Recovering Subjective Probability Distributions

by

Glenn W. Harrison and Eric R. Ulm †

February 2016

ABSTRACT.

An individual reports subjective beliefs over continuous events using a proper scoring rule, such as the Quadratic Scoring Rule. Under mild additional assumption, it is known that these reports reflect latent subjective beliefs if the individual is risk neutral and obeys Subjective Expected Utility (SEU) theory. It is also known that these reports are close to latent subjective beliefs if the individual obeys SEU and has a concave utility function in the range typically observed. We extend these results in three ways. First, we show how to fully recover latent subjective beliefs if the individual obeys SEU and has any concave utility function. Second, and significantly for practical purposes, we demonstrate how to fully recover latent subjective beliefs if the individual is known to distort probabilities into decision weights using Rank Dependent Utility (RDU) theory. We illustrate with a range of beliefs elicited from individuals, and for whom we also have estimates of their risk preferences to allow us to identify SEU and RDU individuals. Third, we generalize all results for the complete class of proper scoring rules. These theoretical results and empirical applications significantly widen the domain of applicability of proper scoring rules for eliciting latent subject belief distributions.

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); and Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, USA (Ulm). Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contacts: gharrison@gsu.edu and eulm@gsu.edu.

An individual reports subjective beliefs over continuous events using a proper scoring rule, such as the popular Quadratic Scoring Rule. Under some mild additional assumption, it has been known since Matheson and Winkler [1976] that these reports reflect latent subjective beliefs if the individual is risk neutral and obeys Subjective Expected Utility (SEU) theory. It is also now known that these reports are "close" to latent subjective beliefs if the individual obeys SEU and has a concave utility function in the range observed over typical payments in experiments.

We extend these theoretical results in three ways. First, we demonstrate how to *exactly* recover latent subjective belief distributions if the individual obeys SEU. Thus one does not have to rely on approximation results from theory that show that these are likely to be "close," and one can demonstrate exactly how close they are on an individual basis. Second, and more significantly, we demonstrate how to recover latent subjective belief distributions if the individual is known to distort probabilities into decision weights using Rank Dependent Utility (RDU) theory. Although this extension is relatively simple to state as a theoretical matter, it provides a constructive basis for exactly recovering latent subjective belief distributions for individuals that do not behave consistently with SEU.[1] Third, we generalize these results to the compete class of proper scoring rules, of which the QSR is just the most popular.

We illustrate the application of these theoretical results by recovering the latent subjective belief distributions from observed reports from individuals in experiments, and for whom we also have individual estimates of their risk preferences. We find that the recovered beliefs for EUT-

[1] This is not the same as eliciting a series of binary subjective probabilities and "knitting together" an elicited subjective belief distribution. The elicitation problem for subjective probabilities over binary events has been well-studied, and operational methods for recovering latent subjective probabilities for risk-dependent scoring rules developed (e.g., Offerman, Sonnemans, van de Kuilen and Wakker [2009] and Andersen, Harrison, Fountain and Rutström [2014]). Our approach is to elicit the distribution in one task, not in a number of independent tasks. Undertaking a series of binary elicitations runs the risk of order effects, or the risk of elicited probabilities not summing to 1. It is also much harder to correctly estimate standard errors for the inferred latent distribution when making a series of independent inferences about binary slices of the underlying distribution. Of course, in future work it would be interesting to compare the consistency of elicited distribution from one task with constructed distribution from a series of binary elicitations.

consistent individuals are, as expected, very close to the observed reports. We also show that the recovered belief distributions of RDU-consistent individuals exhibit first-order differences from the observed reports. The extent of the distortion between observed and recovered beliefs depends on the dispersion of observed beliefs as well as the extent of probability weighting, each of which can vary across different belief questions and individuals.

We focus on the recovery of subjective belief distributions assuming either EUT or RDU as the underlying model of decision-making under risk. Our results extend to recovering subjective belief distributions if the individual exhibits risk preferences consistent with Cumulative Prospect Theory and models of uncertainty and ambiguity aversion. For now we focus on choices in the gain frame and models of risk aversion.

These theoretical results and empirical applications significantly widen the domain of applicability of proper scoring rules for eliciting latent subject belief distributions. Apart from intrinsic interest in knowing the subjective belief distributions of individuals, our methods help make Bayesian inference more operational by recovering latent priors when one cannot be sure that the individual is risk neutral.

1. 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. We do *not* assume symmetric subjective distributions, nor do we assume that the distribution is even unimodal.

A. Background and Notation

Let the decision maker report his subjective beliefs in a discrete version of a QSR for continuous distributions (Matheson and Winkler [1976]).[2] Partition the domain into K intervals, and denote as r_k the report of the likelihood that the event falls 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 reports for each interval, $\{ r_1, r_2, ..., r_k ,..., r_K \}$ such that $r_k \geq 0\ \forall k$ and $\Sigma_{i=1...K} (r_i) = 1$.

If k is the interval in which the actual value lies, then the payoff score is defined by Matheson and Winkler [1976; p.1088, equation (6)]: $S = (2 \times r_k) - \Sigma_{i=1...K} (r_i)^2$. 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.[3]

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 following scoring rule for each report in interval k

$$\alpha + \beta\, [\, (2 \times r_k) - \Sigma_{i=1...K} (r_i)^2 \,], \tag{0}$$

where we initially assumed $\alpha=0$ and $\beta=1$. We can assume $\alpha>0$ and $\beta>0$ to get the payoffs to any positive 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 and 2 illustrate one the QSR, which

[2] Alternative scoring rules could be characterized, and we provide proof that our results generalize to the class of proper scoring rules. The QSR is the most popular scoring rule in practice, and all of the practical issues of recovering beliefs can be directly examined in that context. For instance, Andersen, Fountain, Harrison and Rutström [2014] 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.

[3] Take some examples, assuming $K = 4$. What if the subject has very tight subjective beliefs and allocates all of the weight to 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 tight subjective beliefs that are 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. Assuming that the subject has very diffuse subjective beliefs and allocates 25% of the weight to each interval, the score is less than 1: $S = (2 \times \frac{1}{4}) - ((\frac{1}{4})^2 + (\frac{1}{4})^2 + (\frac{1}{4})^2 + (\frac{1}{4})^2) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4} < 1$. 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.

we will use in experiments, for $\alpha = \beta = 25$ and K=10.

We restate Lemma 1 from Harrison, Martínez-Correa, Swarthout and Ulm [2012]:

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 $\theta(k) = \alpha + \beta 2r_k - \beta \sum_{i=1...K} (r_i)^2$ be the scoring rule that determines earnings θ if state k occurs. Assume that the individual behaves consistently with SEU. If the individual has a utility function $u(\cdot)$ that is continuous, twice differentiable, increasing and concave and maximizes expected utility over actual subjective probabilities, the actual and reported probabilities must obey the following system of equations:

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

Our main theoretical result is a generalization of Lemma 1 for RDU individuals, who distort probabilities and employ "decision weights" when evaluating ranked payoff outcomes.

We state parametric versions of EUT and RDU decision making over objective probabilities, to introduce notation and basic concepts. Nothing hinges on the parametric assumptions, although the parametric forms assumed are standard in the literature.

Assume that utility of income in an elicitation is defined by

$$U(x) = x^{(1-s)}/(1-s) \tag{2}$$

where x is the lottery prize and $s \neq 1$ is a parameter to be estimated. For s=1 assume U(x)=ln(x) if needed. Thus s is the coefficient of CRRA for an EUT individual: s=0 corresponds to risk neutrality, s<0 to risk loving, and s>0 to risk aversion. Of course, risk attitudes under RDU depend on more than the curvature of the utility function.

Let there be J possible outcomes in a lottery defined over objective probabilities commonly implemented in experiments. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i:

$$EU_i = \sum_{j=1,J} [\, p(x_j) \times U(x_j) \,]. \tag{3}$$

The RDU model of Quiggin [1982] extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same parametric specification (2) considered for EUT.[4] To calculate decision weights under RDU one replaces expected utility defined by (3) with RDU

$$RDU_i = \sum_{j=1,J} [w(p(x_j)) \times U(x_j)] = \sum_{j=1,J} [w_j \times U(x_j)] \tag{4}$$

where

$$w_j = \omega(p_j + ... + p_J) - \omega(p_{j+1} + ... + p_J) \tag{5a}$$

for j=1,... , J-1, and

$$w_j = \omega(p_j) \tag{5b}$$

for j=J, with the subscript j ranking outcomes from worst to best, and $\omega(\cdot)$ is some probability weighting function.

We consider three popular probability weighting functions. The first is the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ:

$$\omega(p) = p^\gamma \tag{6}$$

So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \ \ \forall p$ and hence the "RDU EV" weighted by $\omega(p)$ instead of p has to be less than the EV weighted by p.

The second probability weighting function is the "inverse-S" function popularized by Tversky and Kahneman [1992]:

$$\omega(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma} \tag{7}$$

This function exhibits inverse-S probability weighting (optimism for small p, and pessimism for large p) for $\gamma<1$, and S-shaped probability weighting (pessimism for small p, and optimism for large p) for $\gamma>1$.

[4] To ease notation we use the same parameter s because the context always make it clear if this refers to an EUT model or a RDU model.

The third probability weighting function is a general functional form proposed by Prelec [1998] that exhibits considerable flexibility. This function is

$$\omega(p) = \exp\{-\eta(-\ln p)^{\varphi}\}, \tag{8}$$

and is defined for $0<p\leq 1$, $\eta>0$ and $\varphi>0$.[5] When $\varphi=1$ this function collapses to the Power function $\omega(p) = p^{\eta}$.

B. Recovering Beliefs

We generalize Lemma 1 to include individuals that distort probabilities:

Lemma 2: 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 $\theta(k) = \alpha + \beta 2r_k - \beta \sum_{i=1...K} (r_i)^2$ be the scoring rule that determines earnings θ if state k occurs. Assume that the individual uses some probability weighting function $\omega(\cdot)$, leading to decision weights $w(\cdot)$ defined in the standard decumulative fashion of (5a) and (5b). Assume that the individual behaves consistently with RDU, applied to subjective probabilities. If the individual has a utility function $u(\cdot)$ that is continuous, twice differentiable, increasing and concave and maximizes rank-dependent utility over weighted subjective probabilities, the actual and reported probabilities must obey the following system of equations:

$$w(p_k) \times \partial u/\partial\theta \,|_{\theta=\theta(k)} - \sum_{j=1,K} \{ w(p_j) \times r_j \times \partial u/\partial\theta \,|_{\theta=\theta(j)} \} = 0, \forall\, k = 1,..., K \tag{9}$$

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

$$\mathbf{Max}_{\{r\}}\ E_{w(p)} [\, u(\theta) \,] \text{ subject to } \sum_{i=1...K} (r_i) = 1 \tag{10}$$

where $E_{w(p)} [\, u(\theta) \,] = \sum_{j=1...K} w(p_j) \times u[\, \alpha + \beta 2r_j - \beta \sum_{i=1...K} (r_i)^2 \,]$. In some experimental configurations there may be K additional constraints: $r_i \geq 0$ for $i = 1,..., K$. These constraints are not included in (10) because they are automatically satisfied by the solution (9) for both risk-averse and risk-loving individuals.

Problem (10) can be solved by maximizing the Lagrangian

[5] Many apply the Prelec [1998; Proposition 1, part (B)] function with constraint $0 < \varphi < 1$, which requires that the probability weighting function exhibit subproportionality. Contrary to received wisdom, many individuals exhibit estimated probability weighting functions that violate subproportionality, so we use the more general specification from Prelec [1998; Proposition 1, part (C)], only requiring $\varphi > 0$, and let the evidence for an individual determine if the estimates φ lies in the unit interval.

$$\mathcal{L} = \sum_{j=1...K} w(p_j) \times u[\,\alpha + \beta 2r_j - \beta \sum_{i=1...K} (r_i)^2\,] - \lambda\,[\,\sum_{i=1...K} (r_i) - 1] \tag{11}$$

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[6]

$$\partial\mathcal{L}/\partial r_k = \sum_{j=1...K} (\, w(p_j) \times \partial u(\theta(j))/\partial r_k\,) - \lambda = 0, \forall\ k = 1,..., K \tag{12}$$

where $\partial u(\theta(j))/\partial r_k = \partial u/\partial\theta\,|_{\theta=\theta(j)} \times (2\beta\delta_{jk} - 2\beta \times r_k)$ 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 (11) with respect to the Lagrangian constant:

$$\sum_{i=1...K} (r_i) - 1 = 0. \tag{13}$$

We can simplify the K equations in (12) as:

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

or

$$w(p_k) \times (\,\partial u/\partial\theta\,|_{\theta=\theta(k)}\,) - r_k\, E_{w(p)}\,[\,\partial u/\partial\theta\,] = \lambda/2\beta, \forall\ k = 1,..., K. \tag{12$'$}$$

Summing over the K first-order conditions we get

$$E_{w(p)}\,[\,\partial u/\partial\theta\,|_{\theta=\theta(k)}\,] - \sum_{k=1...K} r_k\, E_{w(p)}\,[\,\partial u/\partial\theta\,] = K\lambda/2\beta\,. \tag{14}$$

Notice that $\sum_{k=1...K} r_k\, E_{w(p)}\,[\,\partial u/\partial\theta\,] = E_{w(p)}\,[\,\partial u/\partial\theta\,]$ because the expectation term is a constant and because of (13). Then (14) implies that $K\lambda/2\beta = 0$, which can only be satisfied if $\lambda = 0$ since K>0 and β>0. This result and (12$'$) implies that the solution to problem (10) must satisfy the following K conditions:

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

The application of (9) is straightforward. If the reports r_k are given from observation of experimental data, the partial derivatives are fixed and independent of the decision weights $w(p_k)$, so this is a linear system of equations in the unknown decision weights. These equations can be solved using standard linear algebra techniques. Although it turns out the equations are linearly dependent, we can replace any one of them with $\sum_{k=1,\,K} \{\, w(p_k)\,\} = 1$ to remove the redundancy and obtain a unique solution.

A numerical example illustrates the basic ideas. Assume K=10 bins. An individual reports

[6] The differentiation here is done with respect to reported values. If the reports for a set of bins are precisely equal, the wealth outcomes are equal. In this case the bins are combined and the derived probability is distributed equally among all members of the set.

30, 45 and 25, out of 100 tokens, in bins 3, 4 and 5, leaving 0 tokens in the other 7 bins. Thus we have $r_1 = 0.00$, $r_2 = 0.00$, $r_3 = 0.30$, $r_4 = 0.45$, $r_5 = 0.25$, $r_6 = 0.00$, $r_7 = 0.00$, $r_8 = 0.00$, $r_9 = 0.00$ and $r_{10} = 0.00$. Assume the QSR given by (0) with $\alpha = \beta = 25$, consistent with the experiments reported later. Let the CRRA utility function be given by (2) with s = 0.77, consistent with evidence from a wide array of experiments, so that $\partial u/\partial\theta = \theta^{-s}$.[7] For RDU individuals further assume the inverse-S probability weighting function (7) with $\gamma = 0.5$ and inverse function $p^{\gamma} [(1-p)^{\gamma} + p^{\gamma}]^{-1/\gamma}$.[8]

The derivative of the utility function is only relevant for the three bins with positive reports, since the decision weights will be 0 for the other bins with zero reports. The 3 equations in 3 unknowns are then

$$0.049591 \times w(p_3) - 0.018 \times w(p_4) - 0.02267 \times w(p_5) = 0 \quad (15a)$$
$$-0.03188 \times w(p_3) + 0.032997 \times w(p_4) - 0.034 \times w(p_5) = 0 \quad (15b)$$
$$w(p_3) + w(p_4) + w(p_5) = 1 \quad (15c)$$

The numerical values in these 3 equations are direct applications of (9). Taking (15a) as an example, we have

$$0.049591 = 0.7 \times 0.07084443 = (1-r_3) \times [25 + (50 \times 0.30) - (25 \times (0.30^2 + 0.45^2 + 0.75^2))]^{-0.77}$$
$$-0.018 = -0.3 \times 0.0059994478 = (-r_3) \times [25 + (50 \times 0.45) - (25 \times (0.30^2 + 0.45^2 + 0.75^2))]^{-0.77}$$
$$-0.02267 = -0.3 \times 0.075562439 = (-r_3) \times [25 + (50 \times 0.25) - (25 \times (0.30^2 + 0.45^2 + 0.75^2))]^{-0.77}.$$

We solve (15a), (15b) and (15c) for decision weights $w(p_3) = 0.281486$, $w(p_4) = 0.498589$ and $w(p_5) = 0.219925$. If the individual were an EUT maximizer, we would be finished and these weights would be the individual's implied subjective probabilities. As expected from the results of Harrison, Martinez-Correa, Swarthout and Ulm [2012], and the assumed value of s, the differences between these weights and the observed reports are small. The reported mean is 34.5 if the bins intervals are

---

[7] For the CARA utility function $U(\theta) = \exp(-k\,\theta)$ the partial is $k \exp(-\theta k)$, and for the Expo-Power utility function $U(\theta) = [1-\exp(-\alpha\,\theta^{1-s})]/\alpha$ the partial is $\exp\{-\theta^{(1-s)}\alpha\}\ (1-s)\ \theta^{-s}$.

[8] Since (7) is not monotonic for $\gamma < 0.278$, as noted by Rieger and Wang [2006; §1.2], we assume values of γ for which it is monotonic, and the inverse function is uniquely defined. This is a reasonable *a priori* restriction given the available empirical evidence for values of γ. For the power probability weighting function (6) the inverse function is $p^{(1/\gamma)}$, and for the Prelec probability weighting function (8) the inverse function is $\exp\{ (-1)^{(1+(1/\varphi))}\ \eta^{(-1/\varphi)}\ (\ln p)^{1/\varphi} \}$.

0 to 10, 11 to 20, ..., 91 to 100, and the subjective mean is 34.38439.

