

Scoring Rules for Subjective Probability Distributions

by

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

November 2012

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. Alternatively, one must carefully 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. 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.

[†] 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, 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 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 risk aversion can dramatically affect the incentives to correctly report the true subjective *probability* of a *binary* event.¹ Alternatively, one must carefully 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 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.

Specifically, we can draw the following conclusions:

1. The individual never *reports* having a positive probability for an event that does not have positive subjective probability. So if the individual believes that inflation will never fall below

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 [1960], 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.

- 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 probability 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 Reduction of Compound Lotteries axiom over subjective belief distributions.⁵ 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

⁵ 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.

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.

1. Examples

Let the decision maker report his subjective beliefs in a discrete version of a quadratic scoring rule for continuous distributions (Mathieson and Winkler [1976]). Partition the domain into K intervals, and denote as r_k the report of the density in interval $k = 1, \dots, K$. Assume for the moment that the decision maker is risk neutral, and that the full report consists of a series of reports for each interval, $\{ r_1, r_2, \dots, r_k, \dots, r_K \}$ such that $r_k \geq 0 \forall k$ and $\sum_{i=1 \dots K} (r_i) = 1$.

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

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

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.⁶ 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:

$$S = (2 \times 1/4) - (1/4^2 + 1/4^2 + 1/4^2 + 1/4^2) = 1/2 - 1/4 = 1/4 < 1.$$

So the tradeoff from the last case is that one can always ensure a score of 1/4, 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

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

where we initially assumed $\alpha=0$ and $\beta=1$. We can assume $\alpha>0$ and $\beta>0$ 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 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 CRRA utility function $u(w) = w^{1-\rho}/(1-\rho)$ such that $\rho = 0$ denotes risk neutrality and $\rho > 0$ risk aversion. 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

⁶ This is a point of practical behavioral significance, but is not important for the immediate theoretical point.

[2008] for a review). Here we observe that relative risk aversion of $\rho = 0.65$ causes the individual to under-report the true 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 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 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 an appendix (available on request).

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 and differentiable and maximizes expected utility over actual subjective probabilities, the actual and reported probability must obey the following system of equations:

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

Proof. Suppose an expected utility maximizer has a subjective discrete probability distribution $\{p_1, p_2, \dots, p_k, \dots, p_K\}$ 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 $r = \{r_1, r_2, \dots, r_k, \dots, r_K\}$ solves the following problem:

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

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

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

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

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

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

$$\sum_{i=1 \dots K} (r_i) - 1 = 0. \quad (5)$$

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

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

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

Summing over the K first-order conditions we get

$$E_p [\partial u / \partial w |_{w=w^{(k)}}] - \sum_{j=1 \dots K} r_k E_p [\partial u / \partial w |_{w=w^{(k)}}] = K \lambda / 2\beta \quad (6)$$

Notice that $\sum_{j=1 \dots K} r_k E_p [\partial u / \partial w |_{w=w^{(k)}}] = E_p [\partial u / \partial w |_{w=w^{(k)}}]$ 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:

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

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

$$\epsilon_k = p_k \times \{ \partial u / \partial w |_{w=w^{(k)}} - E_p [\partial u / \partial w] \} / E_p [\partial u / \partial w], \forall k=1, \dots, K. \quad (7)$$

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

$$p_k \times f_k - p_k \times E_p [f] - \epsilon_k \times E_p [f] = 0, \forall k = 1, \dots, K \quad (1')$$

Solving for ϵ_k we get the K conditions stated in Lemma 2:

$$\varepsilon_k = p_k \times \{f_k - E_p[f]\} / E_p[f], \forall k = 1, \dots, K. \quad \blacksquare$$

Theorem 1: 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 $p_i = p_j$ for some i and j , then $r_i = r_j$.

The intuition of this theorem is the following. Suppose that $\{r_1, r_2, \dots, r_k, \dots, r_K\}^*$ 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 of the theorem is 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 $\varepsilon_i > \varepsilon_j$ because

$$r_i = p_i + \varepsilon_i > r_j = p_j + \varepsilon_j \quad (8)$$

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

Therefore

$$\{f_i - E_p[f]\} / E_p[f] < \{f_j - E_p[f]\} / E_p[f] \quad (9)$$

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

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

$$\begin{aligned} E_p[u(w(r_1))] &= 0.5 (2 r_1 - r_1^2 - (1-r_1)^2)^2 + 0.5 (2 (1-r_1) - r_1^2 - (1-r_1)^2)^2 \\ &= 4 r_1^4 - 8 r_1^3 + 8 r_1^2 - 4 r_1 + 1 \\ \partial E_p[u] / \partial r_1 &= 16 r_1^3 - 24 r_1^2 + 16 r_1 - 4 \end{aligned}$$

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

$$r_1^3 - (1/2) r_1^2 + r_1 - 1/4 = 0$$

$$(r_1 - 1/2)(r_1^2 - r_1 + 1/2) = 0$$

Solving for the real root we get $r_1 = 1/2$. By reporting $r_1 = 1/2$, the subjective EU is equal to $1/4$, 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 $r_1 \neq r_2$.

