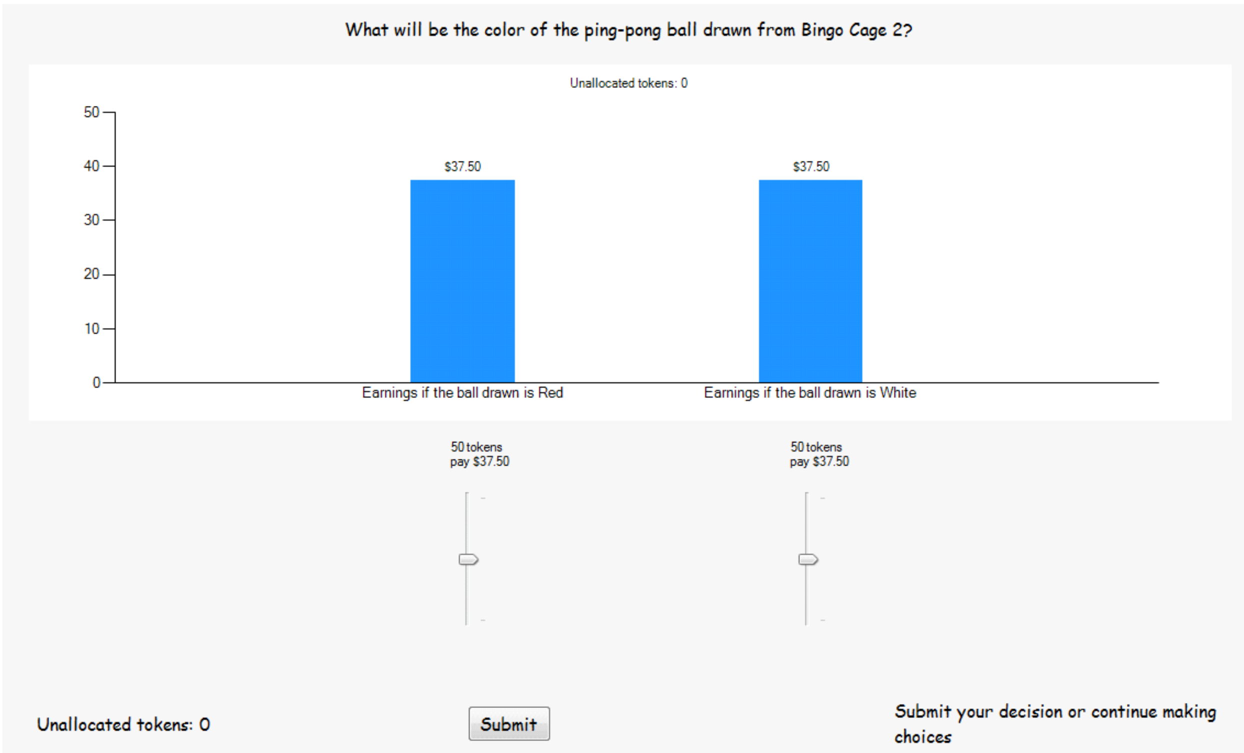
We have selected a Verifier at random solely for the purpose of verifying that we follow the process described in the instructions. When the time comes we will display a summary of the steps the Verifier will have to verify. We remind you that the Verifier must not communicate with anyone in the lab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