The next step is to extract the probabilities from the decision weights if the individual was known to be an RDU maximizer. We first sort the outcomes from lowest payoff to highest payoff. For a given individual and elicitation, this is the same as sorting from lowest to highest report in terms of tokens, or sorting from lowest to highest decision weight. We sort to $w(p_5) = 0.219925$, $w(p_3) = 0.281486$ and $w(p_4) = 0.498589$. We then apply a decumulative process to extract the cumulative distribution function for the probabilities. For example, p_4 produces the largest decision weight, since bin 4 was allocated the most tokens, and the relevant probability of being in bin 4 is then $\omega^{-1}(0.498589) = 0.802313$. The probability of being in bin 3 is then $\omega^{-1}(0.281486 + 0.498589) - \omega^{-1}(0.498589) = 0.179121$, since bin 3 was allocated the second-highest number of tokens, and the residual probability of being in bin 5 is then $\omega^{-1}(0.219925 + 0.281486 + 0.498589) - \omega^{-1}(0.281486 + 0.498589) = 0.018566$. These probabilities must finally be "de-sorted" to connect with the appropriate bin, so $p_3 = 0.179121$, $p_4 = 0.802313$ and $p_5 = 0.018566$. These are significant, first order differences, relative to the second-order effect of risk-aversion. The subjective mean in this case is 33.39445, noticeably different from the reported mean of 34.5.[9]

C. Generalization

Proposition 1: Lemma 1 generalizes to include all proper scoring rules. Hence all of the results that flow from Lemma 1 also generalize.

To prove Proposition 1 we must first prove Theorem 1, below, which is interesting in its own right. Lemmas 1 and 2 then follow for all proper scoring rules. We follow Armantier and Treich [2013] who proved the result for 2 elicitation bins. We prove an analogous theorem for an

[9] It is possible in some cases for the probabilities and weights derived in this fashion to violate first order stochastic dominance. The violations are in most cases small in terms of certainty equivalent, and subjects with extreme, *a priori* unreasonable preferences have been removed from the analysis.

arbitrary number of bins.

Define a scoring rule S where $S_1(r_1,..., r_n)$, $S_2(r_1,..., r_n)$,..., and $S_n(r_1,..., r_n)$ represent the payoffs for each of the possible states of nature 1,..., n. S_k is the payoff if state k is realized after reports r_1,..., r_n, where $r_n = 1 - \sum_{i=1...n-1} r_I$. Let

$$f(p_1,..., p_n; r_1,..., r_n) = \sum_{i=1...n} p_i S_i(r_1,..., r_n).$$

A scoring rule is "proper" if the maximizing arguments are $r_i = p_i$ for all i. Hence a risk-neutral decision maker will report truthfully, bypassing the need for a solution to the "recovery" problem solved by Lemma 2.

Theorem 1: A scoring rule is proper if and only if there exists a function $g(q_1,..., q_{n-1})$ with conditions on the second derivatives guaranteeing uniqueness and maximization such that

$$S_n(q_1,..., q_{n-1}) = g - \sum_{j=1...n-1} q_j \, \partial g/\partial q_j$$

and $$S_j(q_1,..., q_{n-1}) = S_n(q_1,..., q_{n-1}) + \partial g/\partial q_j \text{ for } j \in [1,n\text{-}1].$$

Notice that q_n is not an argument in the functions anymore because the latter is defined by q_1,..., q_{n-1}.

Proof: Necessity (only if).

Let $g(q_1,..., q_{n-1}) = \max_{\{r^*\}} f(q_1,..., q_{n-1}; r_1,..., r_{n-1})$ where $r^* = \{r_1^*, r_2^*,..., r_{n-1}^*\}$ is the vector of reports that maximizes the function f. By the envelope theorem, we see that

$$\begin{aligned}\partial g/\partial q_j &= \partial f(q_1,..., q_{n-1}; r_1,..., r_{n-1})/\partial q_j|_{r^i = q^i \forall i} \\ &= S_j(q_1,..., q_{n-1}) - S_n(q_1,..., q_{n-1}).\end{aligned}$$

Notice that $S_n(q_1,..., q_{n-1})$ comes from a $(1 - \sum_{i=1...n-1} r_i)\, S_n(r_1,..., r_{n-1})$ term. Therefore

$$S_j(q_1,..., q_{n-1}) = S_n(q_1,..., q_{n-1}) + \partial g/\partial q_j.$$

Substituting these into the formula for g, we get

$$g(q_1,..., q_{n-1}) = \max_{\{r^*\}} f(q_1,..., q_{n-1}; r_1,..., r_{n-1}) = f(q_1,..., q_{n-1}; q_1,..., q_{n-1}),$$

since S is a proper scoring rule. Therefore,

$$\begin{aligned}g(q_1,..., q_{n-1}) &= \sum_{j=1...n-1} q_j\, [S_n(q_1,..., q_{n-1}) + \partial g/\partial q_j] + (1 - \sum_{j=1...n-1} q_j)\, S_n(q_1,..., q_{n-1}) \\ &= S_n(q_1,..., q_{n-1}) + \sum_{j=1...n-1} q_j \, \partial g/\partial q_j.\end{aligned}$$

Rearranging terms we get

$$S_n(q_1,..., q_{n-1}) = g(q_1,..., q_{n-1}) - \sum_{j=1...n-1} q_j \, \partial g/\partial q_j.$$

Proof: Sufficiency (if).

$$f(q_1,..., q_{n-1}; r_1,..., r_{n-1}) = \Sigma_{i=1...n-1}\, q_i\, S_i\,(r_1,..., r_{n-1}) + (1 - \Sigma_{i=1...n-1}\, q_i)\, S_n\,(r_1,..., r_{n-1})$$
$$= \Sigma_{i=1...n-1}\, q_i\, [g - \Sigma_{j=1...n-1}\, r_j\ \ \partial g/\partial r_j + \partial g/\partial r_i]$$
$$+ (1 - \Sigma_{i=1...n-1}\, q_i)\,(g - \Sigma_{j=1...n-1}\, r_j\ \ \partial g/\partial r_j)$$

We maximize f by setting the n-1 first order conditions to zero:

$$\partial f/\partial r_k = \Sigma_{i=1...n-1}\, q_i\, [\partial g/\partial r_k - \Sigma_{j=1...n-1}\, r_j\ \ \partial^2 g/\partial r_j\,\partial r_k - \partial g/\partial r_k + \partial^2 g/\partial r_i\,\partial r_k]$$
$$+ (1 - \Sigma_{i=1...n-1}\, q_i)\,(\partial g/\partial r_k - \Sigma_{j=1...n-1}\, r_j\ \ \partial^2 g/\partial r_j\,\partial r_k - \partial g/\partial r_k) = 0.$$

This gives us

$$- \Sigma_{i=1...n-1}\, q_i\ \ \Sigma_{j=1...n-1}\, r_j\,\partial^2 g/\partial r_j\,\partial r_k + \Sigma_{i=1...n-1}\, q_i\,\partial^2 g/\partial r_i\,\partial r_k - \Sigma_{j=1...n-1}\, r_j\,\partial^2 g/\partial r_j\,\partial r_k$$
$$+ \Sigma_{i=1...n-1}\, q_i\ \ \Sigma_{j=1...n-1}\, r_j\,\partial^2 g/\partial r_j\,\partial r_k = 0.$$

Cancelling terms, we obtain

$$\Sigma_{i=1...n-1}\, q_i\,\partial^2 g/\partial r_i\,\partial r_k - \Sigma_{j=1...n-1}\, r_j\,\partial^2 g/\partial r_j\,\partial r_k = 0.$$

Changing the index from j to i in the second summation of the first order condition above we have

$$\Sigma_{i=1...n-1}\,(q_i - r_i)\,\partial^2 g/\partial r_i\,\partial r_k = 0. \tag{16}$$

This system consists of n-1 equations (indexed by k) in the n-1 unknowns $(q_i - r_i)$ indexed by i . One solution is clearly $q_i - r_i = 0$ (or $q_i = r_i$) for all i. Thus, the scoring rule S is *proper*.

There must be conditions on the second derivatives of g such that this solution is *unique* and *maximizes*, rather than minimizes, f. ■

Now we can prove Lemma 1 for general proper scoring rules.

Suppose an individual is now trying to maximize utility $V(p_1,..., p_{n-1}; r_1,..., r_{n-1})$ rather than money $f(p_1,..., p_{n-1}; r_1,..., r_{n-1})$. Suppose a utility function of wealth $u(W)$ and probability weights $w(p)$. We have

$$V(p_1,..., p_{n-1}; r_1,..., r_{n-1}) = \Sigma_{j=1...n-1}\, w(p_j)\, u(S_j\,(p_1,..., p_{n-1})) + w(p_n)\, u(S_n\,(p_1,..., p_{n-1})),$$

where $\Sigma_{j=1...n}\, p_j\ = 1$. We solve the following n-1 first-order conditions to maximize:

$$\partial V/\partial r_k = \ \Sigma_{j=1...n-1}\, w(p_j)\,\partial u/\partial W|_{S_j}\,\partial S_j\,/\partial r_k + w(p_n)\,\partial u/\partial W|_{S_n}\,\partial S_n\,/\partial r_k.$$

Now, since $S_j = S_n + \partial g/\partial r_j$, we see

$$\partial S_j/\partial r_k = \partial S_n/\partial r_k + \partial^2 g/\partial r_j \partial r_k$$

and $$\partial V/\partial r_k = \Sigma_{j=1...n}\, w(p_j)\,\partial u/\partial W|_{S_j}\,\partial S_n/\partial r_k + \Sigma_{j=1...n-1}\, w(p_j)\,\partial u/\partial W|_{S_j}\,\partial^2 g\,/\partial r_j\,\partial r_k = 0$$
$$= \partial S_n/\partial r_k\,\Sigma_{j=1...n}\, w(p_j)\,\partial u/\partial W|_{S_j}\ + \Sigma_{j=1...n-1}\, w(p_j)\,\partial u/\partial W|_{S_j}\,\partial^2 g\,/\partial r_j\,\partial r_k = 0$$

$$= \partial S_n/\partial r_k\, E_{w(p)}\,[\partial u/\partial W] + \Sigma_{j=1\ldots n-1}\, w(p_j)\, \partial u/\partial W|_{s^j}\, \partial^2 g\,/\partial r_j\, \partial r_k = 0$$

where $E_{w(p)}$ [.] denotes the expectations operator under probability measure $w(p) = \{w(p_1), \ldots , w(p_n)\}$. Now, since $S_n = g - \Sigma_{j=1\ldots n-1}\, r_j\, \partial g/\partial r_j$, we get

$$\partial S_n/\partial r_k = \partial g/\partial r_k - \Sigma_{j=1\ldots n-1}\, r_j\, \partial^2 g\,/\partial r_j\, \partial r_k - \partial g/\partial r_k = - \Sigma_{j=1\ldots n-1}\, r_j\, \partial^2 g\,/\partial r_j\, \partial r_k,$$

so $$\partial V/\partial r_k = - \Sigma_{j=1\ldots n-1}\, r_j\, \partial^2 g\,/\partial r_j\, \partial r_k\, E_{w(p)}\,[\partial u/\partial W] + \Sigma_{j=1\ldots n-1}\, w(p_j)\, \partial u/\partial W|_{s^j}\, \partial^2 g\,/\partial r_j\, \partial r_k = 0.$$

Therefore, we obtain

$$\Sigma_{j=1\ldots n-1}\, [\, w(p_j)\, \partial u/\partial W|_{s^j} - r_j\, E_{w(p)}\,[\partial u/\partial W]]\, \partial^2 g\,/\partial r_j\, \partial r_k = 0. \tag{17}$$

Equation (17) looks just like equation (16) except the n-1 unknowns are

$$w(p_j)\, \partial u/\partial W|_{s^j} - r_j\, E_{w(p)}\,[\partial u/\partial W].$$

As before,

$$w(p_j)\, \partial u/\partial W|_{s^j} - r_j\, E_{w(p)}\,[\partial u/\partial W] = 0\ \forall\, j.$$

This is unique and maximizing from the convexity conditions on g. ■

Since Lemma 2 follows from Lemma 1, Proposition 1, that "All results that flow from Lemma 1 also generalize," has been proved.[10]

2. Experimental Design

The theory we have developed implies that we need two experimental tasks: one in which we elicit risk preferences defined over objective lotteries, and one in which we elicit subjective beliefs using the QSR defined over monetary payoffs. We want to ensure that the scale of payoffs in each task is comparable, to avoid extrapolation. We want to have each subject undertake both tasks to allow estimation of risk preferences, and recovery of true latent subjective beliefs at the level of the individual.

In all experiments subjects were recruited from the undergraduate population at Georgia

[10] This also means that Propositions 1 through 7 of Harrison, Martinez-Correa, Swarthout and Ulm [2012], that characterize the beliefs recovered for an SEU decision-maker, also generalize.

State University, spanning several colleges. All subjects received a show-up fee of $7, and no specific information about the task or expected earnings. Apart from the belief tasks that are the focus here, all subjects initially completed a task consisting of 50 binary lottery choices. They were told that one of those choices would be selected at random for payment.[11] Earnings from the selected lottery choice were recorded prior to the belief elicitation task, and subjects were paid for both tasks. Appendix A contains all instructions and lottery parameters. A total of 71 subjects were recruited in July 2012.

The subjective belief questions asked of all subjects were as follows:

- **Q1: Interest Compounding**. "Suppose you had $100 in a savings account and the interest rate is 2% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?" The correct answer is $110.40, and responses were elicited between $100 and $118 in intervals of $2.
- **Q2: Real Interest Rate**. "Suppose you had $200 in a saving account. The interest rate on your saving account was 1% per year and inflation was 2% per year. After 1 year, what would be the value of the money on this account?" The correct answer is $198, and responses were elicited between $196 and $204 in intervals of $1.
- **Q3: Expected Lifetime for Men**. "Based on 2006 statistics, if a man lived to be 20 in the United States, how many more years would he expect to live? Note that this is not the age he would die at, but how many more years he would expect to live." The correct answer is 56.1 years, and responses were elicited in decades (0 to 9 years, 10 to 19 years, ... 90 to 100 years).
- **Q4: Expected Lifetime for Women**. "Based on 2006 statistics, if a woman lived to be 20 in the United States, how many more years would she expect to live? Note that this is not the age she would die at, but how many more years she would expect to live." The correct answer is 61.0 years, and responses were elicited in decades.
- **Q5: Overall Inflation Rate in Atlanta**. "What was the overall inflation rate in Atlanta between February 2012 and February 2013?" The correct answer is 2.1%, and responses were elicited in roughly single percentage points for positive values: negative, between 0.1% to 1%, between 1.1% to 2%, ..., between 7.1% to 8%, and over 8%.
- **Q6: Inflation Rate for Food and Beverages in Atlanta**. "What was the inflation rate for Food and Beverages in Atlanta between February 2012 and February 2013?" The correct answer is 1.9%, and responses were elicited in the same intervals as Q5.
- **Q7: Inflation Rate for Housing Costs in Atlanta**. "What was the inflation rate for Housing Costs in Atlanta between February 2012 and February 2013?" The correct answer is 0.1%, and responses were elicited in the same intervals as Q5.
- **Q8: Inflation Rate for Transportation in Atlanta**. "What was the inflation rate for

[11] Building on a long literature in experimental economics, Harrison and Swarthout [2014] and Cox, Sadiraj and Schmidt [2015] raise new questions about the general validity of the random lottery incentive method when one does not assume SEU. We ignore those concerns when we evaluate alternatives to SEU later.

Transportation in Atlanta between February 2012 and February 2013?" The correct answer is 2.6%, and responses were elicited in the same intervals as Q5.

- **Q9: Death from Heart Disease**. "What fraction of people died from diseases of the heart in the United States in 2007?" The correct answer is 25.4%, and responses were elicited in deciles (0% to 9%, ..., 90% to 100%).
- **Q10: Death from Cancer**. "What fraction of people died from neoplasms (cancers) in the United States in 2007?" The correct answer is 23.2%, and responses were elicited in deciles (0% to 9%, ..., 90% to 100%).
- **Q11: Cancer Deaths to Men from Smoking**. "In the United States, what fraction of deaths due to neoplasms (cancers) in 1995-1999 are attributed to smoking by men?" The correct answer is 71.8%, and responses were elicited in deciles.
- **Q12: Cancer Deaths to Women from Smoking**. "In the United States, what fraction of deaths due to neoplasms (cancers) in 1995-1999 are attributed to smoking by women?" The correct answer is 52.5%, and responses were elicited in deciles.
- **Q13: Heart Disease Deaths from Smoking**. "In the United States, what fraction of deaths due to heart diseases in 1995-1999 are attributed to smoking?" The correct answer is 15.9%, and responses were elicited in deciles.
- **Q14: Deaths from Vehicle Crashes due to Alcohol**. "What fraction of fatal vehicle crashes in 2009 were associated with alcohol-impaired drivers (with blood-alcohol levels of .08% and higher)?" The correct answer is 22.3%, and responses were elicited in deciles.
- **Q15: Deaths from Vehicle Crashes due to Alcohol if Aged Between 21 and 24**."What fraction of fatal vehicle crashes in 2009 were associated with alcohol-impaired drivers aged between 21 and 24 (with blood-alcohol levels of .08% and higher)?" The correct answer is 34.5%, and responses were elicited in deciles.

The order of presentation of questions was held constant for each subject, since several of the questions related to each other, and this ensures maximal control for possible order effects across treatments.

The first two questions are natural extensions of questions asked by Lusardi and Mitchell [2007] in the *Health & Retirement Survey* (HRS) of 2004 in the United States. This survey is naturally representative of Americans over the age of 50. Our Q1 adapts the following question of theirs: "Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, less than $102?" The main difference is that we ask for beliefs about the true answer over a wide range. Our Q2 adapts this question of theirs: "Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?" Lusardi and Mitchell [2012; Table 2.1] report that only 67.1% and 75.2% of their sample gave the correct response to each question, respectively. These fractions drop significantly (their Figures 2.1a and 2.1b) as one considers Black and Hispanic respondents. When the same questions were posed to a nationally representative sample of young Americans, aged between 22 and 28 in Wave 11 of the *National Longitudinal Survey of Youth* conducted in 2007-2008, 79.3% and 54.0% gave the correct responses to the interest rate and inflation questions, respectively (Lusardi, Mitchell and Curto [2010; Table 1, p. 365]). We do not take a position on whether these two questions assess information, in the sense of subject knowledge of a fact, computational skills, or even basic literacy about the language used in the questions.

The next two questions ask about a basic informational input to retirement planning: expected remaining lifetime, conditional on reaching the age of 20.[12] Smith, Taylor and Sloan [2001; p. 1126] call this "the most important subjective risk assessment a person can make," although they were referring to own-mortality. We separate out the question for men and women, to ascertain if the differential expected mortality between the two is recognized by individuals. These questions do not condition on the health, income, or any other relevant characteristics of the individual that would affect expected mortality. One could easily extend these questions to elicit more precise beliefs about someone more closely like the subject.