Corollary 1.1: For the risk-averse individual in Theorem 1, 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 $k = 1, \dots, n$, 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 $i = 1 \dots m-1$ and $p_{m-i} = p_{m+i}$.

By Theorem 1, $r_{m-i} = r_{m+i}$

$$\begin{aligned} E_p[y] &= \sum_{i=1 \dots m-1} p_{m-i} (\mu - \eta_i) + \sum_{i=1 \dots m-1} p_{m+i} (\mu + \eta_i) + p_m \mu \\ &= \sum_{j=1 \dots n} p_j \mu + \sum_{i=1 \dots m-1} [p_{m-i} - p_{m+i}] \eta_i = \mu + 0 = \mu \end{aligned} \quad (10)$$

and

$$\begin{aligned} E_r[y] &= \sum_{i=1 \dots m-1} r_{m-i} (\mu - \eta_i) + \sum_{i=1 \dots m-1} r_{m+i} (\mu + \eta_i) + r_m \mu \\ &= \sum_{j=1 \dots n} r_j \mu + \sum_{i=1 \dots m-1} [r_{m-i} - r_{m+i}] \eta_i = \mu + 0 = \mu \end{aligned} \quad (11)$$

By (10) and (11) we have that $E_p[y] - E_r[y] = 0$. The even case is similar except that $m = n/2$ and the possible outcome of the random variable μ has no weight. ■

Theorem 2: The converse of Theorem 1. Assume an individual with a continuous, differentiable utility function $u(n)$, where risk aversion is not necessary in this case. If $r_i = r_j$ for this individual, then $p_i = p_j$.

Proof. Assume without loss of generality that $r_i = r_j$ and $p_i < p_j$.

$$\begin{aligned}
\partial E_p[u] / \partial r_i &= \sum_{k \neq i, j} p_k \times \partial u / \partial w \Big|_{w=w(k)} \times [-2\beta r_i - 2\beta (1 - \sum_{\ell \neq i, j} r_\ell - r_i)] \\
&\quad + p_i \times \partial u / \partial w \Big|_{w=w(i)} \times [2\beta - 2\beta r_i - 2\beta (1 - \sum_{\ell \neq i, j} r_\ell - r_i)] \\
&\quad + p_j \times \partial u / \partial w \Big|_{w=w(j)} \times [-2\beta - 2\beta r_i - 2\beta (1 - \sum_{\ell \neq i, j} r_\ell - r_i)] \\
&= \sum_{k \neq i, j} p_k \times \partial u / \partial w \Big|_{w=w(k)} \times [-2\beta (1 - \sum_{\ell \neq i, j} r_\ell)] \\
&\quad + p_i \times \partial u / \partial w \Big|_{w=w(i)} \times [2\beta - 2\beta (1 - \sum_{\ell \neq i, j} r_\ell)] \\
&\quad + p_j \times \partial u / \partial w \Big|_{w=w(j)} \times [-2\beta - 2\beta (1 - \sum_{\ell \neq i, j} r_\ell)]
\end{aligned}$$

Since $r_i = r_j$ and $\partial u / \partial w \Big|_{w=w(i)} = \partial u / \partial w \Big|_{w=w(j)}$, we know that

$$\begin{aligned}
\partial E_p[u] / \partial r_i &= \sum_{k=1 \dots K} p_k \times \partial u / \partial w \Big|_{w=w(k)} \times [-2\beta (1 - \sum_{\ell \neq i, j} r_\ell)] \\
&\quad + p_i \times 2\beta \times \partial u / \partial w \Big|_{w=w(i)} - p_j \times 2\beta \times \partial u / \partial w \Big|_{w=w(j)} < 0
\end{aligned}$$

Therefore $r_i = r_j$ cannot be the optimum when $p_i < p_j$. Hence if $p_i \neq p_j$ then $r_i \neq r_j$. So the contrapositive is also true, that is $r_i = r_j$, then $p_i = p_j$. ■

Corollary 2.1: For the individual in Theorem 2, 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 Corollary 1.1, with r_k and $p_k, \forall k$, interchanged at all steps. ■

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

Proof. Using Lemma 2, if $p_i = 0$, then $\varepsilon_i = 0$ and $r_i = 0$. ■

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

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

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

Theorem 5: An individual with sufficiently high risk aversion will have a reported probability arbitrarily close to \hat{p}^* .

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

Therefore

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

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

$$\begin{aligned} w(k) &= \alpha + \beta 2r_k - \beta \sum_{i=1 \dots K} (r_i)^2 \\ w(k) &= \alpha + \beta 2(\hat{p}^* + \delta_k) - \beta \sum_{i \neq k} (\hat{p}^* + \delta_i)^2 - \beta (\hat{p}^* + \delta_k)^2 \end{aligned} \quad (12)$$

and

$$\begin{aligned} w^* &= \alpha + \beta 2r^* - \beta \sum_{i=1 \dots K} (r_i)^2 \\ w^* &= \alpha + \beta 2(\hat{p}^*) - \beta \sum_{i \neq k} (\hat{p}^* + \delta_i)^2 - \beta (\hat{p}^* + \delta_k)^2 \end{aligned} \quad (13)$$