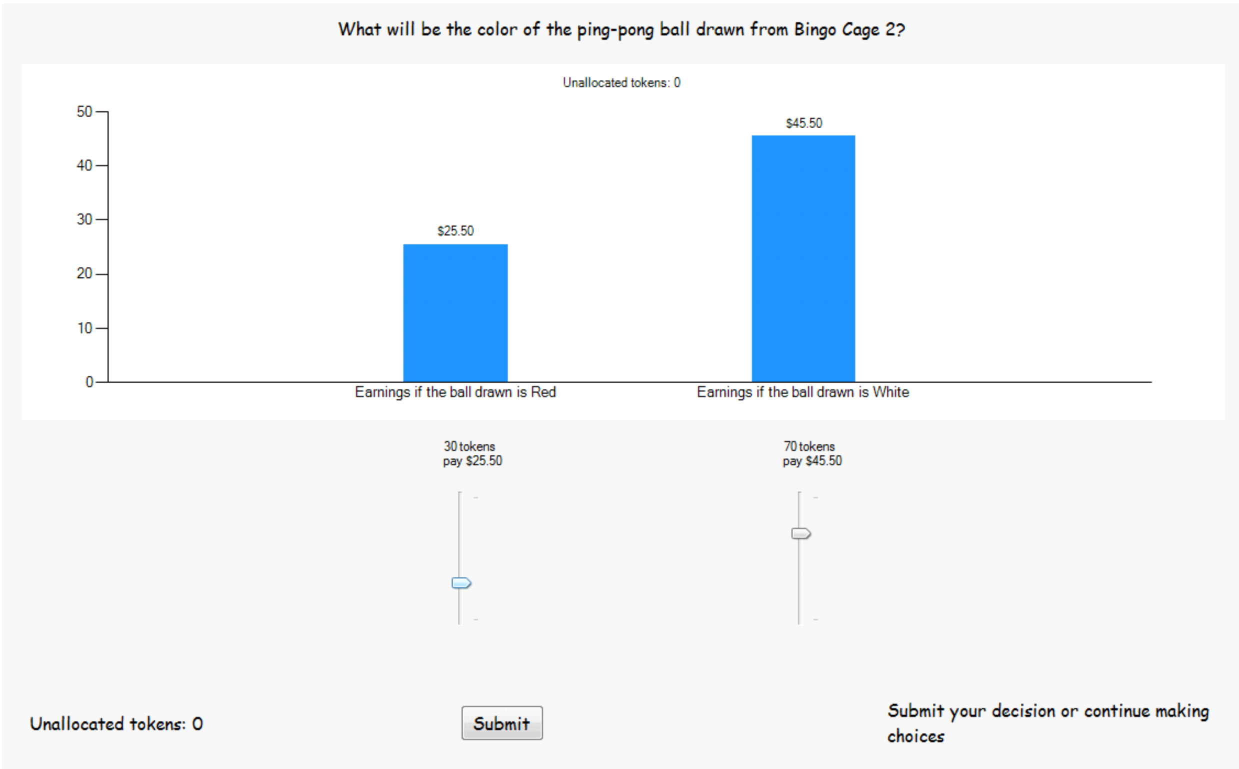


The display on your computer will be larger and easier to read. You have 2 sliders to adjust, shown at the bottom of the screen. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens in order to submit your decision, and we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

You can earn up to $50 in this task.

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

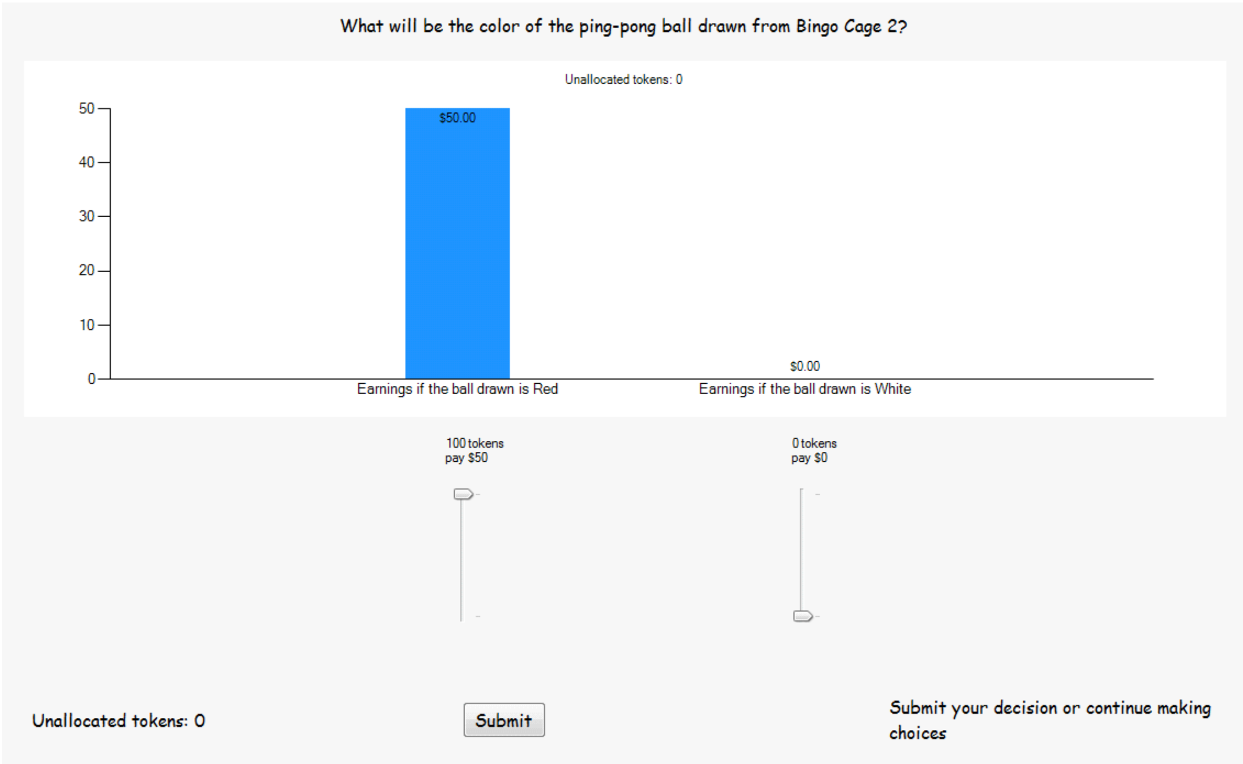
To illustrate how you use these sliders, suppose you think there is a fair chance that there are less red balls than white balls in Bingo Cage 2. Then you might allocate 30 tokens to the first bar, as shown below. Notice that the second bar will be automatically adjusted depending on the number of tokens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