The most widely used evidence on subjective beliefs about longevity come from the *Health and Retirement Survey*, which has asked a simple question since 1992: "With 0 representing absolutely no chance, and 100 absolute certainty, what is the chance that you will live to be 75 years of age or older?" for respondents under the age of 65. A comparable question asks the chance that they would live to be 85, and for respondents over 65 a variant asked the chances of them living 11-15 years

[12] These data come from Table A of the United States Life Tables for 2006, reported in the *National Vital Statistics Reports* (v.58, #21, June 28, 2010) of the Centers for Disease Control & Prevention (CDC) of the U.S. Department of Health & Human Services.

more. In the 2006 wave of the *Health and Retirement Survey* a sub-sample was asked questions that elicited their beliefs about the population life tables: "Out of a group of [men/women] your age, how many do you think will survive to the age of X?" The value of X was 75 for those under 65 themself, and 11-15 years older for those over 65. These questions are closer to those we asked, although we only conditioned on the single age 20. Of course, these questions were not incentivized, and did not elicit information on the confidence of the subjective belief.[13]

Four questions ask for beliefs about inflation rates, which are a critical input to decision making by policy-makers, and contain many puzzles (e.g., Bryan and Venkatu [2001a][2001b] and Engelberg, Manski and Williams [2009]). Our questions focus on the annual rate of inflation in Atlanta in the year prior to the elicitation, since that experience is likely to be most relevant for our population. It considers the inflation rate for all urban residents, and decomposes the overall rate into the three most significant components: Food and Beverages accounts for 14.3% of the expenditures in Atlanta, Housing for 42.7%, and Transportation for 16.5%.[14] It is quite possible that individuals have a poor sense of the overall inflation rate, but do know more precisely the inflation rate for certain categories.

The final six questions elicit beliefs about basic health risks and their correlates. One is the general risk of heart disease, another is the general risk of cancers, the two leading causes of death in the United States.[15] Then we turn to the role of smoking in deaths from cancers, differentiating men

---

[13] Smith, Taylor and Sloan [2001] show that responses to this question are reasonably good predictors of future, actual mortality, even if they do not perfectly reflect new health information when updated. Perozek [2008] makes an even stronger case for the predictive value of these subjective belief questions, arguing that responses to these questions actually outperform population life tables. In contrast, Elder [2013] stresses that only with the 2006 wave can one evaluate the actual predictions, as early respondents reach the target ages of 75 or 85. And in that respect he presents a sharply contrary view, arguing that the evidence supports a "flatness bias," a "tendency for individuals to understate the likelihood of living to relatively young ages while overstating the likelihood of living to ages beyond 80." He attributes this bias to a failure to recognize that mortality risk increases with age.

[14] The data on inflation rates comes from the Detailed CPI Tables of the Bureau of Labor Statistics (BLS) for February 2013, available at http://www.bls.gov/cpi/cpid1302.pdf.

[15] These data on the leading causes of death come from the Mortality Tables of the Division of Vital Statistics, National Center for Health Statistics, Centers for Disease Control & Prevention (CDC), available at

and women.[16] Finally, we examine the role of excessive drinking on vehicle fatalities, in general and for the age group closest to our subjects, those aged between 21 and 24.[17]

3. Results

A. Risk Preferences

To evaluate RDU preferences for individuals we estimate an RDU model for each individual, following procedures explained in Harrison and Rutström [2008]. The formal econometric model is specified in Appendix B. We consider the CRRA utility function (2) and one of three possible probability weighting functions defined earlier by (6), (7) and (8). For our purposes, it does not matter which of these probability weighting functions characterize behavior: the only issue is at what statistical confidence level we can (or cannot) reject the EUT hypothesis that $\omega(p) = p$.

If the sole metric for deciding if a subject were better characterized by EUT and RDU was the log-likelihood of the estimated model, then there were be virtually no subjects classified as EUT since RDU nests EUT.[18] But if we use metrics of a 10%, 5% or 1% significance level on these test of the EUT hypothesis that $\omega(p) = p$, then we classify 31%, 33% or 42% of the 65 subjects with valid estimates as being EUT-consistent. Figure 3 displays these results using the 10% significance level. The left panel shows a kernel density of the *p*-values estimated for each individual and the EUT hypothesis that $\omega(p) = p$; we use the best-fitting RDU variant for each subject, which is normally the

http://www.cdc.gov/nchs/nvss/mortality_tables.htm. We specifically rely on Table LCWK2 for 2007.

[16] These data are extracted from the 2004 report of the Surgeon-General on the health effects of smoking. Those reports are available at http://www.surgeongeneral.gov/library/reports/. Specifically, we rely on data from Table 7.3 of U.S. Department of Health & Human Services [2004].

[17] These data on fatalities come from the U.S. National Highway Traffic Safety Administration, as reported in the *Statistical Abstract of the United States: 2012* of the U.S. Census Bureau (Table 1113, p. 698).

[18] The qualification "virtually" is added because there may be some subjects for whom none of the RDU models can be estimated, for numerical reasons, but for whom the EUT model can be estimated. We could filter the estimates to avoid a handful of numerical outliers: estimates with s too close to 1, estimates with $\gamma < 0.28$ with the inverse-S probability weighting function, estimates with $\gamma > 5$, and estimates with $\eta > 20$. The net effect would be that we have valid estimates for 65 of 71 subjects that participated in the experiment.

general Prelec function (8). The vertical lines show the 1%, 5% and 10% *p*-values, so that one can see that subjects to the right of these lines would be classified as being EUT-consistent. The right panel shows the specific allocation using the 10% threshold. The majority of subjects are classified as RDU in terms of one of the three variants, and 31% are classified as EUT.

We therefore consider the effect of the subjects that are not classified as EUT using these data, on the assumption that they cannot be reliably classified as SEU for the beliefs elicitation task. Again, the maintained assumption here is that evidence against EUT behavior is a useful metric for evidence against SEU for that individual. We classify subjects in a binary manner using this approach.[19]

B. Subjective Beliefs

Measuring Agreement and Disagreement

Any measuring instrument can be compared against another measuring instrument. Examples include weight scales, political opinion polls, or medical judgements about diagnoses. In our case we are interested in the *reported* and *recovered* subjective beliefs about some fact and seek to measure their consistency. In the biostatistics literature a popular concordance index ρ_c has been developed by Lin [1989]. This index combines the familiar notion of correlation from a Pearson inter-class correlation coefficient with allowance for bias, and is virtually identical to measures of intra-class correlation used in psychology and sociology (Krippendorff [1970], Müller and Büttner [1994], Nickerson [1997]). The index is bounded in [-1, 1], with the usual interpretation that $\rho_c = 1$ indicates perfect concordance, and smaller values indicate poorer concordance. We apply the concordance index at the level of the individual's reported and recovered subjective beliefs for a

[19] It would be possible to use the individual *p*-value as the basis for *weighting* the beliefs data in a more quantitatively nuanced manner: one individual might have a *p*-value of 0.09 and another might have a *p*-value of 0.11, and be treated as completely different types using the binary classification.

specific fact. Thus we can make a statement about the agreement or disagreement for each subject and each specific question.[20]

Recovered Beliefs Are Conditional on the Assumed Model of Risk Preferences

Figure 4 illustrates a central theme that goes back to Savage [1971][1972]: one cannot recover subjective beliefs without making some assumptions about the underlying model of risk preferences. Those assumptions might take the form of designing an elicitation procedure that is assumed to "risk neutralize" the individual (e.g., Köszegi and Rabin [2008] and Karni [2009]), applying a payoff procedure that is assumed to "risk neutralize" the individual (e.g., Smith [1961], Harrison, Martínez-Correa and Swarthout [2014] and Harrison, Martínez-Correa, Swarthout and Ulm [2015]), or just assuming, contrary to the evidence, that individuals are risk neutral. In Figure 4 we take one individual and one set of reported beliefs, and recover four distinct sets of subjective belief distributions for each of four distinct models of risk preferences.

In this case there are positive reports for 4 bins, and none of the models of risk preferences assigns any subjective belief to the bins that have zero reports. This is an unsurprising matter of theory. The first bar of each bin in Figure 4 shows the observed report. The second bar of each bin shows the recovered belief assuming that this individual behaved as if an EUT decision-maker, and further had a CRRA coefficient of 0.706. This coefficient was estimated for this individual from the separate task of 50 lottery choices. Again, as expected from the theoretical results of Harrison, Martínez-Correa, Swarthout and Ulm [2012], we do not see a significant difference between the beliefs recovered under EUT from the reported beliefs.

[20] There is a large literature on the significance of disagreement *across* elicited *point* forecasts of different individuals as a measure of uncertainty *in* the forecast. These are different things, forced together solely because elicited *distributions* have not been available, as explained well by Zarnowitz and Lambros [1987] and Engelberg, Manski and Williams [2009]. And this is quite apart from the within-subject nature of our evaluations of disagreement.

The third, fourth and fifth bars of each bin show the dramatic effect of assuming different RDU models, where the difference derives solely from different assumptions about the probability weighting function. The largest effect is if we assume the individual is an RDU decision maker with a power probability weighting function: the recovered belief for the fifth (sixth) bin is much lower (higher) than the reported belief.

As it happens, this individual is best characterized by the RDU model that assumes the flexible Prelec probability weighting function shown in Figure 5. This happens to be an inverse-S probability weighting function, with a pattern of overweighting low probabilities and underweighting high probabilities that many regard as standard (we disagree with that conventional wisdom, but that is of no importance here). The effect of this probability weighting pattern is to overweight the smaller payoff when only two bins have positive reports, and to overweight extreme payoffs when there are more than two bins with positive reports. The right panel of Figure 5 shows the implied decision weights, using *equiprobable* reference lotteries with 2, 3 or 4 prizes to illustrate the pure effect of probability weighting.[21] The "prizes" in our case are payoffs for each bin that received a positive report, so there could be up to 10 prizes in the implied subjective belief elicitation lottery.

Hence the bottom line for this subject and his recovered subjective beliefs about this fact is to compare the reported belief to the final bar within each bin. We infer that he actually attaches roughly the same subjective belief to bin 5 as the observed report, we infer a much higher subjective belief for bin 6 compared to the observed report, and we infer virtually no subjective belief for bins 7 and 8 compared to the observed report. Without doing the arithmetic, it is apparent that the average *recovered* belief here would have to be lower than the average *reported* belief.

These examples in Figure 4 illustrate the effect of beliefs being conditional on different

[21] In other words, the two-prize reference lottery has true probabilities of ½ and ½, the three-prize reference lottery has true probabilities of ⅓, ⅓ and ⅓, and the four-prize reference lottery has true probabilities of ¼, ¼, ¼ and ¼.

models of risk preferences, but entail one simplification: we assume that there are no standard errors on the estimates of the model of risk preferences. However, the EUT parameter ρ has an estimated standard error of 0.14, and a 95% confidence interval between 0.42 and 0.98. The obvious solution here is to bootstrap the calculation of recovered subjective beliefs, and we do so later. When the risk preferences model has more than one parameter, which is the case for the RDU models, we use the estimated covariance matrix for this bootstrapping exercise. The classification of this individual as an RDU decision maker with a Prelec probability weighting function *does* properly take into account the statistical nature of these estimates, since it is based on a *p*-value of 0.0014 from a test of the null hypothesis that $\omega(p) = p$.

Recovered Beliefs Differ From Reports at the Individual Level

Because of the heterogeneity of risk preferences across individuals, it is perhaps no surprise that the average effect of correcting for RDU risk preferences might "wash out" in the sense that the reported and recovered beliefs for all subjects look similar. If the fraction of EUT subjects in the sample is sufficiently high, we expect this approximate agreement between reported and recovered beliefs as a theoretical matter, and indeed for each EUT individual as well as for the sample as a whole. This is why we use a 10% significance level in our exposition, to generate a small fraction of individuals classified as EUT, as shown in Figure 3.

Figure 6 illustrates this phenomenon, looking at one belief question. The top left panel shows reported and recovered beliefs for all subjects for whom a risk preference model is estimated. In each case we use the preferred risk preference model for the individual, so for some individuals this is EUT, and for others it is one of the RDU models: see the right panel of Figure 3 for the distribution across risk preference models. In the pooled case the reported and recovered distributions look close, even though a minority of subjects are classified as EUT-consistent.

However, the most important part of Figure 6 is the display of reported and recovered beliefs for each of 8 individuals. In each case we show the preferred type of risk preferences (EUT or RDU) and the implied concordance correlation ρ_c between reported and recovered beliefs. Subject #3 is an EUT subject, so theory and our numerics show that recovered beliefs and reported beliefs are very close, resulting in a concordance correlation ρ_c of 0.99. However, subject #5, an RDU subject has a concordance correlation of 0.88, and subject #6 has a concordance correlation of only 0.15. It is perfectly possible for an RDU subject to have a concordance correlation of 1, as subject #4 illustrates: this would occur if all tokens are in one bin, or those bins that have tokens allocated to them have the same number of tokens allocated.[22] In effect, the disparity between reports and recovered beliefs depends on the dispersion of reports, whether the individual is characterized as an EUT or RDU decision maker, and finally the degree of probability weighting conditional on being an RDU decision maker.

Figure 7 shows the distribution of concordance correlations across all subjects for the question illustrated in Figure 6. The summary statistics of this distribution are also shown. It is apparent that the average is less than 1, although the negative skew is pronounced. Since the reported and recovered beliefs must perfectly agree for the bins with zero reports, and individuals generally only report positive beliefs for a subset of the 10 bins, there is bound to be some concordance between reported and recovered beliefs, even under RDU risk preferences. Figure 8 decomposes the aggregate results of Figure 7 in a striking manner. The bottom panel of Figure 8, again, just reflects what theory tells us from Harrison, Martínez-Correa, Swarthout and Ulm [2012], albeit in exact numeric form. The top panel of Figure 8 is the value added here, recovered the true, latent beliefs from RDU individuals, where the recovered beliefs are distinctly not the same as the reported beliefs.

[22] In this case, as inspection of the 3rd and 5th bin of Figure 1 illustrates, the bin receives the same payoff and hence, under RDU, the same rank.

Figure 9 summarizes the distribution of individual concordance coefficients across the belief questions.[23] Since these are the same subjects, armed with the same risk preferences, reporting their beliefs across all questions, the differences result solely from differences in the dispersion of reported beliefs and the interaction with risk preferences. In the case of the Interest Compounding question Q1, we observe a significant mode of subjects allocating 100 tokens to the correct bin, so risk preferences play no role in this case. At the other end of the spectrum, there is considerable dispersion in reported beliefs about the historical inflation rate in Atlanta, so RDU risk preferences play more of a role here.

Imprecision About Risk Preferences

Risk preferences are estimated statistically. This entails two types of imprecision: the determination of the correct *model* of risk preferences, and the estimation of specific *parameters* conditional on the model.

Figure 10 illustrates the importance of determining the correct model of risk preferences. These are the elicited and recovered beliefs of one person about the fraction of deaths due to cancer that the CDC attributes to smoking. This subject is a young, female smoker, so these are important beliefs for her health decisions. If we use a 1% significance level we characterize her as an EUT decision-maker, and her recovered beliefs closely track the reported beliefs. However, if we use a 5% or higher significance level, we characterize her as an RDU decision-maker, and recover latent subjective beliefs that are much closer to the true facts. She underestimates the risk of smoking no matter how we characterize her risk preferences, but the misperception is clearly greater if we view her as an EUT decision maker. So from a *qualitative* perspective we do not need to know whether her risk preferences are EUT or RDU, but to ascertain the size of her misperception we do need to

[23] We omit the three specific inflation questions Q6, Q7 and Q9, which generate virtually the same distribution as the general inflation question Q5.

correctly know those preferences.

As it happens, as an empirical matter we observe relatively little imprecision of this kind. Most of our subjects are either EUT or RDU for all three popular significance levels used to evaluate the EUT null hypothesis that $\omega(p) = p$. Two subjects are classified as EUT at the 1% and 5% significance levels, and then as RDU at the 10% level; six subjects were similarly classified at the 1% and 5% levels, respectively.[24]

The other source of imprecision in risk preferences is the statistical sampling error, conditional on each model. In this case we can bootstrap the recovered beliefs, using the covariance matrix of estimates for each subject and model. Figure 11 shows detailed result for the subject and belief task considered earlier in Figure 4, using 10,000 bootstraps.[25] The average beliefs for the interval spanning 60 to 69 years is 0.29, close to the report of 0.31. However, the average belief for the interval spanning 70 to 79 years is 0.68, much higher than the report of 0.47. These averages match the displays for the point estimates shown in Figure 4.

The striking result from the top two panels of Figure 11 is that the dispersion of inferred beliefs about these two intervals are substantial. The standard deviation in these distributions is 0.10 (60 to 69 years) and 0.12 (70 to 79 years). This dispersion arises solely from the fact that the estimates of risk preferences have standard errors. In Figure 11 we focus only on the descriptively best model of risk preferences for this subject, the RDU model with a Prelec probability weighting function.

The distributions in each of the top panels in Figure 11 are not independent. For any bootstrap draw of parameters, the inferred probabilities over all intervals sums to 1. Hence higher

---

[24] One could also consider a mixture model for each individual, following Harrison and Rutström [2009], with the fraction of choices consistent with EUT and RDU providing some basis for weighting each model.

[25] We pool the beliefs for the intervals spanning 70 to 79 years and 80 to 89 years. Each has a low report of 0.11 from this subject, and each have inferred beliefs that average 0.018.

inferred probabilities for one interval implies lower inferred probabilities for some other interval(s). The bottom right panel of Figure 11 illustrates this negative correlation, focusing on the two intervals in the top panels.

Across all subjects and questions, Figure 12 displays the relative size of standard errors of inferred beliefs as a function of the average inferred beliefs and, critically, whether the descriptively preferred model of risk preferences is EUT or RDU. We find that the inferred dispersion is very low when the EUT model best descriptively characterizes the subject, but is much larger when the RDU model best characterizes the subject. The reason for this difference is not that the RDU model has "worse" estimates than the EUT model in general; indeed, we only use the RDU model for an individual when it better describes the risk preferences for the individual. The cause of the sharp difference in Figure 12 is that, for a given degree of imprecision in point estimates, inferences are simply more sensitive to any probability weighting compared to utility function concavity. This is a corollary of our earlier results on how EUT changes inferred probabilities in a "second order" manner from observed reports, but how RDU changes those inferred probabilities in a "first order" manner.

An implication of Figure 12 is that any improvement in the econometric estimation of the RDU models would improve the ability to recover tighter subjective beliefs for those individuals that are best characterized by RDU. In a logical sense this is self-evident, since the recovered subjective beliefs are conditional on those estimates. Figure 12 provides a sense of the quantitative gains in inferences about subjective beliefs that can come from better specifications (e.g., from semi-parametric models and/or larger samples of choices to estimate from).