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

$$\partial u / \partial w |_{w=v(k)} = 1 - 2c[w(k) - w^*] = 1 - 2c[\beta 2\delta_k] = 1 - 4c\beta\delta_k \quad (14)$$

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

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

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 $\lim_{c \rightarrow \infty} \delta_k = 0$ 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

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

If $p_k = 0$, then $\lim_{c \rightarrow 0} \delta_k = p_k - p^*$ and $\lim_{K \rightarrow \infty} r_k = p^* + \delta_k = p_k = 0$. 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 $\lim_{c \rightarrow \infty} r_k = \lim_{c \rightarrow \infty} p^* + \delta_k = 1/K$. That is, the reported probabilities approach a uniform distribution over the outcomes where the subjective probability is non-zero. ■

Theorem 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. 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:

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

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

$$\epsilon_k = \delta p_k \times \{\partial u^* / \partial w |_{w=n(k)} - E_p [\partial u^* / \partial w]\} / \{1 + \delta E_p [\partial u^* / \partial w]\}, \forall k = 1, \dots, K \quad (15)$$

Assume a random variable y with K possible states of nature. Define $E_r[y] = \sum_{k=1 \dots K} r_k y_k$ and $E_p[y] = \sum_{k=1 \dots K} p_k y_k$. Then the difference of the expected value of y under measures $\{r_1, r_2, \dots, r_k, \dots, r_K\}$ and $\{p_1, p_2, \dots, p_k, \dots, p_K\}$ is equal to $E_r[y] - E_p[y] = \sum_{k=1 \dots K} \epsilon_k y_k$. Substituting for ϵ_k using (15), 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

$$E_r[y] - E_p[y] = \delta \times \sum_{K=1 \dots K} p_k \{\partial u^* / \partial w |_{w=n(k)} - E_p [\partial u^* / \partial w]\} y_k$$

$$\begin{aligned}
&= \delta \times \{E_p [\partial u^*/\partial w \times y] - E_p [\partial u^*/\partial w] E_p [y]\} \\
&= \delta \times Cov_p[\partial u^*/\partial w, y] = Cov_p[\partial u/\partial w, y].
\end{aligned}$$



3. Some Evidence

These theoretical results were 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, Theorem 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 (Corollary 1.1). This result is important for tests of the Reduction of Compound Lotteries axiom of Subjective Expected Utility theory. In the laboratory we can present a stimulus for beliefs which provides a minimal basis for asymmetric beliefs, and for which we know, by design, the true mean of the stimulus. It is then a simple matter to compare that true stimulus with the average elicited belief.

Our experiments have a simple design. In 8 sessions we elicit subjective belief *distributions* about the true fraction of red balls in an urn filled with red and white balls. Although we know from other experiments that our subjects are generally risk averse over the stakes used here, in 4 of these sessions we also compare behavior with elicited subjective *probabilities* that a single red ball would be drawn from the *same* urn. This probability elicitation task is known to be one in which risk averse subjects would rationally and significantly distort their reports towards 1/2. By comparing reports in these 4 subjective probability elicitation experiments with the corresponding 4 subjective belief

distribution elicitation experiments, we can assess the practical significance of our claims about the weak effects of risk aversion on optimal reports in the latter setting.

Figures 7 and 8 illustrate the software interface we developed for subjects, allowing them to allocate 100 tokens across ten intervals corresponding to their beliefs on the true fraction of red balls in an urn. The stimulus was constructed in a standard manner. One urn was filled with ping pong balls numbered 1 through 99. This was common knowledge to all subjects, as the numbered balls were loaded in sight of all subjects. A “verifier” was selected from the subject pool at random, and paid to observe and verify that all procedures followed what was stated in the instructions. One of these numbered balls was selected, at random, from behind a screen, and the experimenter then filled a second urn with the number of red balls from the number drawn.⁷ The urn was filled with white balls so that the total number of balls was 100. The urn with 100 red and white balls was covered from view with a blanket, and placed in a prominent location for all subjects to see. After subjects were alerted to pay attention, the blanket was quickly removed and the urn spun for 10 seconds, then covered again. This visual display was the information that each subject received.

The computer interface in Figure 7 was then presented to subjects, allowing them to allocate tokens to reflect their subjective beliefs. The complete instructions are provided in an appendix, and were read out word for word prior to any choices being made. The subjects were told that there were no other salient, rewarded choices for them to make, avoiding possible confounds with the experimental payment protocol (see Harrison and Swarthout [2012] for discussion). The interface implements the quadratic scoring rule discussed earlier, with $\alpha=\beta=25$. Subjects could move the

⁷ The verifier observed everything taking place behind the screen. Further, the ball indicating the number of red balls in the second urn was placed in an envelope at the front of the room in full view of all subjects. The subjects knew that this ball would be revealed at the end of the experiment, along with a count of the number of red and white balls in the second urn. Both of these procedures were followed to convince subjects that the experimenter was not manipulating the process.

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.

Over 8 sessions we elicited subjective beliefs from 123 subjects, recruited from a general subject pool of undergraduates at Georgia State University. The stimulus, the number of red balls in the second urn, 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.

To illustrate the data, Figure 9 illustrates the elicited beliefs pooled over 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.⁸

Figure 10 reports the results across all eight sessions. With one exception, the elicited averages closely tracks the true averages. 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

⁸ 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.