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

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

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



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

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

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

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

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

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

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

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

Are there any questions?

D. Instructions for the 10-Bin Scoring Rule Defined Over Points

10p

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 composition of 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, 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. Your winnings will depend on your choices and the actual number of red balls in Bingo Cage 2.

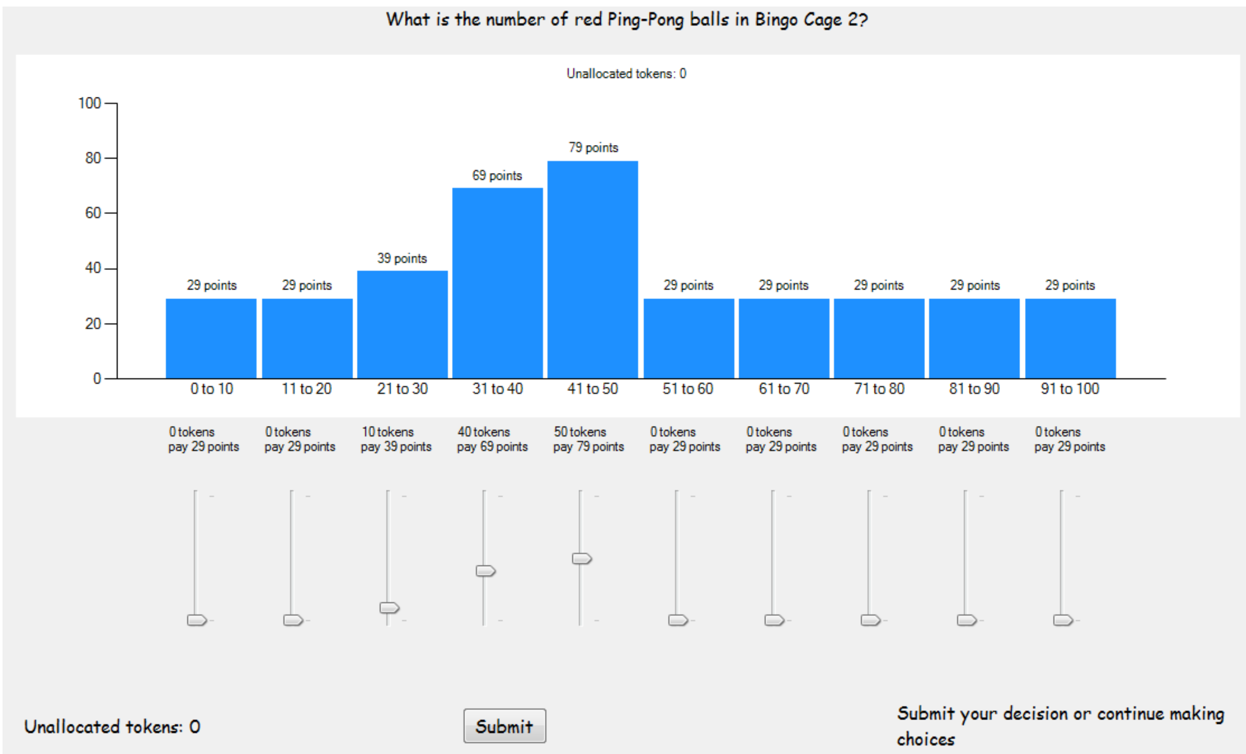
Now we will explain how you will actually make your choices. To make your choices, you will use a computer screen like the one shown below.