4. Limitations and Extensions

A. Stakes Outside the Experiment

A related, maintained assumption is that the individual evaluates the payoffs from the belief elicitation task independently of stakes *outside* the experiment, first noted as the "no stakes" condition by Kadane and Winkler [1988]. The issue of outside stakes includes situations in which the individual's wealth outside the lab is known and perfectly integrated with scoring rule payoffs from the experiment (Karni and Safra [1995]), as well as situations where wealth outside the lab is state-dependant on the state of nature whose beliefs are being elicited (Jaffray and Karni [1999]). One example is a surgeon whose reported beliefs about the outcome of an operation might be affected by the fact that her reputation depends on the underlying event (Karni [1999; p.480). Although some formal procedures exist which mitigate these issues, none are practical or cost-effective, and we simply note this issue and assume that the utility function of our decision maker satisfies the no-stakes condition by being additively separable.

B. Hedging Within the Experiment

One maintained assumption is that the individual evaluates the payoffs from the belief elicitation task independently of other choices *within* the experiment. If beliefs are being elicited about some event that affects the choices the individual might make in some other task, there is an immediate possibility of "hedging" causing a distortion in reported beliefs.[26] An example in experimental economics is where one elicits beliefs about the choices some opponent in a strategic game is about to make, and also asks the subject to choose a strategy that depends, under reasonable strategic equilibrium predictions, on that belief (e.g., Rutström and Wilcox [2009]). The evidence

[26] We paid a given subject for their risk aversion choices and for their belief elicitation choices. This combination does not raise hedging issues as defined here, but does raise issues about whether the subject viewed these two sets of choices as one portfolio choice.

from Blanco, Engelman, Koch and Normann [2010] and Armentier and Treich [2013] suggests that this hedging problem is only an issue if the incentives are "transparent and strong."[27] One might view the questions we elicited beliefs over longevity risk of men and women, or cancer mortality risk from smoking of men and women, as likely to be correlated. One solution, tested by Blanco, Engelman, Koch and Normann [2010] and used in our experiments, is simply to apply the random lottery incentive mechanism over the tasks involving hedging opportunities. They find that it does effectively mitigate hedging.

C. Other Models of Risk Preferences or Uncertainty Aversion

Our approach to recovering subjective belief distributions from reported beliefs can be readily extended beyond EUT or RDU. The one constraint, and it is an important one, is to determine the parameters of the appropriate models for an individual independently of the belief elicitation exercise.

In terms of alternative models of risk aversion, alternatives such as Cumulative Prospect Theory (Tversky and Kahneman [1992]) or Disappointment Aversion (Gul [1991]) could be applied. A different type of extension would be to consider "uncertainty aversion," as defined by Schmeidler [1989; p.582] and often referred to as "ambiguity aversion." For instance, the "smooth ambiguity model" of Klibanoff, Marinacci and Mukerji [2005] would be relatively straightforward, as would the α-maxmin EU model of Ghirardoto, Maccheroni and Marinacci [2004], generalizing the maxmim EU model of Gilboa and Schmeidler [1989].[28]

---

[27] For instance, Armentier and Trech [2013; p.24] elicit beliefs over some binary event using a QSR, and pair that with a bet that the subject can make with house money on whether *exactly* the same event occurs.

[28] The α-maximin EU and maxmin EU models are properly considered models of ambiguity aversion, since they do not assume anything about the shape of the subjective belief distribution other than at the extremities. This information might be known to the decision-maker, in which case they are models of uncertainty aversion, but they need not be known, in which case they are properly models of ambiguity aversion.

Our implementation also assumes that individuals are "probabilistically sophisticated" in the sense of Machina and Schmeidler [1992][1995]. Again, if one had some structural model that relaxes this assumption, and operational tests of that alternative, we could extend our approach to relax that assumption. We see this extensibility as an attractive feature of the structural approach, even if one does not implement every possible extension in every application.

D. Source Independence

A critical, identifying assumption of our approach is that the "type" of decision maker can be reliably identified from decisions defined over *objective* lotteries and applied to infer latent *subjective* beliefs. The assumption that the "type" for objective probabilities is the same type for subjective probabilities, and further that the parameters of the utility function and probability function are the same, is a strong form of "source independence." The notion of "source" here was developed by Tversky and Fox [1995] and refers to different sources of imprecision, spaning objective sources such a die rolls as well as subjective sources such as personal knowledge of facts.

The difficulty of dispensing with this assumption is apparent. One would then need to simultaneously recover latent subjective beliefs as well as the utility function and/or probability weighting function. If we just focus on probability weighting function, and the characterization of convex or concave functions as "optimistic" or "pessimistic," the identification problem becomes obvious.

Of course, the importance of the assumption does not make it true. Some have argued that it is false. Abdellaoui, Baillon, Placido and Wakker [2011] conclude that different probability weighting functions are used when subjects face risky processes with known probabilities and uncertain processes with subjective processes. They correctly refer to this "source dependence," where the notion of a source is relatively easy to identify in the context of an artefactual laboratory experiment, and hence provides the tightest test of this proposition. Unfortunately, their conclusions are an artefact of estimation procedures that do not worry about sampling errors.[29]

However, even if the claimed evidence for source dependence is missing, this does not mean that the behavioral phenomenon is missing. Indeed, it is intuitively plausible once one moves to the domain of subjective probabilities, or where objective probabilities are presumed to arise from some inferential process.[30] But we should not mistake our intuition for the evidence.

E. Related Literature

Our approach directly extends the structural approach to recovering latent subject *probabilities* for *binary* events developed by Andersen, Fountain, Harrison and Rutström [2014]. Our focus is on the complete belief distribution for continuous events. Of course, when we consider the discretized version in which one has K intervals of that continuous event space, it would be possible to undertake K-1 elicitations of probabilities for binary events and compile these into an elicited belief distribution.[31] Although formally feasible, this approach would quickly become cumbersome for the subject, particularly if there are several events that one is interested in eliciting belief distributions for.

One other study attempts to recover elicited probabilities from observed choices over binary events, calibrating for non-linear utility functions and/or probability weighting: Offerman, Sonnemans, van de Kuilen and Wakker [2009]. Like us, they consider the recovery of true subjective beliefs when the agent may be risk averse in the narrow sense of EUT, as well as the broader sense

[29] It can be shown that their experiments provide no statistically significant evidence for source dependence when one applies valid statistical procedures.

[30] For example, by the application of Bayes Rule or the reduction of compound lotteries.

[31] For good behavioral reasons one might want to elicit K subjective probabilities and check if the inferred subjective probabilities sum to 1. It is then possible to infer a normalized distribution, noting the possible need for that extra normalization step. Offerman, Sonnemans, van de Kuilen and Wakker [2009] elicit a subjective probability for some outcome, such as a stock price, falling in disjoint intervals S and T, and then elicit the subjective probability of the union outcome in which the stock price is in S or T. They then check for additivity bias in responses.

implied by an allowance for probability weighting. Their preferred approach has a reduced form simplicity, and is agnostic about which structural model of decision making under risk one uses. Our approach is explicitly structural, and generates inferences about subjective beliefs that are conditional on the assumed model of decision making under risk. We see these as complementary approaches, and both have strengths and weaknesses.

One method Offerman et al. [2009] consider is by estimating or eliciting the functional forms of a model of choice under risk (e.g., EUT, RDU or CPT), then observing beliefs over a natural event in some task, and econometrically recovering the implied subjective probability by using the estimated model of choice under risk to "back out" the subjective probability that must have been used in the belief elicitation task. They dismiss this approach, which is the one we follow (for subjective belief distributions). They claim, without further discussion, that estimating or eliciting the functional forms is "laborious" and that it involves "complex multi-parameter estimations." It is certainly true that the joint likelihood involves several parameters, but such estimation is standard fare with maximum likelihood modeling, so that is hardly a concern. It is not clear in what sense this is a "complex" undertaking. The labor involved depends on how one undertakes the estimation or elicitation. In our case the subjects need to do one task, which consists of 50 binary choices over lotteries, and then all of the labor involved is by the computer estimating maximum likelihood models that have been well-studied for years (e.g., Harrison and Rutström [2008; §2] for a survey).[32]

The empirical method they use instead has an attractive reduced form simplicity. For a given subject, it uses the QSR to elicit reported probabilities for naturally occurring events, and then uses the QSR in a calibration task to elicit a "risk correction function" that allows them to recover the

[32] On the other hand, if one uses other elicitation procedures, such as the Trade-Off design of Wakker and Deneffe [1996], Fennema and van Assen [1998], Abdellaoui [2000] and Abdellaoui, Bleichrodt and Paraschiv [2007], then the procedures can indeed become laborious *for the subject*. There are other reasons not to use these methods, the most significant of which is their lack of incentive compatibility as conventionally applied (Harrison and Rutström [2008; §1.5]). But these methods are not needed, and the stated concerns with this approach to recovering subjective beliefs are not substantial.

subjective probability that generated the report for the *naturally occurring event*. The risk correction function simply elicits reports for "objective probabilities," such as the chance that a single roll of a 100-sided die will come up between 1 and 25. Assume the subject reports 0.30 for this event. Then, if the subject ever reported a 0.30 in the initial task for the naturally occurring event, they would infer that he had a subjective probability of 0.25 underlying it, since that was the objective probability that generated this report using the (same) scoring rule. Thus the difference between the report of 0.30 in the calibration task and the true underlying probability is attributed solely to the effects of non-linear utility and/or probability weighting. By eliciting a risk correction function for a wide range of probabilities, and with a sufficiently fine grid, one can recover any report with some reasonable accuracy.

This approach is attractive because it avoids the need for the researcher to "take a stand" on which model of choice under uncertainty determines betting behavior. To see the key assumption underlying their approach, let φ be the *actuarial* probability that the calibration event will occur. For some artefactual events, such as tossing coins and rolling die, φ is well defined, but for other events it is not so well defined. Let $\pi(\varphi)$ be the function that summarizes the subjective belief that the subject actually holds that the calibration event will occur, and let $R(\pi(\varphi))$ be the function transforming $\pi(\varphi)$ into a report using the QSR, or any appropriate scoring rule. Offerman et al. [2009] first assume that $\pi(\varphi) = \varphi$ *in the calibration task*, so that the only reason that $R(\pi(\varphi)) \neq \varphi$ is that the subject has non-linear utility and/or undertakes probability weighting.[33] Why might $\pi(\varphi) \neq \varphi$, for such simple tasks?

[33] So there is no allowance for subjects to make decision errors in the calibration task, or the elicitation task for the naturally occurring event for that matter. These errors could be subsumed into some sampling error on estimates of $R(\pi(\varphi))$ as an empirical function of φ, but then one is relying on the errors being well-behaved statistically. In fact, Offerman et al. [2009; equation (19), p. 1475] do allow for an additive error term which they assume to be truncated normal to ensure that reported probabilities lie between 0 and 1. Their pooled estimates indicate that there is a need for some correction for non-linear utility, but that it is not so clear that probability weighting is an issue (the log-likelihood which allows for both effects is virtually identical to the log-likelihood in which no probability weighting is assumed). Although they allow for errors to vary with an incentives treatment applied between-subjects, it would be useful to extend their statistical analysis of the pooled data to allow for correlated errors at the level of the individual subject, rather than implicitly assume homoskedasticity.

Apart from concerns with loaded die, or certain cultures imbuing randomizing devices or colors on chips with some animist intent, we would be concerned with psychological editing processes based on similarity relations. To take a simplistic example, someone might "round down" to the nearest increment of 0.05 or 0.10 and then decide how to report using this subjectively edited probability $\pi(\varphi)$ as the basis for any adjustments due to non-linear utility or probability weighting. Is the actuarial probability φ the one we really want to compare $R(\pi(\varphi))$ to in such a case, or is it $\pi(\varphi)$?

This might seem to be nit-picking when it comes to the rolling of a 100-sided die in the calibration task, and perhaps viewed as part of a latent structural psychological story underlying the notion of probability weighting. But it is surely more significant for naturally occurring events. Here is where the second assumption comes in: that $\pi(\varphi) = \varphi$ *in the belief elicitation task* where φ is defined (or not) over naturally occurring events. Thus, what if we accept that $\pi(\varphi) = \varphi$ is a reasonable assumption for the calibration task with the artefactual event, but cannot be so sure for the task with the naturally occurring event? Our position is that we are recovering $\pi(\varphi)$, "warts and all" in terms of how the subject conceives of the event and defines the probability φ. Offerman et al. [2009] would appear to be recovering the "touched up" image of $\pi(\varphi)$, φ, after the warts have been removed.

5. Conclusions

We demonstrate how to recover latent subjective beliefs if an individual is known to distort probabilities into decision weights using Rank Dependent Utility theory. Our specific results were for the popular Quadratic Scoiring Rule, but are proven to generalize to the class of proper scoring rules. We show that the effect on recovered beliefs from probability distortions is significant, with large changes in the location and shape of subjective belief distributions. These effects stand in stark contrast to the minimal effects of risk preferences under Subjective Expected Utility Theory. Our results allow the recovery of subjective belief distributions for a much wider class of risk preferences, enhancing the practicality of inferring subjective belief distributions.

Figure 1: Belief Elicitation Interface

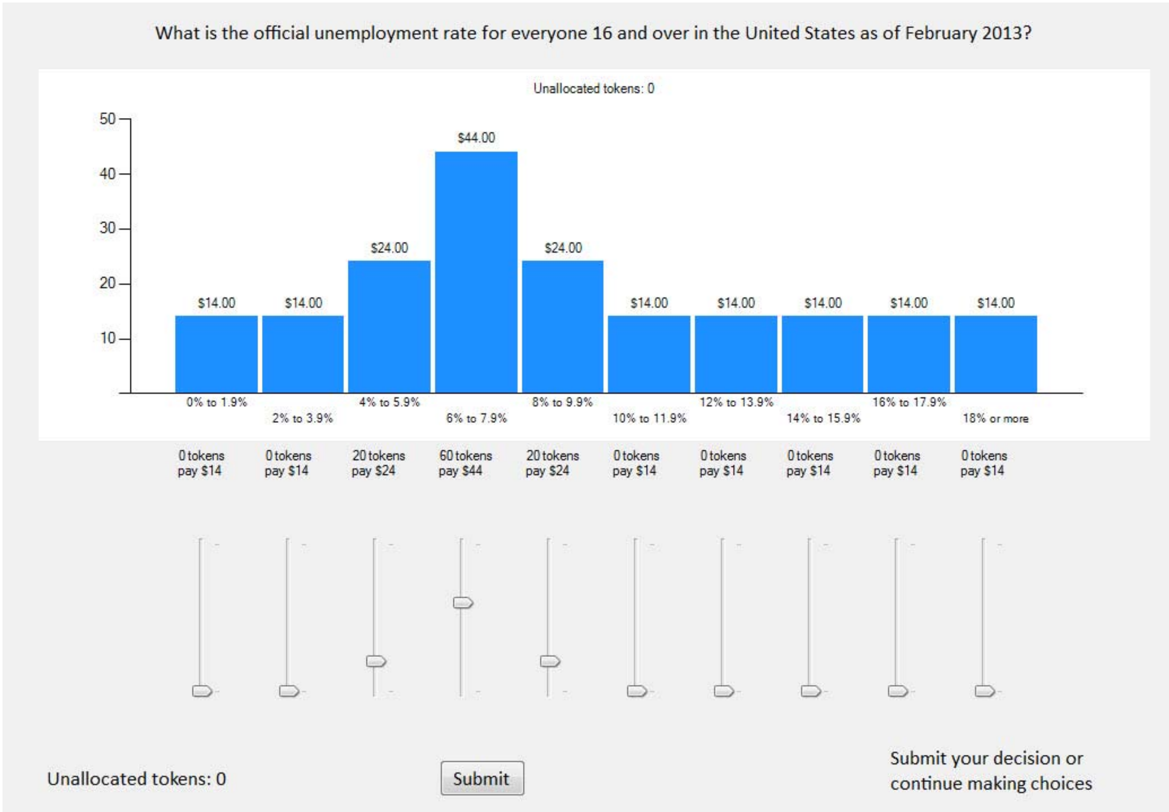


Figure 2: Possible Belief Elicitation Response

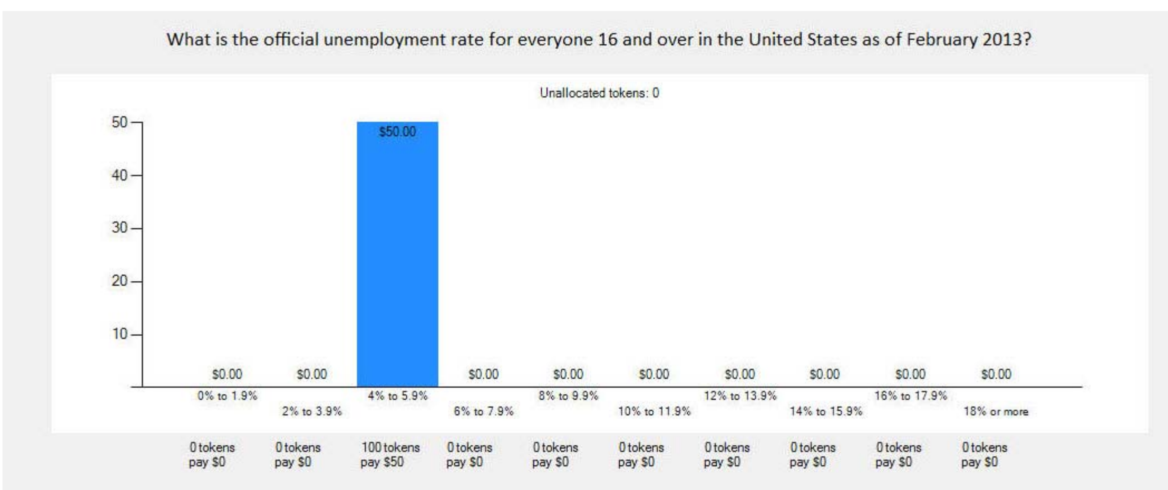


Figure 3: Classifying Subjects as EUT or RDU

N=65, one *p*-value per individual
Estimates for each individual of EUT and RDU specifications