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.⁹ 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.

We should also add that 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 Corollary 1.1 is not due to the risk neutrality of the subjects over these stakes.

Indeed, we can 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 quadratic scoring rule 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 also had separate subjects facing the same stimulus, but responding to the quadratic scoring rule for binary events shown in Figure 11.¹⁰ In this case the binary event was a single draw from the urn containing the red and white balls. Although all subjects within a given session are presented with the same physical stimuli, the two

⁹ 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%.

¹⁰ These subjects were briefly taken out of the room as the instructions specific to the scoring rule for the 10-event elicitation procedure, and *vice versa*. All subjects were in the room at the same time for the initialization of the urns and the later realizations of the event they were betting on.

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 in the binary scoring rule to be closer to 50% than to the true stimuli. This is perfectly consistent with our predictions, since risk aversion has a significantly biasing effect for the binary scoring rule, and virtually none for the 10-event scoring rule.

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 $\rho = 0$ (risk neutral)

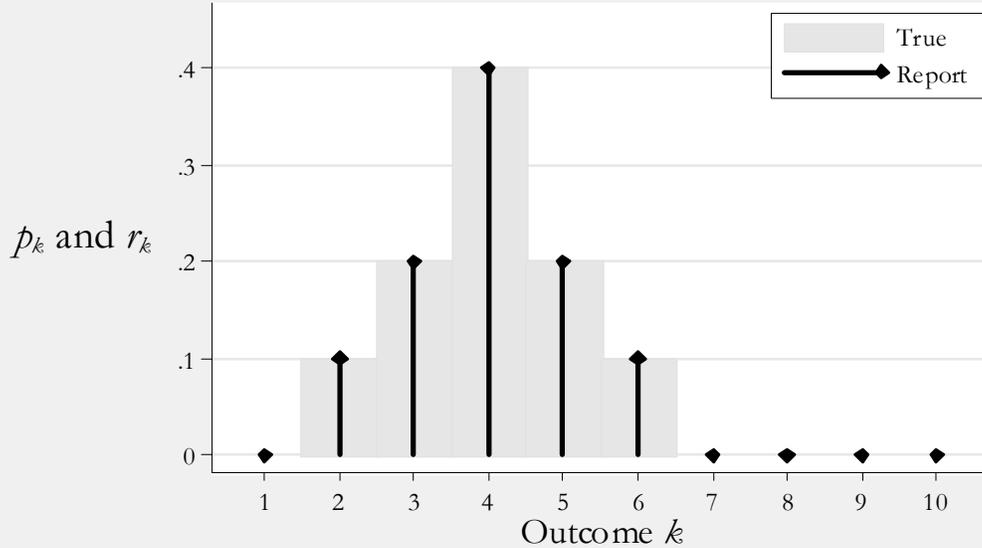


Figure 2: Optimal Reports Assuming Risk Aversion and Subjective Expected Utility

CRRA utility function with $\rho = 0.65$ (risk averse)

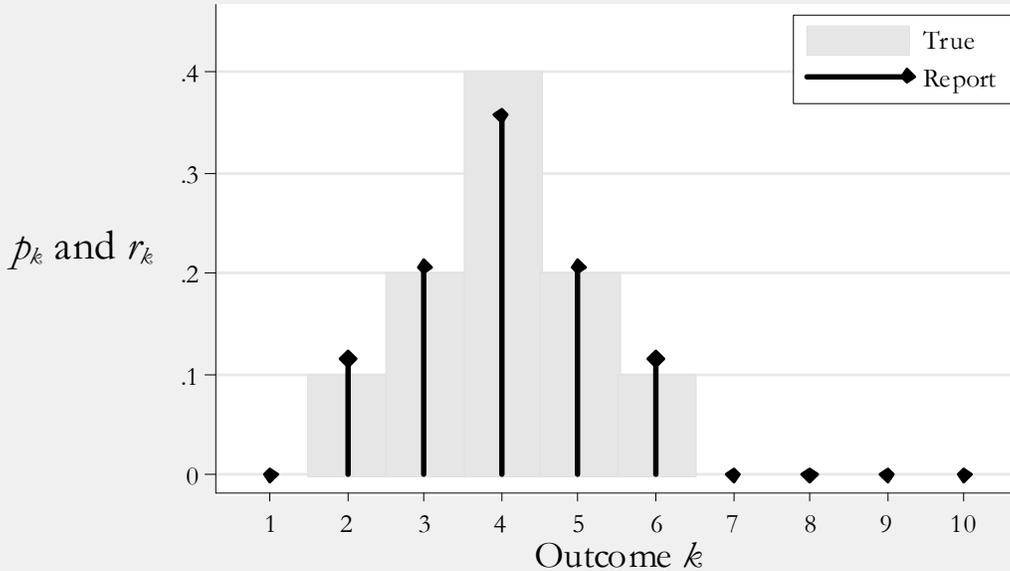


Figure 3: Optimal Reports Assuming Unimodal, Symmetric Beliefs and Subjective Expected Utility