The display on your computer will be larger and easier to read. You have 10 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 10 tokens being allocated to each slider. The payoffs shown on the screen only apply when you allocate all 100 tokens. As you allocate tokens, by adjusting sliders, the 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 number of red Ping-Pong balls in Bingo Cage 2. 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 first bar here corresponds to your belief that the number of red balls in the bingo cage is between 0 and 10, the second bar corresponds to your belief that the number of red balls is between 11 and 20, and so on. Each bar here shows the amount of points you earn if the actual number of red balls in the bingo cage is in the interval shown under the bar.

To illustrate how you use these sliders, suppose you think there is a fair chance the number of red balls is just under 50. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 41 to 50, 40 tokens to the interval 31 to 40, and 10 tokens to the interval 21 to 30. So you can see in the picture below that if indeed the number of red balls in the bingo cage is between 41 and 50 you would earn 79 points. You would then earn less than 79 points for any other outcome. You would earn 69 points if the number of red balls is between 31 and 40, 39 points if it is between 21 and 30, and for any other number of red balls you would earn 29 points.

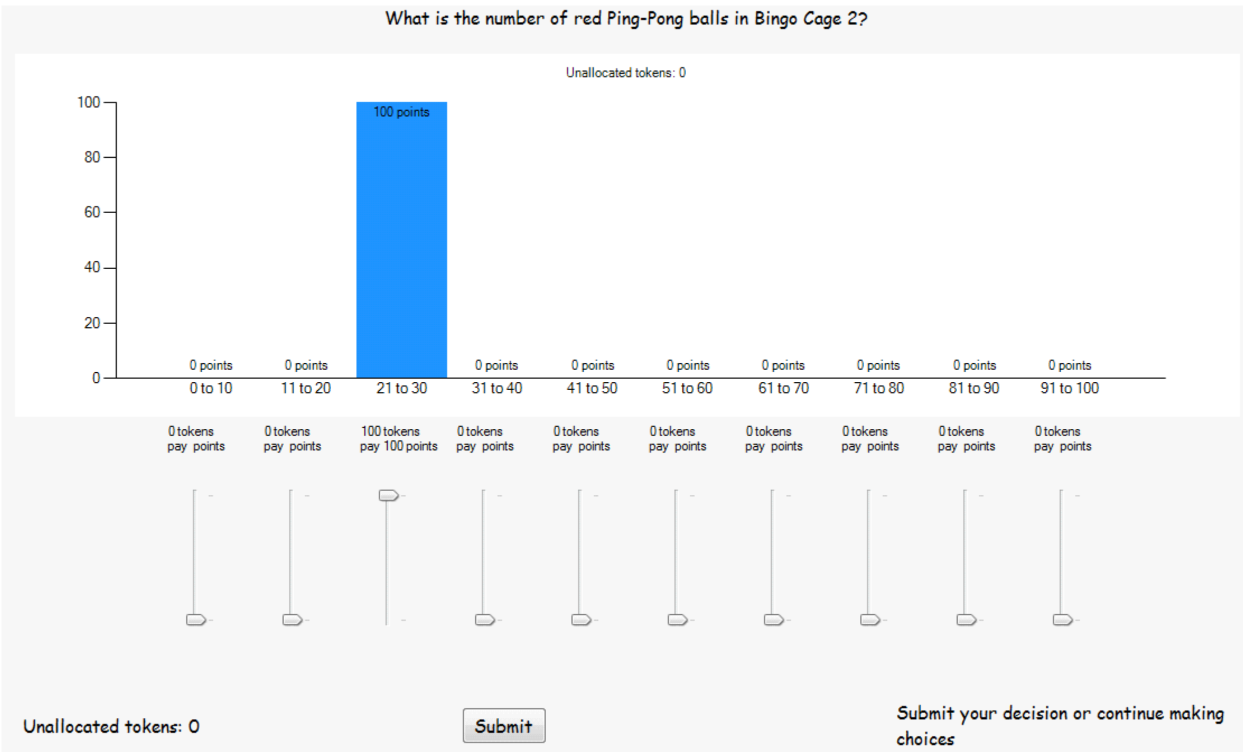


You can adjust this as much as you want to best reflect your personal beliefs about the composition of the bingo cage.

Your earnings depend on your reported beliefs, the true answer and the outcome of a dice roll. For instance, suppose you allocated your tokens as in the figure shown above. In this case the true composition of the bingo cage might have been 25 red balls and 75 white balls. Then you would earn 39 points. Now suppose that you rolled a 35 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 39, 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 you had put all of your eggs in the true basket, and allocated 100 tokens to the interval corresponding to 25 red balls? Then you would have faced the earnings outcomes shown below.