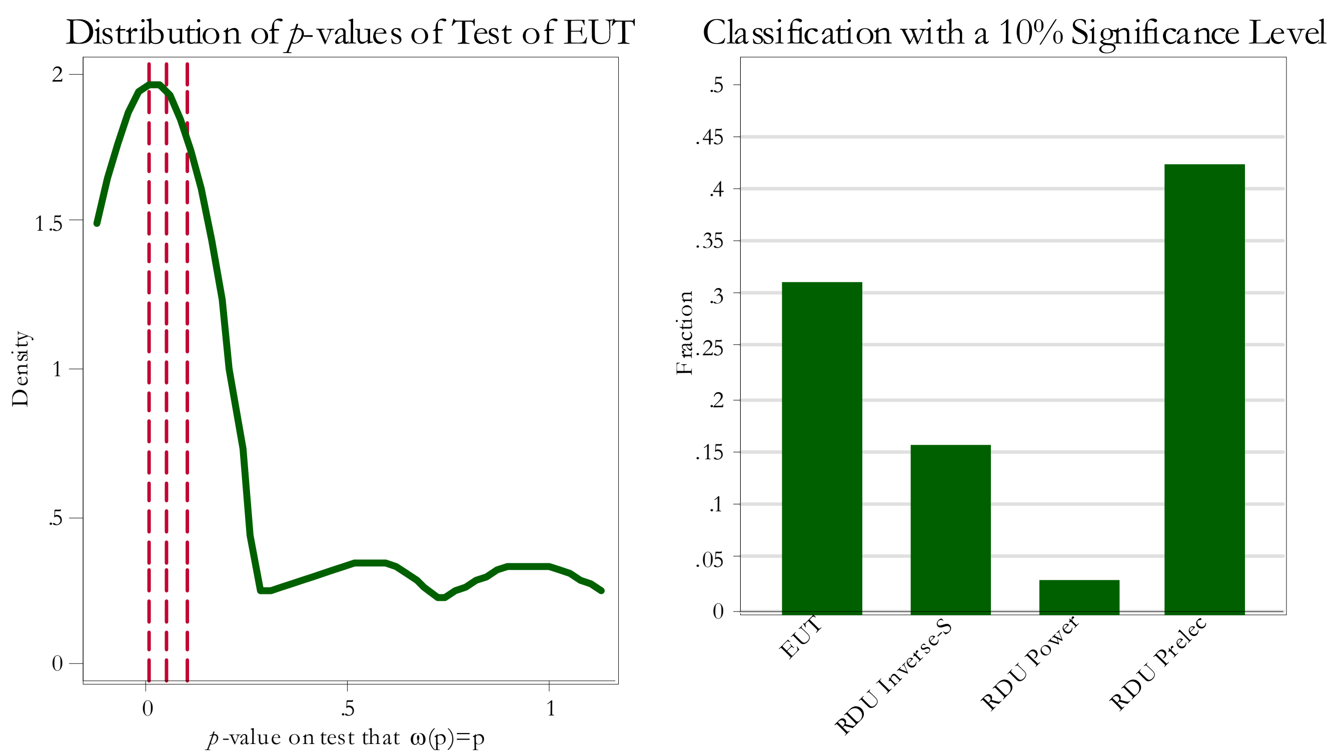


Figure 4: Recovered Beliefs for One Subject

Reported and recovered beliefs for subject 1 and *Longevity for Men* question

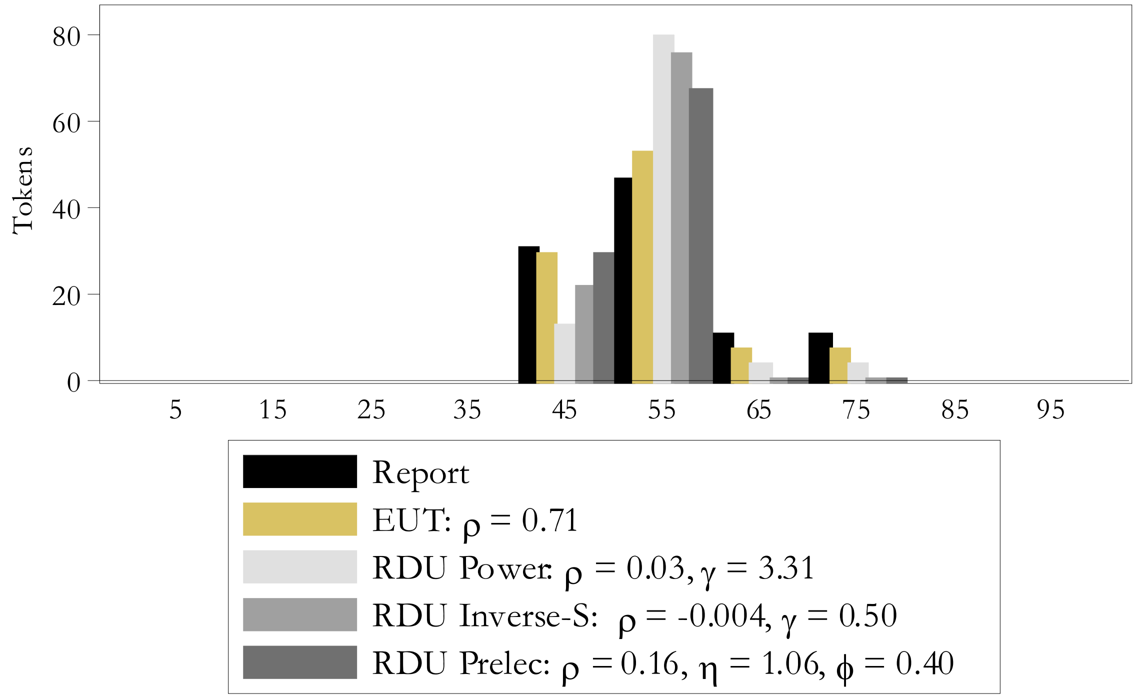


Figure 5: Prelec Probability Weighting and Implied Decision Weights

Based on equi-probable reference lotteries

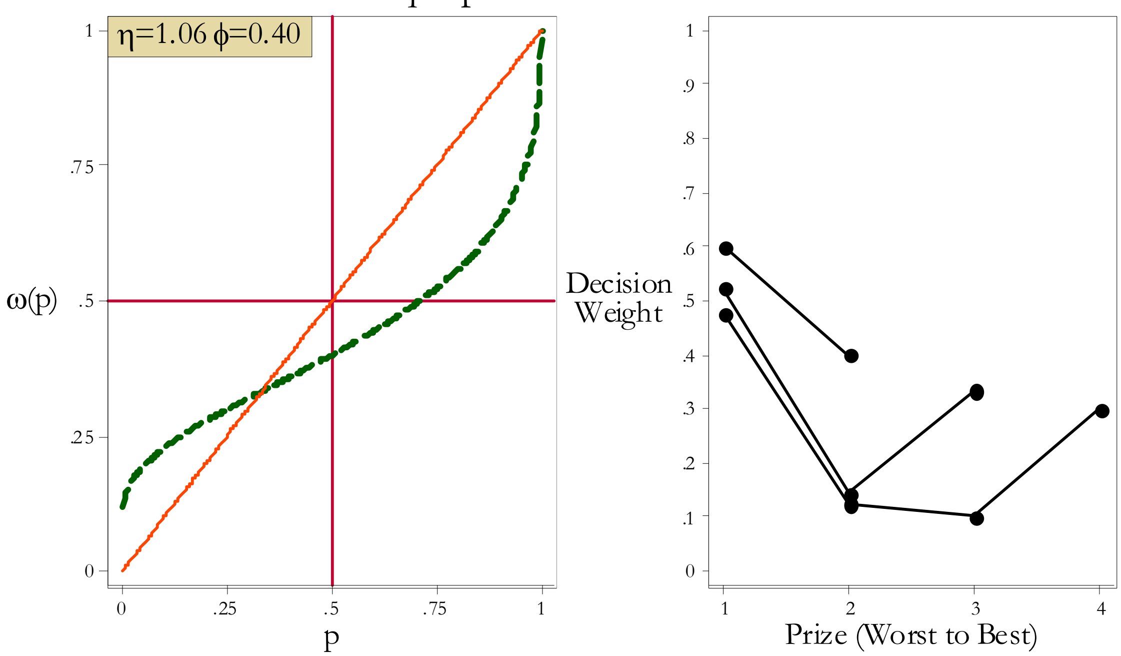


Figure 6: Reported and Recovered Beliefs about Longevity for Men

True number of remaining years was 56.1 according to the *CDC*

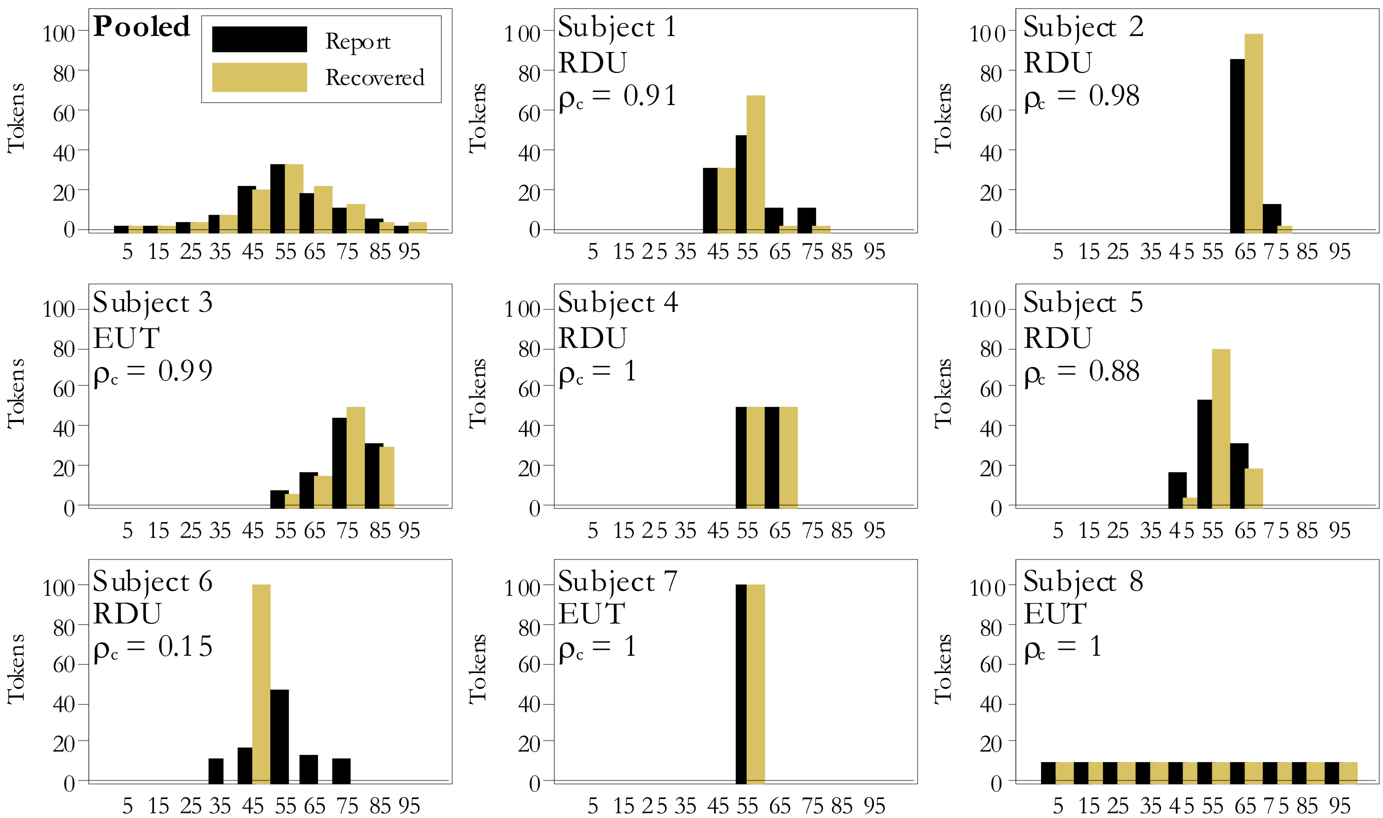


Figure 7: Distribution of Individual Concordance Coefficients for Reported and Recovered Beliefs about Longevity for Men

Concordances for 64 subjects: *M* = 0.82 *SD* = 0.24 *Skew* = -1.62 *Kurtosis* = 4.90

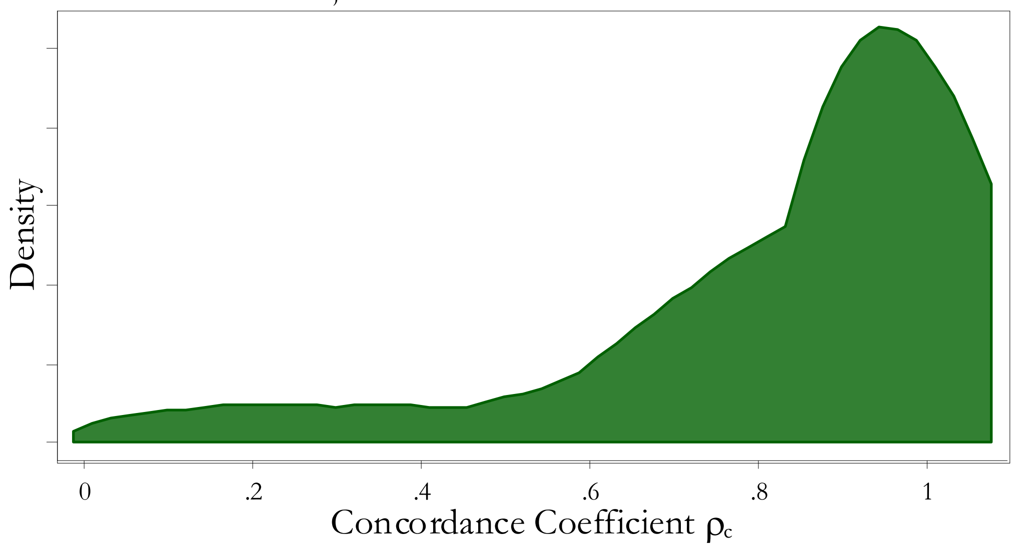


Figure 8: Distribution of Individual Concordance Coefficients for Reported and Recovered Beliefs about Longevity for Men

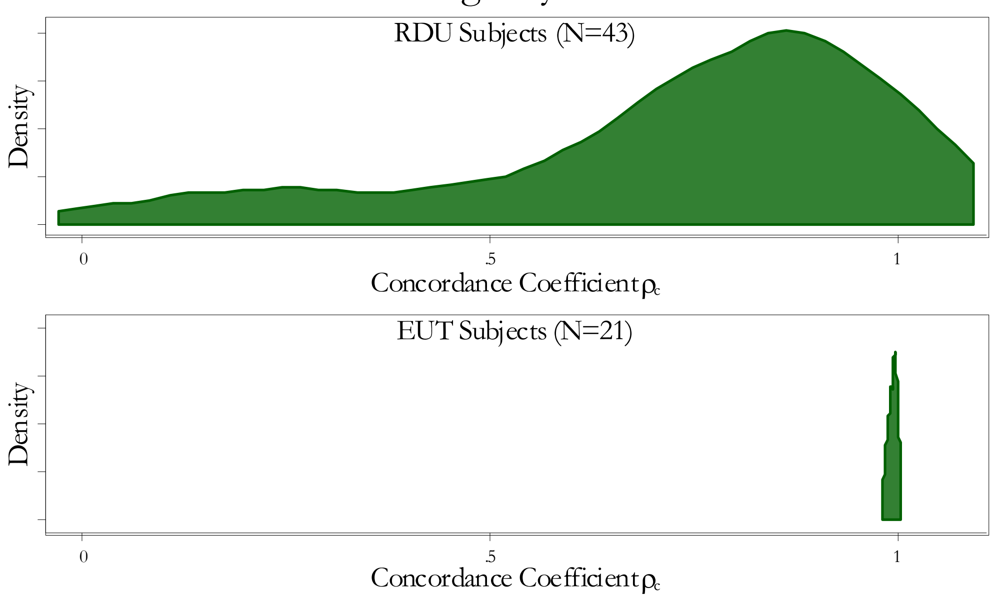


Figure 9: Distributions of Individual Concordance Coefficients for Reported and Recovered Beliefs

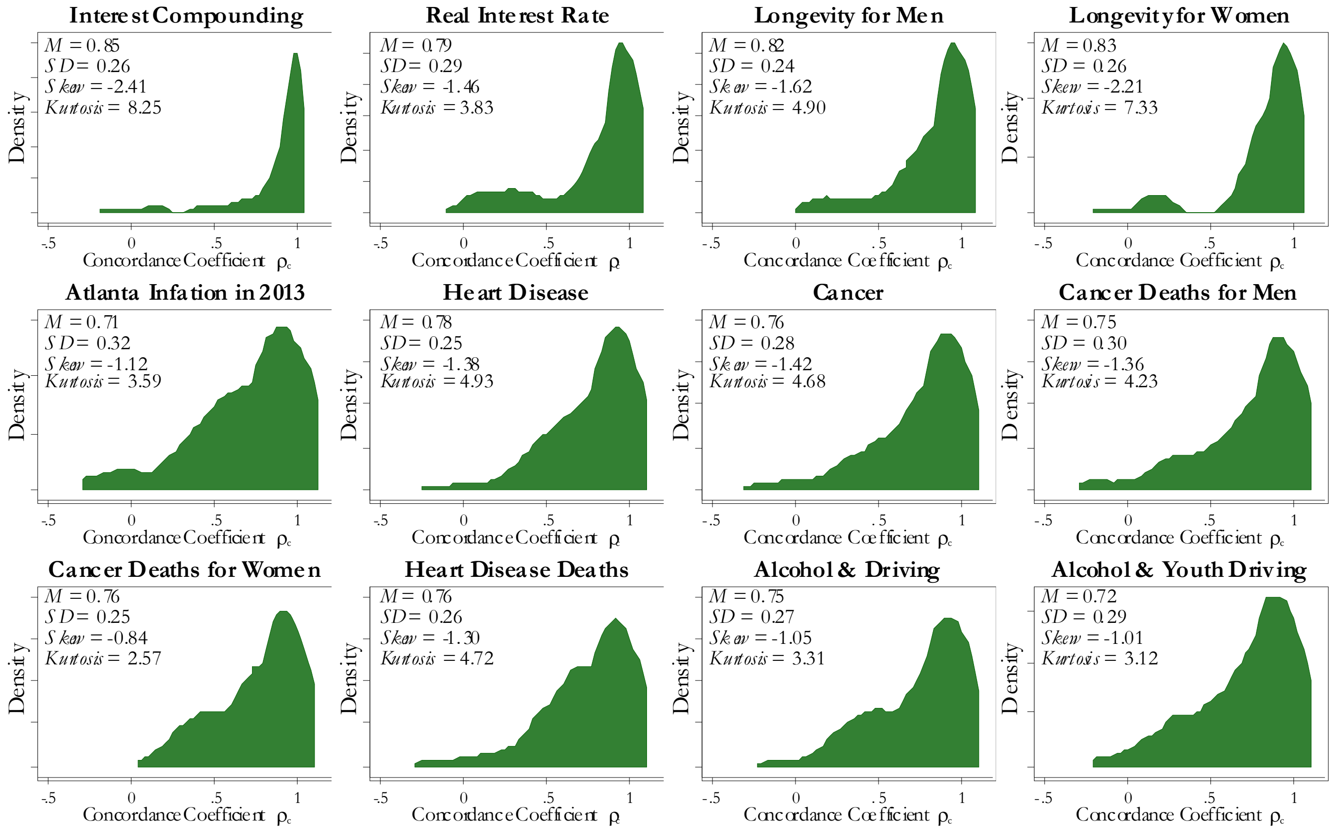


Figure 10: Importance of Modeling Risk Preferences Correctly for a Young, Female Smoker

Reported and recovered beliefs for subject 56

Fraction of deaths due to cancers in 1995-1999 attributed to smoking by women?