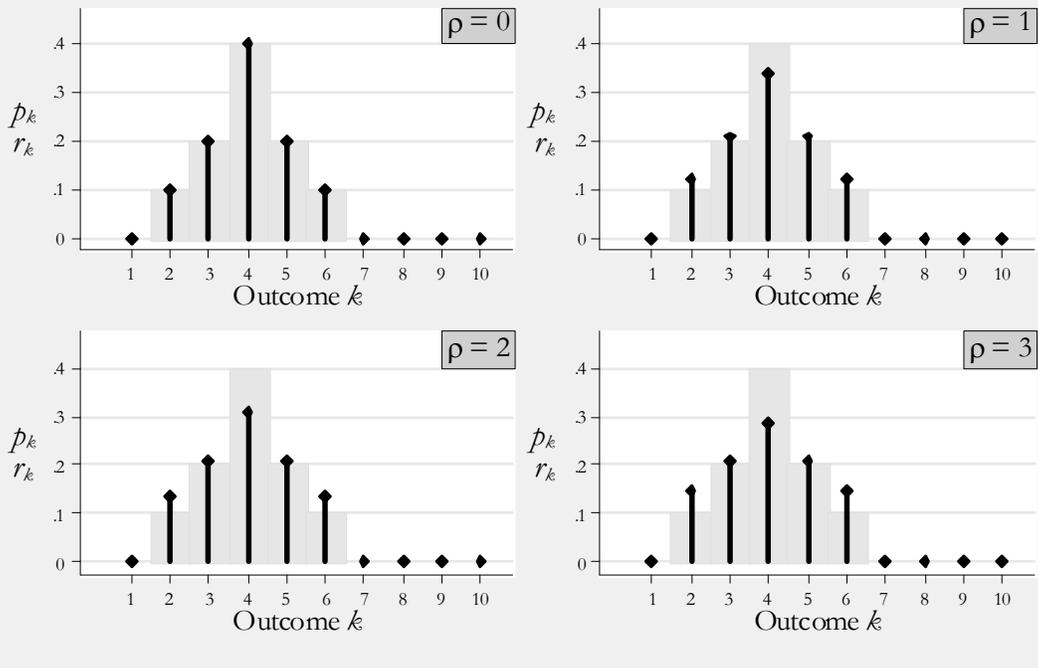


Figure 4: Optimal Reports Assuming Unimodal, Asymmetric Beliefs and Subjective Expected Utility

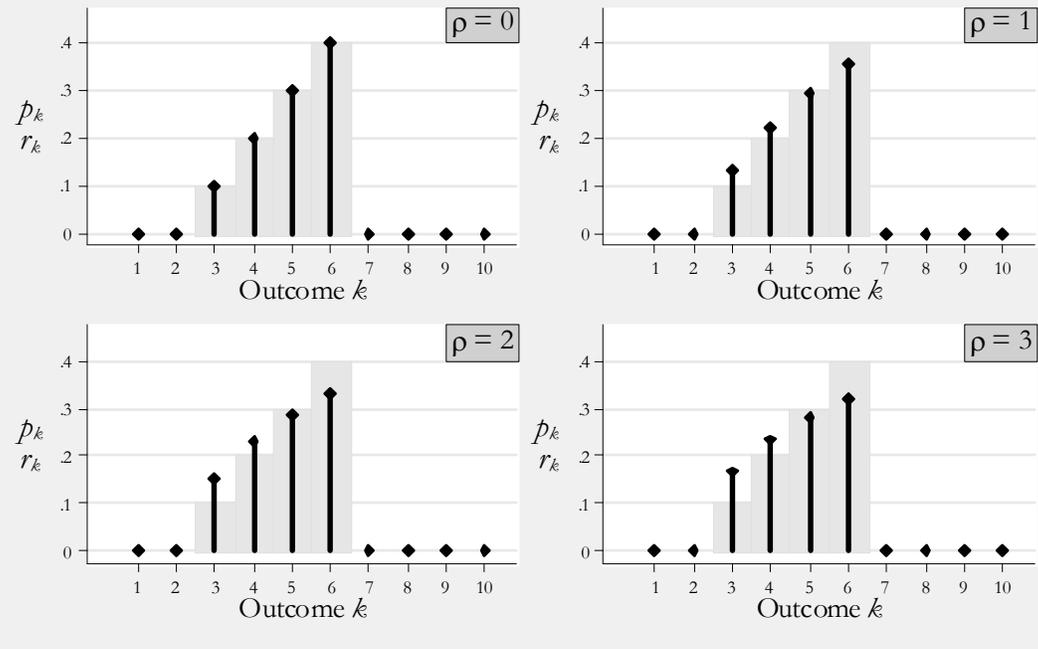


Figure 5: Optimal Reports Assuming Bimodal, Symmetric Beliefs and Subjective Expected Utility