Note the "good news" and "bad news" here. If the number of red balls in Bingo Cage 2 is indeed between 21 and 30, you earn the maximum payoff, shown here as 100 points. But if the number of red balls had been 35 instead of 25, 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 number of red balls in 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 reveal the number of red balls in Bingo Cage 2 by showing the numbered ball in the sealed envelope and counting the red and white balls in Bingo Cage 2. Then an experimenter will come to you and record your earnings according to the correct number of red balls in Bingo Cage 2 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?

E. Protocol for Constructing the Bingo Cage for the 10-Bin and 2-Bin Scoring Rules

Protocol to select Bingo Cage 2.

1. Review the following steps to be verified by the Verifier

   a. Bingo Cage 1 has 99 balls numbered 1 through 99.
   b. The experimenter draws one and only one numbered ball from Bingo Cage 1.
   c. The experimenter puts in Bingo Cage 2 the number of red balls indicated on the numbered ball drawn from Bingo Cage 1.
   d. The experimenter puts the correct number of white balls into Bingo Cage 2, so that the number of white plus red balls equals 100.
   e. The experimenter puts the chosen numbered ball from Bingo Cage 1 into the envelope.

2. Cage 1:

   a. Turn on camera
   b. Load numbered balls on camera
   c. Blank projection
   d. Spin the cage three times before putting it behind the screen
   e. Move cage behind screen
   f. Select a numbered ball from cage
   g. Write number on post-it in front of red balls
   h. Write 100 – number on post-it in front of white balls
   i. Place selected numbered ball in envelope
   j. Tack it to the top of the white board

3. Cage 2:

   a. Load cage with red and white balls as indicated on post-it notes -- making sure to record on post-it notes the number of balls added each time
   b. Cover cage with blanket or box
   c. Carry it to platform with handle facing the whiteboard
   d. Uncover it and fairly rapidly rotate cage COUNTER-CLOCKWISE for 10 secs/turns
   e. Cover the cage and leave it on platform

4. Launch the software and have subjects make decisions.

5. Once all decisions have been made:

   a. Uncover cage 2 and draw one ball from cage 2 -- making sure everyone sees
   b. Unblank projection
   c. Remove ball from envelope and project on camera
   d. Uncover cage 2 and remove all balls, verifying the count of red and white balls
   e. Go to each subject with record sheet and determine earnings

F. Protocol for Constructing the Bingo Cage for the 10-Bin Scoring Rules

Protocol to select Bingo Cage 2.

1. Review the following steps to be verified by the Verifier
   a. Bingo Cage 1 has 99 balls numbered 1 through 99.
   b. The experimenter draws one and only one numbered ball from Bingo Cage 1.
   c. The experimenter puts in Bingo Cage 2 the number of red balls indicated on the numbered ball drawn from Bingo Cage 1.
   d. The experimenter puts the correct number of white balls into Bingo Cage 2, so that the number of white plus red balls equals 100.
   e. The experimenter puts the chosen numbered ball from Bingo Cage 1 into the envelope.

2. Cage 1:

   a. Turn on camera
   b. Load numbered balls on camera
   c. Blank projection
   d. Spin the cage three times before putting it behind the screen
   e. Move cage behind screen
   f. Select a numbered ball from cage
   g. Write number on post-it in front of red balls
   h. Write 100 – number on post-it in front of white balls
   i. Place selected numbered ball in envelope
   j. Tack it to the top of the white board

3. Cage 2:

   a. Load cage with red and white balls as indicated on post-it notes -- making sure to record on post-it notes the number of balls added each time
   b. Cover cage with blanket or box
   c. Carry it to platform with handle facing the whiteboard
   d. Uncover it and fairly rapidly rotate cage COUNTER-CLOCKWISE for 10 secs/turns
   e. Cover the cage and leave it on platform

4. Launch the software and have subjects make decisions.

5. Once all decisions have been made:

   a. Unblank projection
   b. Remove ball from envelope and project on camera
   c. Uncover cage 2 and remove all balls, verifying the count of red and white balls
   d. Go to each subject with record sheet and determine earnings (simultaneous with d)

G. Verification Sheet

Verification Sheet – Steps to Verify

1. Bingo Cage 1 has 99 balls numbered 1 through 99.
2. The experimenter draws one and only one numbered ball from Bingo Cage 1.
3. The experimenter puts the chosen numbered ball from Bingo Cage 1 into the envelope.
4. The experimenter puts in Bingo Cage 2 the number of red balls indicated on the numbered ball drawn from Bingo Cage 1.
5. The experimenter puts the correct number of white balls into Bingo Cage 2, so that the number of white plus red balls equals 100.