True fraction was 52.5% according to vital statistics compiled by the *CDC*

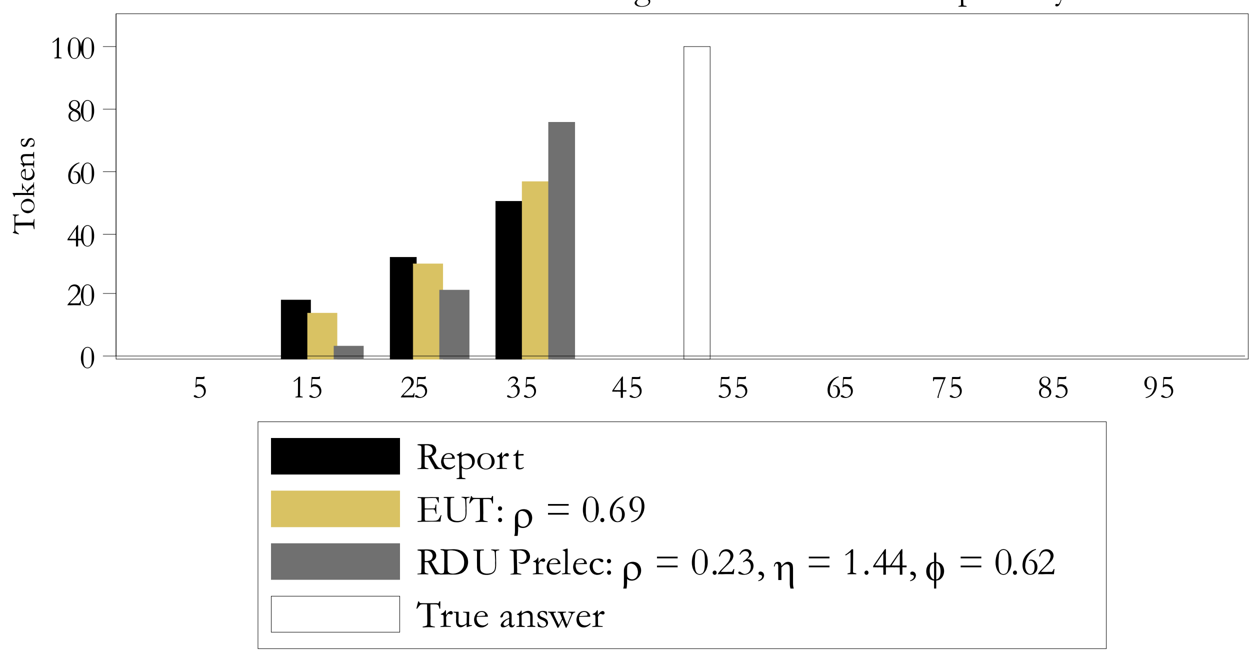


Figure 11: Distribution of Inferred Subjective Beliefs

10,000 bootstrap simulations from estimated risk preferences
Red dashed line is observed report

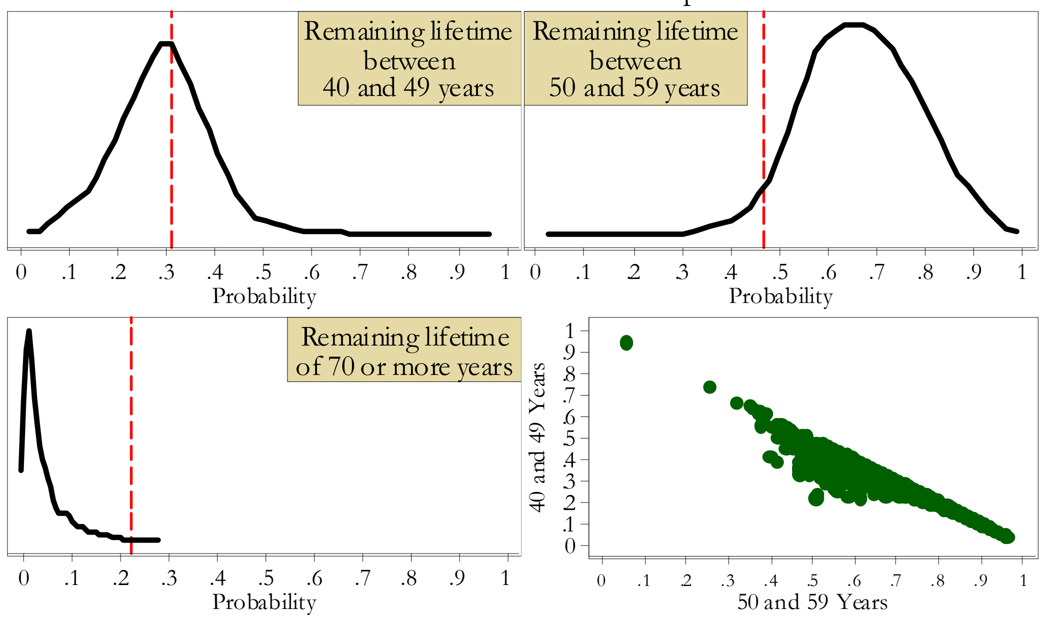
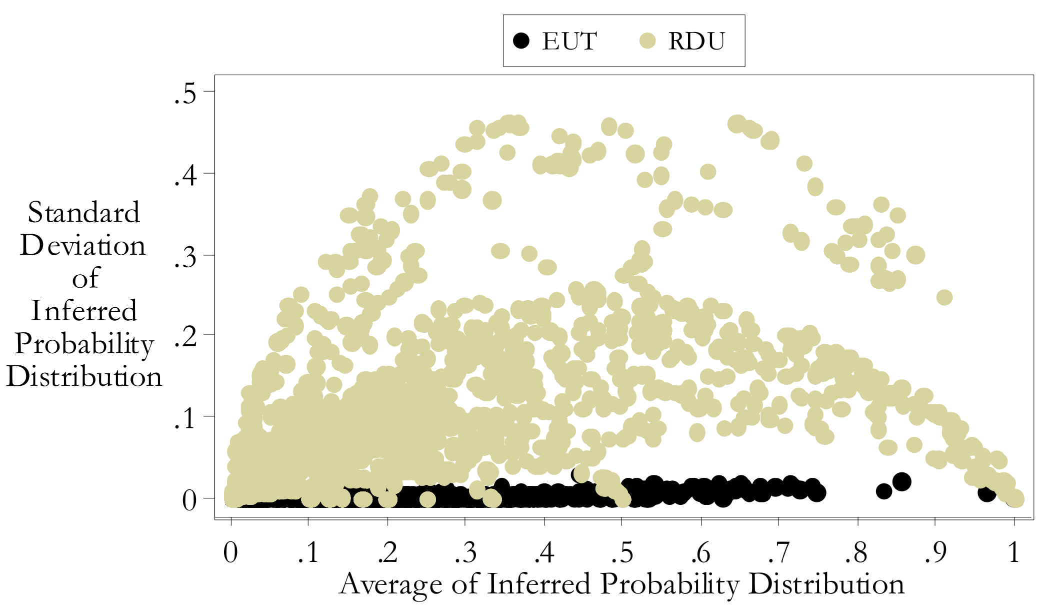


Figure 12: Dispersion of Inferred Probabilities

Only inferences based on winning model of risk preferences
Inferred distributions based on 10,000 bootstrap simulations



References

Abdellaoui, Mohammed, "Parameter-Free Elicitation of Utilities and Probability Weighting Functions," *Management Science*, 46, 2000, 1497-1512.

Abdellaoui, Mohammed; Baillon, Aurélien; Placido, Lætitia and Wakker, Peter P., "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review*, 101, April 2011, 695-723.

Abdellaoui, Mohammed; Bleichrodt, Han, and Paraschiv, Corina, "Measuring Loss Aversion under Prospect Theory: A Parameter-Free Approach," *Management Science*, 53(10), October 2007, 1659-1674.

Andersen, Steffen; Fountain, John; Harrison, Glenn W., and Rutström, E. Elisabet, "Estimating Subjective Probabilities," *Journal of Risk & Uncertainty*, 48, 2014, 207-229.

Armentier, Olivier, and Treich, Nicolas, "Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging," *European Economic Review*, 62, 2013, 17-40.

Blanco, Mariana; Engelman, Dirk; Koch, Alexander K., and Normann, Hans-Theo, "Belief Elicitation in Experiments: Is There a Hedging Problem?" *Experimental Economics*, 13, 2010, 412-438.

Bryan, Michael F., and Venkatu, Guhan, "The Demographics of Inflation Opinion Surveys," *Commentary*, Federal Reserve Bank of Cleveland, October 2001a.

Bryan, Michael F., and Venkatu, Guhan, "The Curiously Different Inflation Perspectives of Men and Women," *Commentary*, Federal Reserve Bank of Cleveland, November 2001b.

Cox, James C.; Sadiraj, Vjollca, and Schmidt, Ulrich, "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics*, 18, 2015, 215-250.

Elder, Todd E., "The Predictive Validity of Subjective Mortality Expectations: Evidence from the Health and Retirement Study," *Demography*, 50(2), 2013, 569-589.

Engelberg, Joseph; Manski, Charles F., and Williams, Jared, "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters," *Journal of Business & Economic Statistics*, 27(1), January 2009, 30-41.

Fennema, Hein, and van Assen, Marcel A.L.M., "Measuring the Utility of Losses by Means of the Tradeoff Method," *Journal of Risk and Uncertainty*, 17, 1998, 277-295.

Ghirardoto, Paolo; Maccheroni, Fabio, and Marinacci, Massimo, "Differentiating Ambiguity and Ambiguity Attitude," *Journal of Economic Theory*, 118, 2004, 133-173.

Gilboa, Itzhak and Schmeidler, David, "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 1989, 141-153.

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

140.

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

Harrison, Glenn W.; Martínez-Correa, Jimmy; Swarthout, J. Todd, and Ulm, Eric R., "Eliciting Subjective Probability Distributions with Binary Lotteries," *Economics Letters*, 127, 2015, 68-71.

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 Rutström, E. Elisabet, "Expected Utility *And* Prospect Theory: One Wedding and a Decent Funeral," *Experimental Economics*, 12(2), June 2009, 133-158.

Harrison, Glenn W., and Swarthout, J. Todd, "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision*, 77, 2014, 423-438.

Jaffray, Jean-Yves, and Karni, Edi, "Elicitation of Subjective Probabilities when Initial Endowment is Unobservable," *Journal of Risk & Uncertainty*, 8, 1999, 5-20.

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, "Elicitation of Subjective Probabilities when Preferences are State-Dependent," *International Economic Review*, 40, 1999, 479-486.

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.

Klibanoff, Peter; Marinacci, Massimo, and Mukerji, Sujoy, "A Smooth Model of Decision Making Under Ambiguity," *Econometrica* 73(6), November 2005, 1849-1892.

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).

Krippendorff, Klaus, "Bivariate Agreement Coefficients for Reliability of Data," *Sociological Methodology*, 2, 1970, 139-150.

Lin, Lawrence I-Kuei, "A Concordance Correlation Coefficient to Evaluate Reproducibility," *Biometrics*, 45, 1989, 255–268.

Lusardi, Annamaria, and Mitchell, Olivia S., "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth," *Journal of Monetary Economics*, 54, 2007, 205-224.

Lusardi, Annamaria, and Mitchell, Olivia S., *Financial Literacy: Implications for Retirement Security and the Financial Marketplace* (New York: Oxford University Press, 2012).

Lusardi, Annamaria; Mitchell, Olivia S., and Curto, Vilsa, "Financial Literacy Among the Young," *Journal of Consumer Affairs*, 44(2), Summer 2010, 358-380.

Machina, Mark J., and Schmeidler, David, "A More Robust Definition of Subjective Probability," *Econometrica*, 60(4), July 1992, 745-780.

Machina, Mark J., and Schmeidler, David, "Bayes without Bernoulli: Simple Conditions for Probabilistically Sophisticated Choice," *Journal of Economic Theory*, 67(106), 1995, 106-128.

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

Müller, Reinhold, and Büttner, Petra, "A Critical Discussion of Intraclass Correlation Coefficients," *Statistics in Medicine*, 13, 1994, 2465-2476.

Nickerson, Carol A. E., "A Note on 'A Concordance Correlation Coefficient to Evaluate Reproducibility'," *Biometrics*, 53, December 1997, 1503-1507.

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, 2009, 1461-1489.

Perozek, Maria, "Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something We Don't Know?" *Demography*, 45(1), February 2008, 95-113.

Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66, 1998, 497-527.

Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.

Rieger, Marc Oliver, and Wang, Mei, "Cumulative Prospect Theory and the St. Petersburg Paradox," *Economic Theory*, 28, 2006, 665-679.

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

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

Schmeidler, David, "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57(3), May 1989, 571-587.

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

Smith, V. Kerry; Taylor, Donald H., and Sloan, Frank A., "Longevity and Expectations: Can People Predict Their Own Demise?" *American Economic Review*, 91, September 2001, 1126-1134.

Tversky, Amos, and Fox, Craig R., "Weighing Risk and Uncertainty," *Psychological Review*, 12(2), 1995, 269-283.

Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representations of Uncertainty," *Journal of Risk & Uncertainty*, 5, 1992, 297-323.

U.S. Department of Health and Human Services, *The Health Consequences of Smoking: A Report of the Surgeon General* (Washington, D.C.: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004).

Wakker, Peter P., and Deneffe, Daniel, "Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown," *Management Science*, 42, 1996, 1131-1150.

Zarnowitz, Victor, and Lambros, Louis A., "Consensus and Uncertainty in Economic Prediction," *Journal of Political Economy*, 95(3), June 1987, 591-621.

Appendix A: Instructions (Online Working Paper)

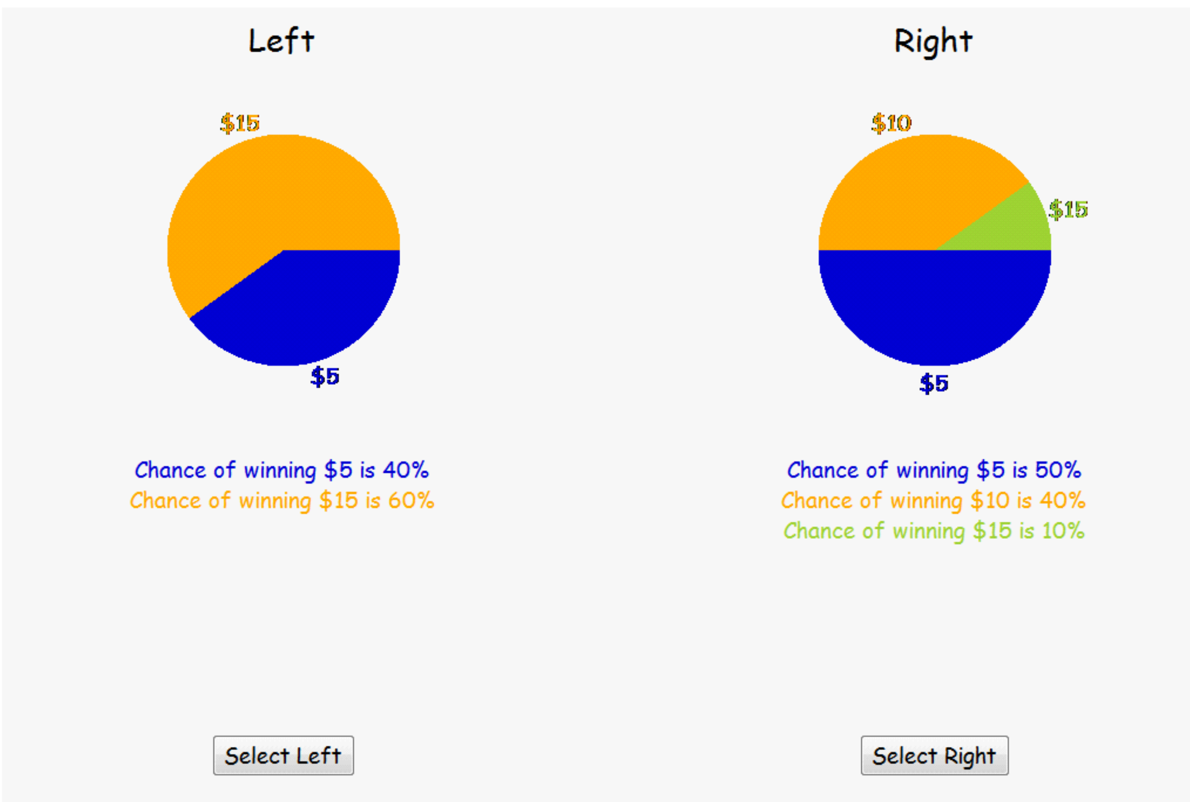
The instructions below for risk preferences assumes 50 lottery choices. The appropriate text was changed for the slight variations in which we had 57 or 60 lottery choices.

A.1. Risk Preferences

Choices Over Risky Prospects

This is a task where you will choose between prospects with varying prizes and chances of winning each prize. You will be presented with a series of pairs of prospects where you will choose one of them. There are 50 pairs in the series. For each pair of prospects, you should choose the prospect you prefer. You will actually get the chance to play one of these prospects, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer.

Here is an example of what the computer display of a pair of prospects will look like.