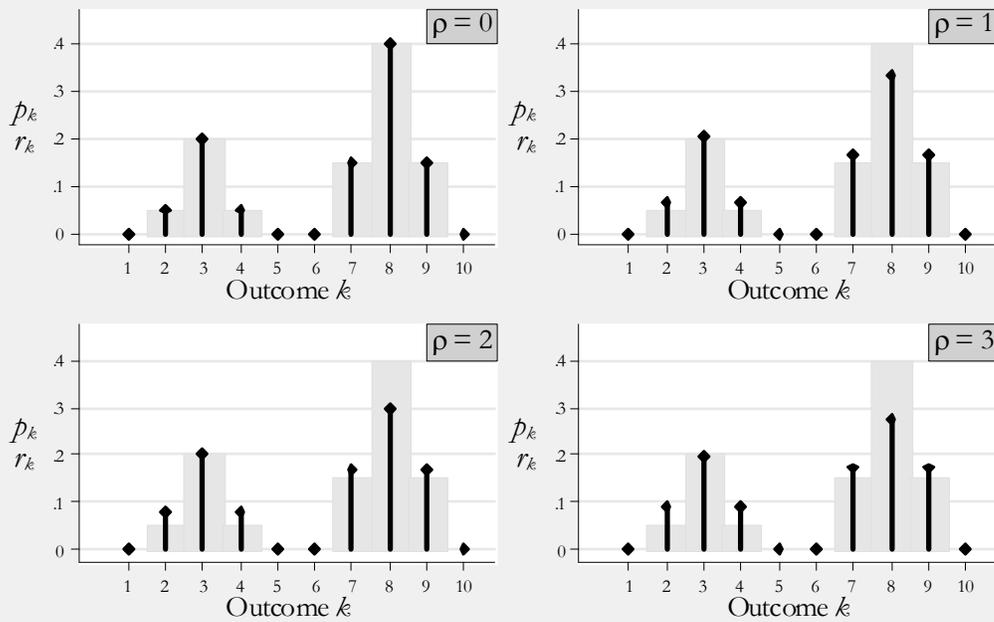


Figure 6: Difference Between True Average and Reported Average with Asymmetric Beliefs and Varying Relative Risk Aversion