The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

In the above example the left prospect pays five dollars ($5) if the number drawn is between 1 and 40, and pays fifteen dollars ($15) if the number is between 41 and 100. The blue color in the pie chart corresponds to 40% of the area and illustrates the chances that the number drawn will be between

1 and 40 and your prize will be $5. The orange area in the pie chart corresponds to 60% of the area and illustrates the chances that the number drawn will be between 41 and 100 and your prize will be $15. When you select the lottery to be played out the computer will confirm what die rolls correspond to the different prizes.

Now look at the pie chart on the right. It pays five dollars ($5) if the number drawn is between 1 and 50, ten dollars ($10) if the number is between 51 and 90, and fifteen dollars ($15) if the number is between 91 and 100. As with the prospect on the left, the pie slices represent the percentage of the possible numbers which yield each payoff. For example, the size of the $15 pie slice is 10% of the total pie, and is thus 10 numbers out of 100.

Each pair of prospects is shown on a separate screen on the computer. On each screen, you should indicate which prospect you prefer by clicking on one of the buttons beneath the prospects.

After you have worked through all of the pairs of prospects, raise your hand and an experimenter will come over as soon as they are available. You will then roll two 10-sided dice to determine which pair of prospects will be played out. You roll the die until a number between 1 and 50 comes up. Since there is a chance that any of your 50 choices could be played out for real earnings, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will again roll the two ten-sided dice to determine the outcome of the prospect you chose.

For instance, suppose you picked the prospect on the left in the above example and it was the pair chosen to be played. If the random number from your rolls of the dice was 37, you would win $5; if it was 93, you would win $15. If you picked the prospect on the right and drew the number 37, you would win $5; if it was 93, you would win $15.

Therefore, your payoff is determined by three things:

- **which prospect you selected, the left or the right, for each of these 50 pairs;**
- **which prospect pair is chosen to be played out in the series of 50 pairs using the two 10-sided dice; and**
- **the outcome of that prospect when you roll the two 10-sided dice again.**

Which prospects you prefer is a matter of personal choice. The people next to you may be presented with different prospects, and may have different preferences, so their responses should not matter to you or influence your decisions. Please work silently, and make your choices by thinking carefully about each prospect.

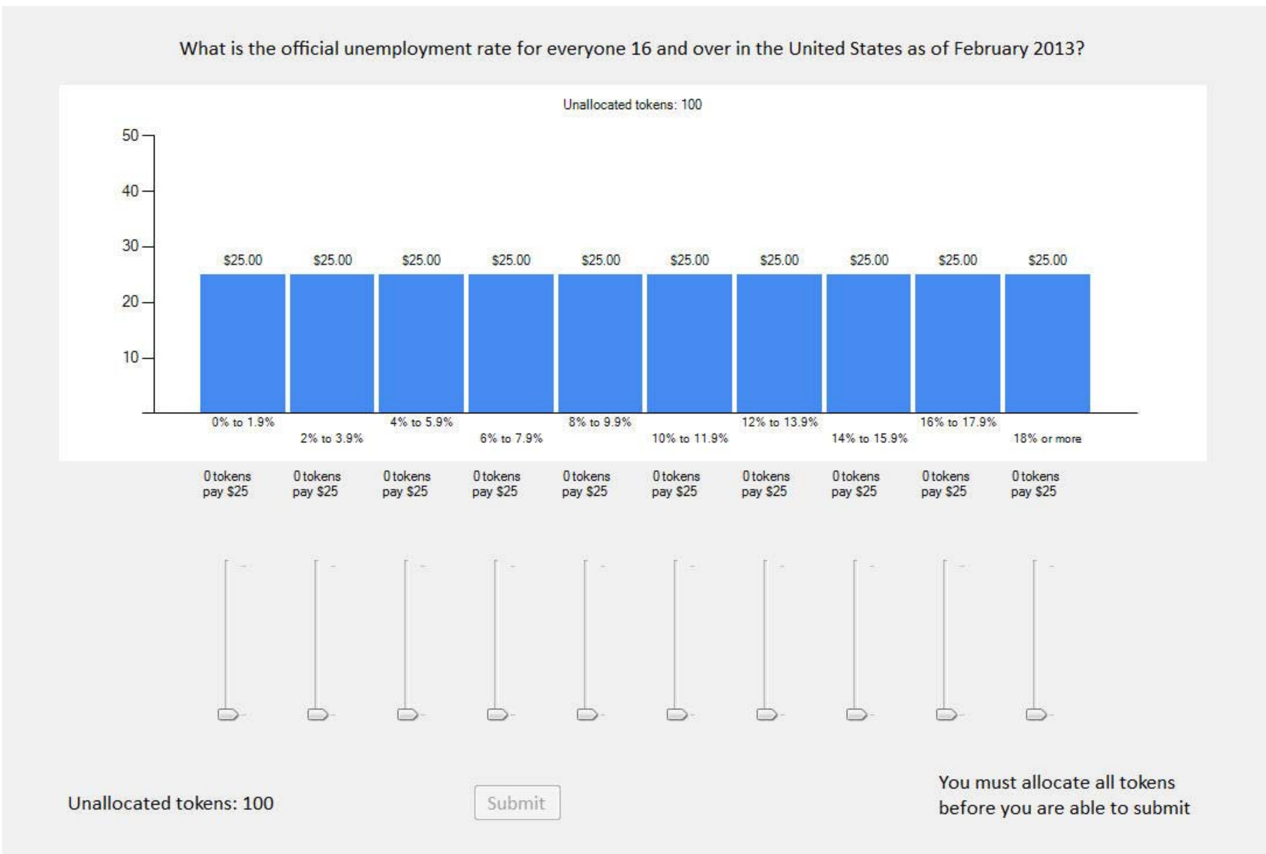
All payoffs are in cash, and are in addition to the show-up fee that you receive just for being here, as well as any other earnings in other tasks from the session today.

A.2. Subjective Beliefs

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 15 questions and asked to place some bets on your beliefs about the answers to each question. You will actually get the chance to be rewarded for your answers to one of the questions, so you should think carefully about your answer to each question.

Here is an example of what the computer display of such a question might look like.

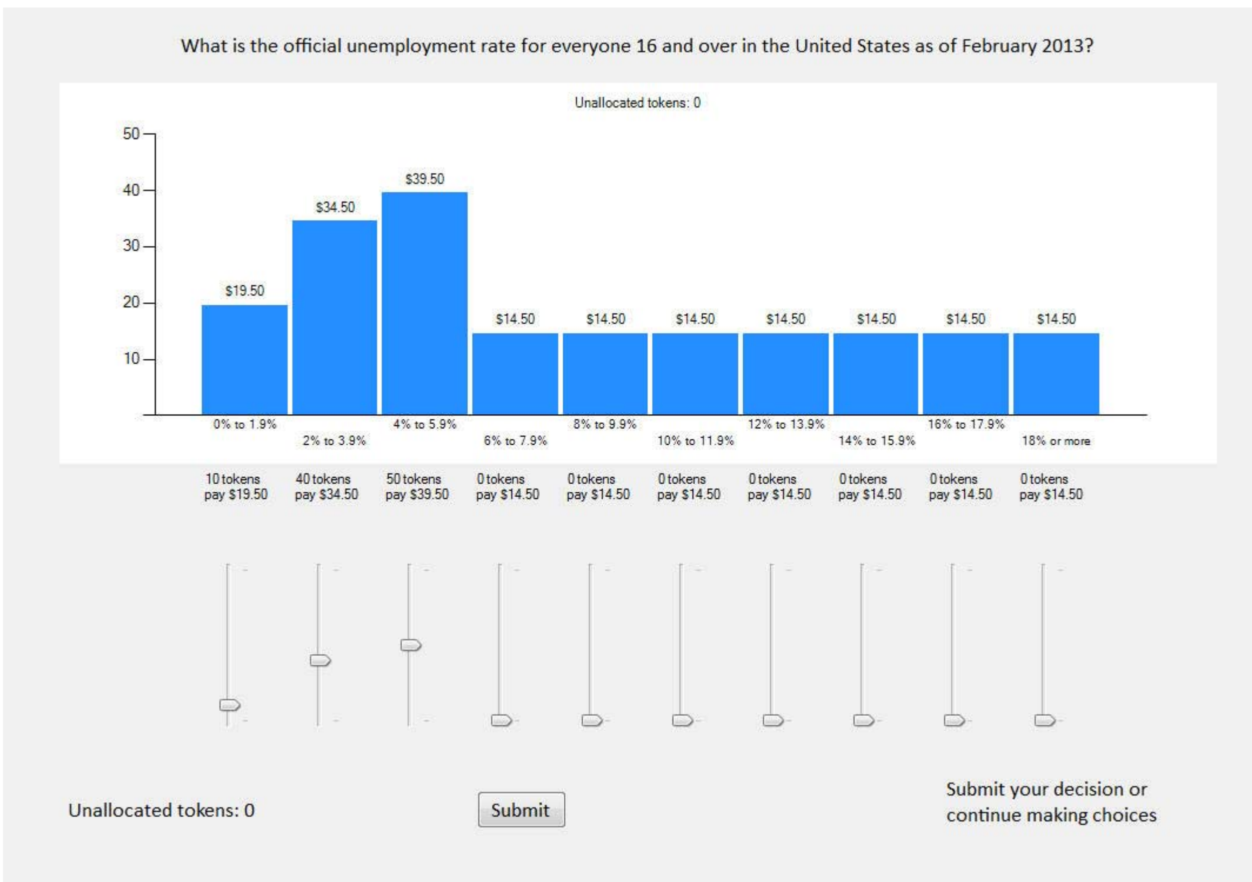


The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens, and in this example we start with 10 tokens allocated to each slider. 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 correct answer to the question. In the above example the tokens you allocate to each bar will naturally reflect your beliefs about the official unemployment rate for everyone 16 and over in February 2013. The first bar corresponds to your belief that the unemployment rate is between 0% and 1.9%. The second bar corresponds to your belief that the unemployment rate is between 2% and 3.9%, and so on. Each bar shows the amount of money you earn if the official unemployment rate is in the interval shown under the bar.

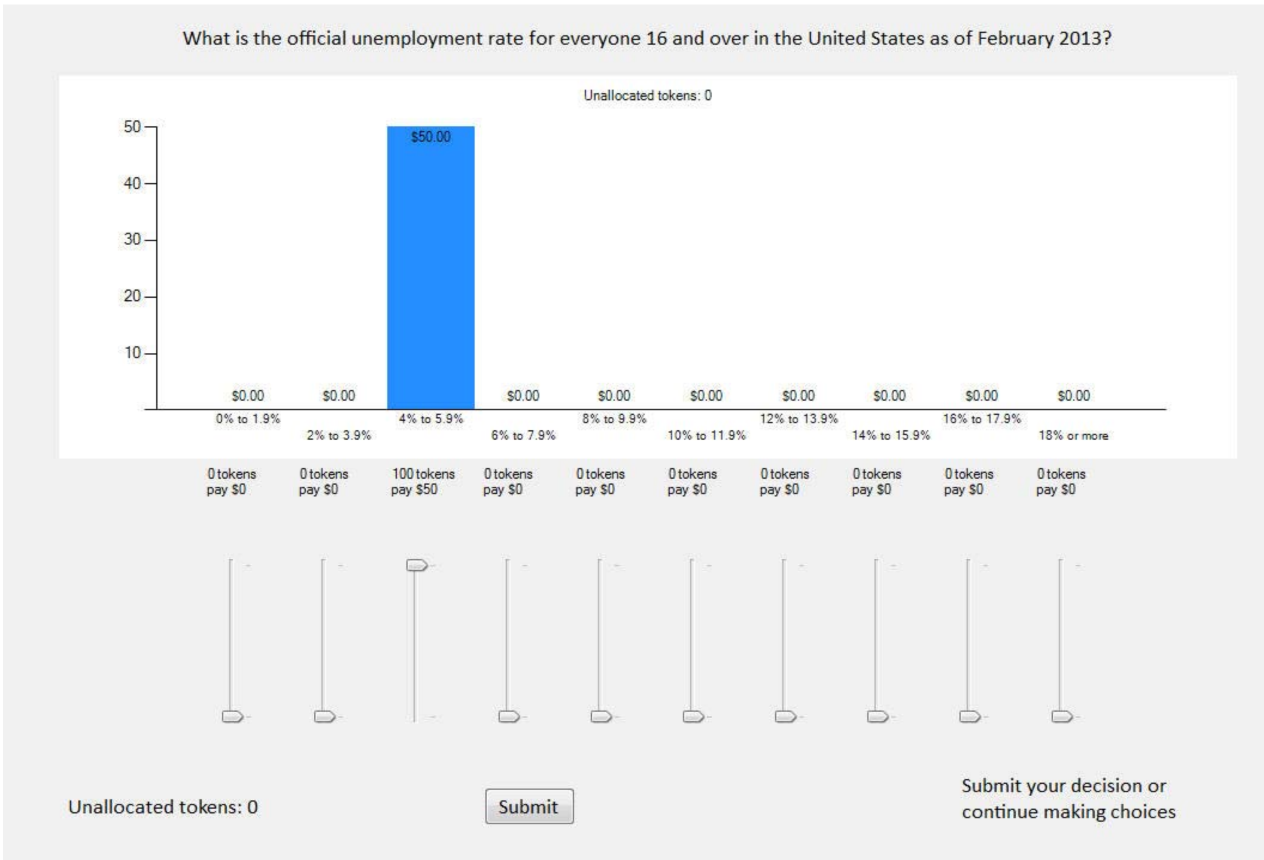
To illustrate how you use these sliders, suppose you think there is a fair chance the unemployment rate is just under 5%. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 4% to 5.9%, 40 tokens to the interval 2% to 3.9%, and 10 tokens to the interval 0% to 1.9%. So you can see in the picture below that if indeed the unemployment rate is between 4% and 5.9% you would earn $39.50. You would earn less than $39.50 for any other outcome. You would earn $34.50 if the unemployment rate is between 2% and 3.9%, $19.50 if it is between 0% and 1.9%, and for any other unemployment rate you would earn $14.50.



You can adjust the allocation as much as you want to best reflect your personal beliefs about the unemployment rate.

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. The true unemployment rate is actually 7.7%, according to the *Bureau of Labor Statistics*. So if you had reported the beliefs shown above, you would have earned $14.50.

Suppose you had put all of your eggs in one basket, and for example allocated 100 tokens to the interval corresponding to unemployment rates between 4% and 5.9%. Then you would have faced the earnings outcomes shown below.



Note the "good news" and "bad news" here. If the unemployment rate is indeed between 4% and 5.9%, you earn the maximum payoff, shown here as $50. But the true unemployment rate is 7.7%, so 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 correct answer to each question is a personal judgment that depends on the information you have about the topic of the question.**
- **Depending on your choices and the correct answer you can earn up to $50.**
- **Your choices might also depend on your willingness to take risks or to gamble.**

The decisions you make are a matter of personal choice. Please work silently, and make your choices by thinking carefully about the questions 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 roll a 30-sided die until a number between 1 and 15 comes up to determine which question will be played out. The experimenter will record your earnings according to the correct answer and the choices you made.

All payoffs are in cash, and are in addition to the show-up fee that you receive just for being here as well as any other earnings.

Are there any questions?

A.3. Lottery Parameters

Each subject is asked to make 50 binary choices between lotteries with objective probabilities. The battery of 50 lottery pairs are carefully selected for our purpose of identifying if the individual subject behaves consistently with EUT or RDU, as explained in the main text. We employ lotteries from two batteries designed to test RDU.

Wakker, Erev and Weber [1994] constructed lotteries to carefully test the "comonotonic independence" axiom of RDU. Their battery of 32 lottery pairs contains 24 that directly test this axiom, and 8 "filler" pairs added to avoid risk-neutral subjects employing a simple heuristic of selecting at random (since the 24 of interest had similar expected values within each pair). The main lottery pairs consist of 6 sets of 4 pairs. The logic of their design can be seen by examining the first set, from Wakker, Erev and Weber [1994; Figure 3.1] and reproduced here:

```
1st pair: (.55, $0.5;  .25, $6.0;  .20, $7.0) (safe) versus
          (.55, $0.5;  .25, $4.5;  .20, $9.0) (risky)

2nd pair: (.55, $3.5;  .25, $6.0;  .20, $7.0) (safe) versus
          (.55, $3.5;  .25, $4.5;  .20, $9.0) (risky)

3d  pair: (.55, $6.5;  .25, $6.0;  .20, $7.0) (safe) versus
          (.55, $6.5;  .25, $4.5;  .20, $9.0) (risky)

4th pair: (.55, $9.5;  .25, $6.0;  .20, $7.0) (safe) versus
          (.55, $9.5;  .25, $4.5;  .20, $9.0) (risky)
```

The second and third prizes in each pair stay the same within the set. The only thing that varies from pair to pair is the monetary value of the first prize, and that is common to the "safe" and "risky" lottery within each pair.[34] Since the first listed prize is a common consequence in both lotteries within a pair, it should not affect choices under EUT. In the 1st pair the first prize is only $0.50, and is the lowest ranked prize for both lotteries. The first prize increases to $3.50 for the 2nd pair, and is again the lowest ranked prize for both lotteries: so rank-dependence should have no effect on choice patterns as the subject moves from the 1st to the 2nd pair. But when we come to the 3rd pair the first prize is $6.50, which makes it the second highest ranked prize for both lotteries; this is where RDU *could* have a different prediction than EUT, depending on the extent and nature of probability weighting. Finally, in the 4th pair the common consequence is the highest ranked prize for both lotteries, again *allowing* RDU to predict something different from EUT (and from the choices in the 3rd pair). Note that this design does not formally *require* an RDU decision-maker to choose differently than an EUT decision-maker; it simply encourages it for *a priori* reasonable levels of probability weighting. We employ all 24 of their main lottery pairs, and scale the prizes considerably so that the highest prize is $78.

The remaining 26 lottery pairs for each subject were drawn from 69 pairs developed by Wilcox

[34] What is "safe" and what is "risky" is not so obvious when one allows for probability weighting, but this is how the lotteries are labeled.

[2010] for the purpose of robust estimation of EUT and RDU models.[35] These lottery pairs span five monetary prize amounts, $5, $10, $20, $35 and $70, and five probabilities, 0, ¼, ½, ¾ and 1. The prizes are combined in ten "contexts," defined as a particular triple of prizes.[36] These lotteries also contain a number of pairs in which the "EUT-safe" lottery has a *higher* EV than the "EUT-risky" lottery: this is designed deliberately to evaluate the extent of risk premia deriving from probability pessimism rather than diminishing marginal utility. Wilcox [2010] documents a wide variety of probability weighting functions from choices from the complete battery, based on estimates at the individual level.