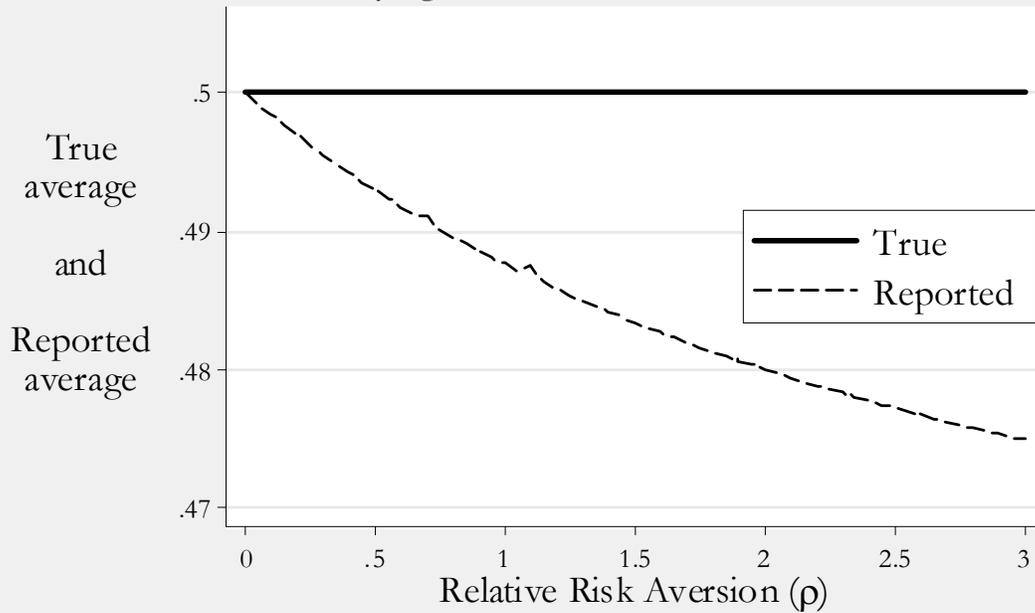


Figure 7: Initial Belief Elicitation Interface



Figure 8: Typical Belief Elicitation Response

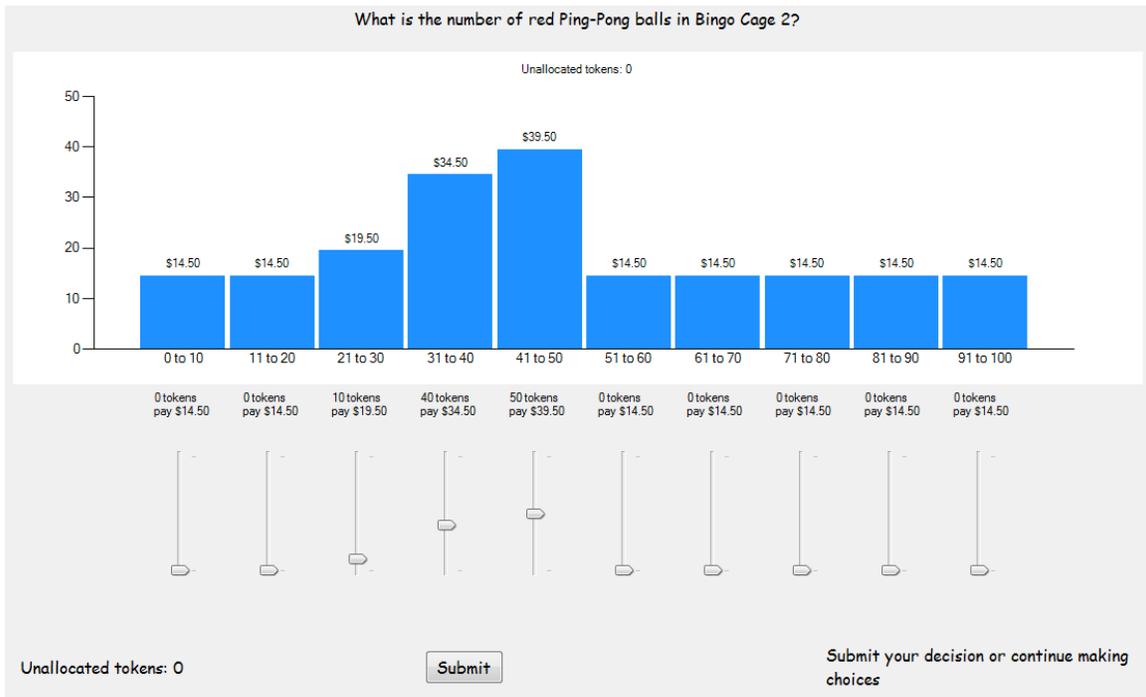


Figure 9: Elicited Subjective Distribution Pooled Over 15 Subjects and With True Percent of 69%

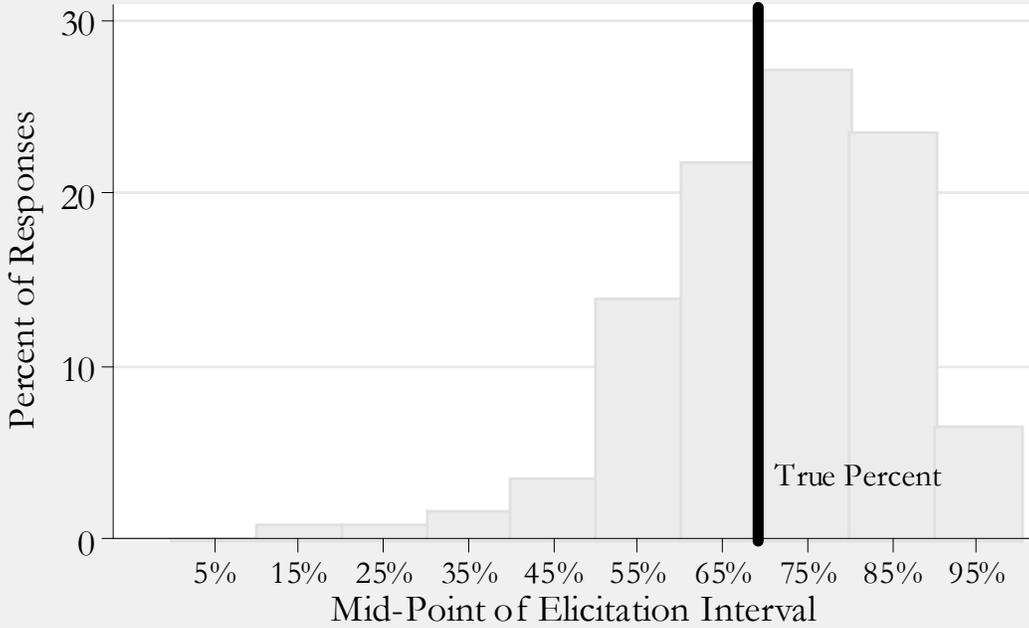


Figure 10: Average Elicited Subjective Belief Distribution

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

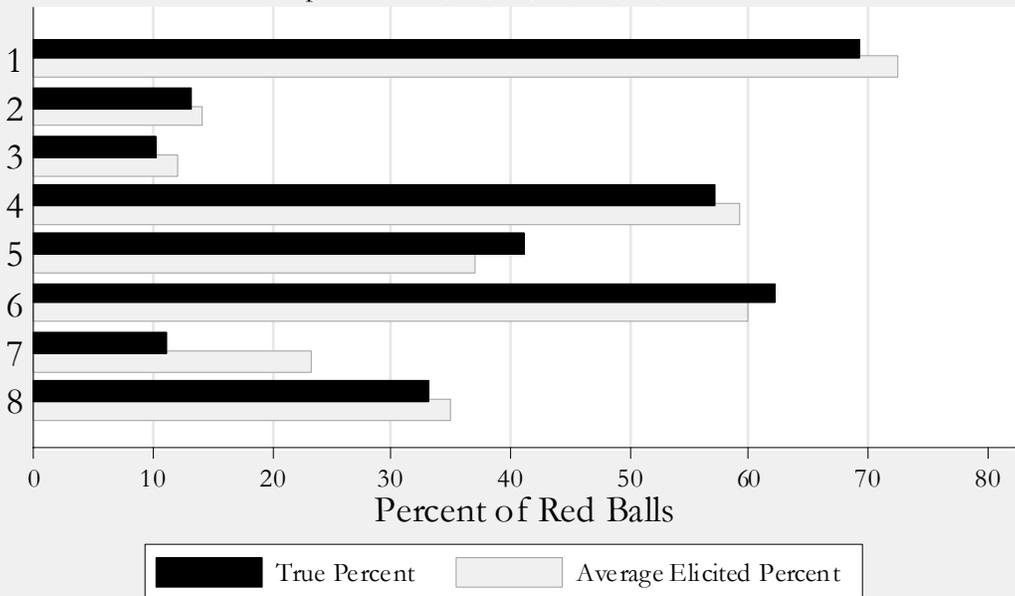


Figure 11: Subjective Probability Elicitation Interface

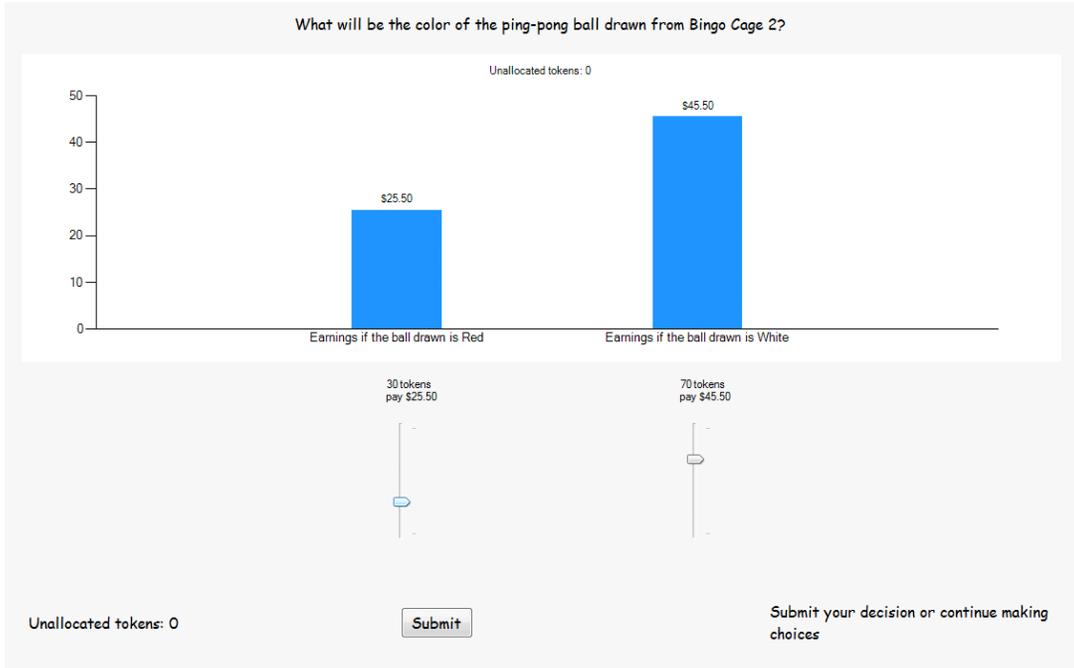
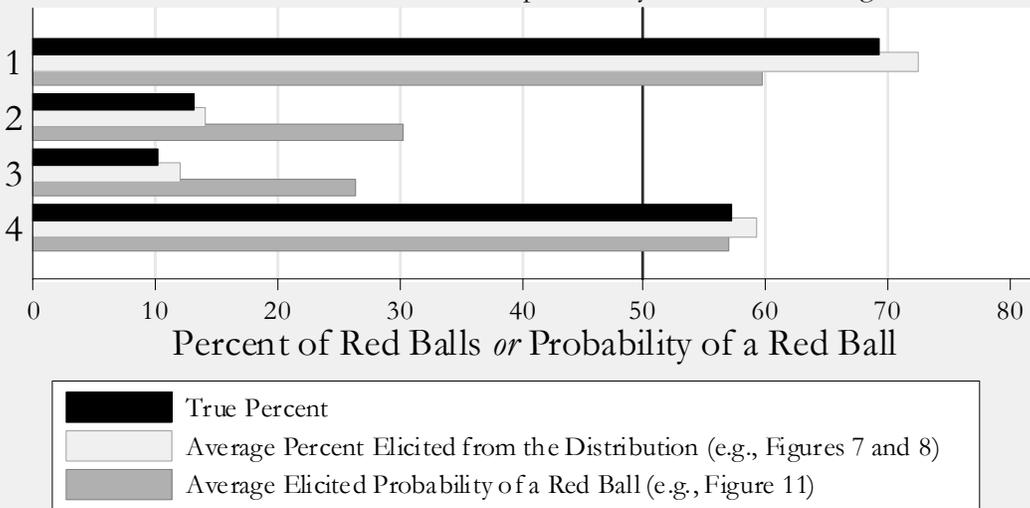


Figure 12: Average of Elicited Subjective Belief Distributions and Average Elicited Subjective Probabilities

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.



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.
- 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.
- 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, 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 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.
- Köszegi, Botond, and Rabin, Matthew, "Revealed Mistakes and Revealed Preferences," in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook* (New York: Oxford University Press, 2008).
- Mathieson, James E., and Winkler, Robert L., "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22(10), June 1976, 1087-1096.
- 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.
- 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.