The lottery parameters are documented below. The labels start at the 2nd outcome because of internal coding requirements to list outcomes in the order of prizes, and to have the highest prize be the 4th outcome. This is immaterial for the analysis, but means that for these lotteries the 1st prize and 1st probability default to 0, and need not be displayed. The "qid" label is the internal reference name for each lottery. Each subject received the chosen lotteries in random order, to mitigate any possible order effects.

Additional References

Wakker, Peter P.; Erev, Ido, and Weber, Elke U., "Comonotonic Independence: The Critical Test Between Classical and Rank-Dependent Utility Theories," *Journal of Risk and Uncertainty*, 9, 1994, 195-230.

Wilcox, Nathaniel T., "A Comparison of Three Probabilistic Models of Binary Discrete Choice Under Risk," *Working Paper*, Economic Science Institute, Chapman University, March 2010.

Wilcox, Nathaniel T., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 162(1), May 2011, 89-104.

[35] The original battery includes repetition of some choices, to help identify the "error rate" and hence the behavioral error parameter, defined later. In addition, the original battery was designed to be administered in its entirety to every subject. We decided *a priori* that 50 choice tasks was the maximum that our subject pool could focus on in any one session, given the need in some sessions for there to be later tasks. We also have some evidence that 50 choices of this kind is sufficient to reliably estimate EUT and RDU models at the level of the individual.

[36] For example, the first context consists of lotteries defined over the prizes $5, $10 and $20, and the tenth context consists of lotteries defined over the prizes $20, $35 and $70. The significance of the prize context is explained by Wilcox [2010][2011].

Lotteries from Wakker, Erev and Weber [1994]

qid	prob2L	prob3L	prob4L	prize2L	prize3L	prize4L	evL	prob2R	prob3R	prob4R	prize2R	prize3R	prize4R	evR
rWEW1	.55	.25	.2	3	36	42	19.05	.55	.25	.2	3	27	54	19.2
rWEW2	.55	.25	.2	21	36	42	28.95	.55	.25	.2	21	27	54	29.1
rWEW3	.25	.55	.2	36	39	42	38.85	.25	.55	.2	27	39	54	39
rWEW4	.25	.2	.55	36	42	57	48.75	.25	.2	.55	27	54	57	48.9
rWEW5	.65	.2	.15	3	21	33	11.1	.65	.2	.15	3	18	36	10.95
rWEW6	.2	.65	.15	21	15	33	18.9	.65	.2	.15	15	18	36	18.75
rWEW7	.2	.15	.65	21	33	27	26.7	.2	.65	.15	18	27	36	26.55
rWEW8	.2	.15	.65	21	33	39	34.5	.2	.15	.65	18	36	39	34.35
rWEW9	.4	.4	.2	3	15	36	14.4	.4	.4	.2	3	9	45	13.8
rWEW10	.4	.4	.2	15	18	36	20.4	.4	.4	.2	9	18	45	19.8
rWEW11	.4	.4	.2	15	33	36	26.4	.4	.4	.2	9	33	45	25.8
rWEW12	.4	.2	.4	15	36	48	32.4	.4	.2	.4	9	45	48	31.8
rWEW13	.7	.1	.2	15	33	63	26.4	.7	.1	.2	15	21	75	27.6
rWEW14	.1	.7	.2	33	36	63	41.1	.1	.7	.2	21	36	75	42.3
rWEW15	.1	.7	.2	33	57	63	55.8	.1	.7	.2	21	57	75	57
rWEW16	.1	.2	.7	33	63	78	70.5	.1	.2	.7	21	75	78	71.7
rWEW17	.5	.5	0	0	12	0	6	.6	.4	0	0	18	0	7.2
rWEW18	1	0	0	12	0	0	12	.1	.5	.4	0	12	18	13.2
rWEW19	.5	0	.5	12	0	24	18	.1	.4	.5	0	18	24	19.2
rWEW20	.5	0	.5	12	0	36	24	.1	.4	.5	0	18	36	25.2
rWEW21	.5	.5	0	12	24	0	18	.6	.4	0	12	30	0	19.2
rWEW22	1	0	0	24	0	0	24	.1	.5	.4	12	24	30	25.2
rWEW23	.5	0	.5	24	0	36	30	.1	.4	.5	12	30	36	31.2
rWEW24	.5	0	.5	24	0	48	36	.1	.4	.5	12	30	48	37.2

Lotteries from Wilcox [2010]

qid	prob2L	prob3L	prob4L	prize2L	prize3L	prize4L	evL	prob2R	prob3R	prob4R	prize2R	prize3R	prize4R	evR
r1	0	1	0	5	10	20	10	.25	0	.75	5	10	20	16.25
r2	.25	.75	0	5	10	20	8.75	.5	0	.5	5	10	20	12.5
r3	0	1	0	5	10	20	10	.5	0	.5	5	10	20	12.5
r4	.5	.5	0	5	10	20	7.5	.75	0	.25	5	10	20	8.75
r5	0	1	0	5	10	20	10	.25	.5	.25	5	10	20	11.25
r6	.25	.5	.25	5	10	20	11.25	.5	0	.5	5	10	20	12.5
r7	0	.5	.5	5	10	20	15	.25	0	.75	5	10	20	16.25
r8	0	.75	.25	5	10	20	12.5	.5	0	.5	5	10	20	12.5
r9	.25	.75	0	5	10	20	8.75	.75	0	.25	5	10	20	8.75
r10	0	1	0	5	10	20	10	.75	0	.25	5	10	20	8.75
r11	0	1	0	5	10	35	10	.5	0	.5	5	10	35	20
r12	0	.75	.25	5	10	35	16.25	.25	0	.75	5	10	35	27.5
r13	.25	.75	0	5	10	35	8.75	.75	0	.25	5	10	35	12.5
r14	0	.5	.5	5	10	35	22.5	.25	0	.75	5	10	35	27.5
r15	0	.75	.25	5	10	35	16.25	.5	0	.5	5	10	35	20
r16	0	1	0	5	10	35	10	.75	0	.25	5	10	35	12.5
r17	.25	.75	0	5	10	70	8.75	.5	0	.5	5	10	70	37.5
r18	0	1	0	5	10	70	10	.5	0	.5	5	10	70	37.5
r19	.5	.5	0	5	10	70	7.5	.75	0	.25	5	10	70	21.25
r20	0	1	0	5	10	70	10	.75	0	.25	5	10	70	21.25
r21	0	1	0	5	20	35	20	.25	0	.75	5	20	35	27.5
r22	0	.75	.25	5	20	35	23.75	.25	0	.75	5	20	35	27.5
r23	0	.5	.5	5	20	35	27.5	.25	0	.75	5	20	35	27.5
r24	0	1	0	5	20	35	20	.5	0	.5	5	20	35	20
r25	.5	.5	0	5	20	35	12.5	.75	0	.25	5	20	35	12.5
r26	0	.75	.25	5	20	35	23.75	.5	0	.5	5	20	35	20
r27	.25	.75	0	5	20	35	16.25	.75	0	.25	5	20	35	12.5
r28	.25	.75	0	5	20	70	16.25	.5	0	.5	5	20	70	37.5
r29	0	.75	.25	5	20	70	32.5	.25	0	.75	5	20	70	53.75
r30	.5	.5	0	5	20	70	12.5	.75	0	.25	5	20	70	21.25
r31	.25	.5	.25	5	20	70	28.75	.5	0	.5	5	20	70	37.5
r32	.25	.75	0	5	20	70	16.25	.75	0	.25	5	20	70	21.25
r33	0	.5	.5	5	20	70	45	.25	0	.75	5	20	70	53.75
r34	0	1	0	5	35	70	35	.25	0	.75	5	35	70	53.75
r35	.25	.75	0	5	35	70	27.5	.5	0	.5	5	35	70	37.5
r36	0	.75	.25	5	35	70	43.75	.25	0	.75	5	35	70	53.75
r37	.5	.5	0	5	35	70	20	.75	0	.25	5	35	70	21.25
r38	0	.5	.5	5	35	70	52.5	.25	0	.75	5	35	70	53.75
r39	0	.75	.25	5	35	70	43.75	.5	0	.5	5	35	70	37.5
r40	.25	.75	0	5	35	70	27.5	.75	0	.25	5	35	70	21.25
r41	0	1	0	5	35	70	35	.75	0	.25	5	35	70	21.25
r42	0	1	0	10	20	35	20	.25	0	.75	10	20	35	28.75
r43	.25	.75	0	10	20	35	17.5	.5	0	.5	10	20	35	22.5
r44	0	1	0	10	20	35	20	.25	.25	.5	10	20	35	25
r45	0	1	0	10	20	35	20	.5	0	.5	10	20	35	22.5

r46	0	1	0	10	20	35	20	.25	.5	.25	10	20	35	21.25
r47	0	.75	.25	10	20	35	23.75	.5	0	.5	10	20	35	22.5
r48	0	1	0	10	20	35	20	.5	.25	.25	10	20	35	18.75
r49	.25	.75	0	10	20	35	17.5	.75	0	.25	10	20	35	16.25
r50	0	1	0	10	20	35	20	.75	0	.25	10	20	35	16.25
r51	.25	.75	0	10	20	70	17.5	.5	0	.5	10	20	70	40
r52	.5	.5	0	10	20	70	15	.75	0	.25	10	20	70	25
r53	.25	.75	0	10	20	70	17.5	.75	0	.25	10	20	70	25
r54	0	1	0	10	35	70	35	.25	0	.75	10	35	70	55
r55	.25	.75	0	10	35	70	28.75	.5	0	.5	10	35	70	40
r56	0	.5	.5	10	35	70	52.5	.25	0	.75	10	35	70	55
r57	0	.75	.25	10	35	70	43.75	.5	0	.5	10	35	70	40
r58	0	1	0	20	35	70	35	.25	0	.75	20	35	70	57.5
r59	.25	.75	0	20	35	70	31.25	.5	0	.5	20	35	70	45
r60	0	.75	.25	20	35	70	43.75	.25	0	.75	20	35	70	57.5
r61	0	1	0	20	35	70	35	.5	0	.5	20	35	70	45
r62	.5	.5	0	20	35	70	27.5	.75	0	.25	20	35	70	32.5
r63	0	1	0	20	35	70	35	.25	.5	.25	20	35	70	40
r64	.25	.5	.25	20	35	70	40	.5	0	.5	20	35	70	45
r65	0	.5	.5	20	35	70	52.5	.25	0	.75	20	35	70	57.5
r66	0	1	0	20	35	70	35	.5	.25	.25	20	35	70	36.25
r67	.25	.75	0	20	35	70	31.25	.75	0	.25	20	35	70	32.5
r68	0	.75	.25	20	35	70	43.75	.5	0	.5	20	35	70	45
r69	0	1	0	20	35	70	35	.75	0	.25	20	35	70	32.5

Appendix B: Estimating RDU Models of Decision-Making (Online Working Paper)

We write out the formal econometric specifications for EUT and RDU models, to be applied to determine the probability that individual subjects behave consistently with EUT. The notation here may differ slightly from the main text, and repeat certain equations so as to be self-contained.

A. Expected Utility

Assume that utility of income is defined by

$$U(x) = x^{(1-s)}/(1-s) \tag{B1}$$

where x is the lottery prize and $s \neq 1$ is a parameter to be estimated. For s=1 assume U(x)=ln(x) if needed. Thus s is the coefficient of CRRA: s=0 corresponds to risk neutrality, s<0 to risk loving, and s>0 to risk aversion. Let there be J possible outcomes in a lottery. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i:

$$EU_i = \sum_{j=1,J} [\, p(x_j) \times U(x_j) \,]. \tag{B2}$$

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

$$\nabla EU = EU_R - EU_L \tag{B3}$$

calculated, where EU_L is the "left" lottery and EU_R is the "right" lottery as presented to subjects. This latent index, based on latent preferences, is then linked to observed choices using a standard cumulative normal distribution function $\Phi(\nabla EU)$. This "probit" function takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the probit link function,

$$\text{prob(choose lottery R)} = \Phi(\nabla EU) \tag{B4}$$

Even though this "link function" is common in econometrics texts, it is worth noting explicitly and understanding. It forms the critical statistical link between observed binary choices, the latent structure generating the index ∇EU, and the probability of that index being observed. The index defined by (B3) is linked to the observed choices by specifying that the R lottery is chosen when $\Phi(\nabla EU) > \frac{1}{2}$, which is implied by (B4).

Thus the likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of ρ given the above statistical specification and the observed choices. The "statistical specification" here includes assuming some functional form for the cumulative density function (CDF). The conditional log-likelihood is then

$$\ln L(s; y, \mathbf{X}) = \sum_i [\, (\ln \Phi(\nabla EU) \times \mathbf{I}(y_i = 1)) + (\ln (1-\Phi(\nabla EU))) \times \mathbf{I}(y_i = -1)) \,] \tag{B5}$$

where $\mathbf{I}(\cdot)$ is the indicator function, $y_i = 1(-1)$ denotes the choice of the right (left) lottery in risk aversion task i, and $\mathbf{X}$ is a vector of individual characteristics reflecting age, sex, race, and so on.

Harrison and Rutström [2008; Appendix F] review procedures that can be used to estimate structural models of this kind, as well as more complex non-EUT models. The goal is to illustrate how researches can write explicit maximum likelihood (ML) routines that are specific to different structural choice models. It is a simple matter to correct for multiple responses from the same subject ("clustering"), as needed for the pooled estimation results we present.

An important extension of the core model is to allow for subjects to make some *behavioral* errors. The notion of error is one that has already been encountered in the form of the statistical assumption that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery. This assumption is clear in the use of a non-degenerate link function between the latent index ∇EU and the probability of picking one or other lottery; in the case of the normal CDF, this link function is $\Phi(\nabla EU)$. If there were no errors from the perspective of EUT, this function would be a step function: zero for all values of $\nabla EU<0$, anywhere between 0 and 1 for $\nabla EU=0$, and 1 for all values of $\nabla EU>0$.

We employ the error specification originally due to Fechner and popularized by Hey and Orme [1994]. This error specification posits the latent index

$$\nabla EU = (EU_R - EU_L)/\mu \tag{B3$'$}$$

instead of (B3), where μ is a structural "noise parameter" used to allow some errors from the perspective of the deterministic EUT model. This is just one of several different types of error story that could be used, and Wilcox [2008] provides an excellent review of the implications of the alternatives. As $\mu \rightarrow 0$ this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries; but as μ gets larger and larger the choice essentially becomes random. When $\mu=1$ this specification collapses to (B3), where the probability of picking one lottery is given by the ratio of the EU of one lottery to the sum of the EU of both lotteries. Thus μ can be viewed as a parameter that flattens out the link functions as it gets larger.

An important contribution to the characterization of behavioral errors is the "contextual error" specification proposed by Wilcox [2011]. It is designed to allow robust inferences about the primitive "more stochastically risk averse than," and posits the latent index

$$\nabla EU = [\, (EU_R - EU_L)/\nu \,]/\mu \tag{B3$''$}$$

instead of (B3$'$), where ν is a new, normalizing term for each lottery pair L and R. The normalizing term ν is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair. The value of ν varies, in principle, from lottery choice pair to lottery choice pair: hence it is said to be "contextual." For the Fechner specification, dividing by ν ensures that the *normalized* EU difference $[(EU_R - EU_L)/\nu]$ remains in the unit interval. The term ν does not need to be estimated in addition to the utility function parameters and the parameter for the behavioral error tern, since it is given by the data and the assumed values of those estimated parameters.

The specification employed here is the CRRA utility function from (B1), the Fechner error specification using contextual utility from (B3$''$), and the link function using the normal CDF from (B4). The log-likelihood is then

$$\ln L(s, \mu; y, \mathbf{X}) = \textstyle\sum_i [\, (\ln \Phi(\nabla EU)\times\mathbf{I}(y_i = 1)) + (\ln (1-\Phi(\nabla EU))\times\mathbf{I}(y_i = -1)) \,] \tag{B5$''$}$$

and the parameters to be estimated are s and μ given observed data on the binary choices y and the lottery parameters in **X**.

It is possible to consider more flexible utility functions than the CRRA specification in (1), but that is not essential for present purposes.

B. Rank-Dependent Utility

The RDU model of Quiggin [1982] extends the EUT model by allowing for decision weights on lottery outcomes. The specification of the utility function is the same parametric specification (B1) considered for EUT. To calculate decision weights under RDU one replaces expected utility defined by (B3) with RDU

$$RDU_i = \sum_{j=1,J} [\ w(p(x_j)) \times U(x_j)\ ] = \sum_{j=1,J} [\ w_j \times U(x_j)\ ] \tag{B3$'$}$$

where

$$w_j = \omega(p_j + ... + p_J) - \omega(p_{j+1} + ... + p_J) \tag{B6a}$$

for j=1,... , J-1, and

$$w_j = \omega(p_j) \tag{B6b}$$

for j=J, with the subscript j ranking outcomes from worst to best, and $\omega(\cdot)$ is some probability weighting function.

We consider three popular probability weighting functions. The first is the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ:

$$\omega(p) = p^\gamma \tag{B7}$$

So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \ \ \forall p$ and hence the "RDU EV" weighted by $\omega(p)$ instead of p has to be less than the EV weighted by p. The rest of the ML specification for the RDU model is identical to the specification for the EUT model, but with different parameters to estimate.

The second probability weighting function is the "inverse-S" function popularized by Tversky and Kahneman [1992]:

$$\omega(p) = p^\gamma \,/\, (\ p^\gamma + (1-p)^\gamma\ )^{1/\gamma} \tag{B8}$$

This function exhibits inverse-S probability weighting (optimism for small p, and pessimism for large p) for $\gamma<1$, and S-shaped probability weighting (pessimism for small p, and optimism for large p) for $\gamma>1$.

The third probability weighting function is a general functional form proposed by Prelec [1998] that exhibits considerable flexibility. This function is

$$\omega(p) = \exp\{-\eta(-\ln p)^\varphi\}, \tag{B9}$$

and is defined for $0<p\leq 1$, $\eta>0$ and $\varphi>1$. When $\varphi=1$ this function collapses to the Power function $\omega(p) = p^{\eta}$.

The construction of the log-likelihood for the RDU model with Power or Inverse-S probability weighting follows the same pattern as for EUT, with the parameters s, γ and μ to be estimated. The log-likelihood for the RDU model with the Prelec probability weighting requires the estimation of the parameters s, η, φ and μ to be estimated.

Additional References

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Wilcox, Nathaniel T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison," in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Wilcox, Nathaniel T., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 162(1), May 2011, 89-104.