Recovering Subjective Probability Distributions:

A Bayesian Approach

by

Glenn W. Harrison, Brian Monroe and Eric R. Ulm †

September 2022

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. We demonstrate how to fully recover latent subjective beliefs if the individual is known to distort probabilities into decision weights using Rank Dependent Utility theory, and generalize all results for the complete class of proper scoring rules. As an immediate corollary, we can then exactly recover latent subjective beliefs using Expected Utility Theory. Using Bayesian econometric methods, we demonstrate how to recover belief distributions for individuals, and illustrate these results by recovering belief distributions about the addictiveness of smoking. These results significantly widen the domain of applicability of proper scoring rules for eliciting latent subject belief distributions.

KEYWORDS: Subjective belief distributions; scoring rules; experimental methods; Bayesian inference
JEL CODES: D81, D83, C11, C81, C91

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Research Unit in Behavioural Economics and Neuroeconomics, University of Cape Town, South Africa (Monroe); and School of Economics and Finance, Wellington School of Business and Government, Victoria University of Wellington, New Zealand (Ulm). Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contacts: gharrison@gsu.edu, brianalbertmonroe@gmail.com and Eric.Ulm@vuw.ac.nz.

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 known since Harrison, Martínez-Correa, Swarthout and Ulm [2017] 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. This extension 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 complete class of proper scoring rules, of which the QSR is just the most popular.

These theoretical results assume that one knows the risk preferences of an individual. However, those preferences are also latent, and typically generated using econometric methods which must allow

[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 various 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 test hypotheses that span the full latent distribution, such as hypotheses about bias and confidence, by making a series of independent inferences about binary slices of the underlying distribution.

for statistical imprecision in the estimates.[2] To account for this statistical imprecision when recovering beliefs a natural approach is to utilize Bayesian econometric methods. One reason this approach is more natural is that the posterior distribution from Bayesian inference these days invariably takes the form of a simulated set of M samples using Markov Chain Monte Carlo (MCMC) algorithms, and our theoretical results apply *exactly* to each of those M samples.[3] Hence our theoretical results allow us to *directly* derive a Bayesian posterior predictive distribution for the subjective beliefs of an individual.[4] Moreover, this derivation is virtually instantaneous computationally, allowing for easy application using common software.

Another reason it is natural to use Bayesian methods is that we want to recover beliefs at the level of the individual, and Bayesian Hierarchical Models (BHM) provide an attractive way to estimate risk preferences at the level of the individual. There are well-understood limitations of classical methods at the level of the individual, particularly when sample sizes for individuals are not designed for individual-level estimation due to time constraints that limit the number of risky choices one can expect to a subject to make. Gao, Harrison and Tchernis [2022] discuss these issues, the way in which a BHM mitigates them, and also provide software templates for application.

We demonstrate the application of these theoretical results by recovering the latent subjective belief distributions from observed reports by individuals in an incentivized experiment, and for whom

[2] The notion of statistical imprecision means fundamentally different things to classical and Bayesian econometricians.

[3] One does not have to be using Bayesian econometric methods to generate estimates using simulations. This is common in econometric methods using so-called "random coefficients," also known as "mixed estimation." For example, Andersen, Hole, Harrison, Lau and Rutström [2012] use random coefficient methods to simulate risk preferences for *pooled* samples. The Bayesian approach is more natural for generating estimates for individual subjects.

[4] This is formally the same as the calculations of the Bayesian posterior predictive distribution for the *normative* welfare effects of insurance purchase decisions in Gao, Harrison and Tchernis [2022]. Here the application is *descriptive*, closer to the joint estimation of risk preferences and subjective probabilities for binary events of Andersen, Fountain, Harrison, Lau and Rutström [2014] using classical econometric methods.

we also have individual estimates of their risk preferences using a BHM. We show that the recovered belief distributions of RDU-consistent individuals exhibit first-order differences from observed reports. The extent of the difference between observed and recovered beliefs intuitively 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.

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 beliefs when one cannot be certain that the individual is risk neutral.

Section 1 briefly reviews different approaches to eliciting subjective belief distributions, section 2 provides new theoretical results for belief recovery under RDU, and section 3 illustrates the application of our Bayesian approach to beliefs about the addictiveness of smoking. Online appendices provide proofs of formal results, instructions for the experiments, and documentation of computer software we provide to undertake these calculations.

1. Eliciting Belief Distributions

It is becoming popular to elicit subjective belief *distributions*, since they provide information on the bias of beliefs as well as the confidence with which those beliefs are held. There are three ways to elicit beliefs, each with strengths and weaknesses for different applications.

Following Manski [2014], belief distributions might be elicited using hypothetical surveys. The key feature here is the absence of any incentives for responses. The advantage of this approach is that it is cheap, easier to explain to subjects, easier to implement in software, and allows questions about events that cannot be verified. Considerable attention has been paid to the manner in which belief questions are presented, particularly in field settings: see Delavande, Giné and McKenzie [2011]. The disadvantage of this approach is that the results might exhibit hypothetical bias, and it is easy to document that this bias can be a significant one (Harrison [2016]). What "hypothetical bias" means here is that one gets different results, usually on a between-subjects basis, when asking the same question with no incentives compared to asking with incentives. In the nature of subjective beliefs, there is no true answer that one can look to: fidelity with some objective outcome is no metric of value here, unless one wants to impose some "rational expectations" constraint on beliefs, which we do not. So the expression "hypothetical bias" is just a short-hand for the results being different with and without incentives, and the prior that incentivized responses are likely to reflect more effort and willingness to respond truthfully than non-incentivized responses. Although this is our prior, we appreciate that others might not share it.

For us, the primary remaining advantage of hypothetical eliciation is the ability to ask questions about non-verifiable events. For example, if someone is interested in longevity risk, one could ask a series of incentivized questions about how long the subject thinks people in their country will live, or someone with their gender, or someone with their gender and race, or someone with their gender, race and income level, and so on. Each of these could be incentivized with respect to official mortality tables. But then a question about how long the specific subject will live cannot be so incentivized, and is what we might be interested in. To take this example, we would ask all of the incentivized questions and then one non-incentivized question, and evaluate the bias and confidence of all but the last in relation to verifiable data. This would then give us a prior on the bias and confidence for the final hypothetical question. In this manner we see hypothetical and incentivized belief questions as complementary.

A second approach to belief elicitation starts by recognizing that risk preferences affect the rational responses of subjects to the usual scoring rules, but that subjects should rationally report their beliefs if they are risk neutral. Hence one can append an experimental payment procedure, called the Binary Lottery Procedure (BLP), to "risk neutralize" the responses of the subject and view reports directly as beliefs. The BLP was developed by Smith [1961], and has been widely used in various settings in which risk preferences are a confound, such as the bargaining experiments of Roth and Malouf [1979].[5] The first statements of this mechanism, joining the QSR and the BLP, appear to be Allen [1987] and McKelvey and Page [1990]. Schlag and ven der Weel [2013] and Hossain and Okui [2013] examine the same extension of the QSR, along with certain generalizations, renaming it a "randomized QSR" and "binarized scoring rule," respectively. Harrison, Martínez-Correa, Swarthout and Ulm [2015] evaluated the QSR with the BLP in terms of the elicitation of subjective belief distributions, not just some summary statistic of that distribution. Harrison and Phillips [2014] applied this method to elict the beliefs about financial risk by Chief Risk Officers of major companies, all of whom had advanced training in statistical and actuarial methods.

The clear strength of this approach is that it allows the reports of the subject to be evaluated directly as beliefs, under the maintained assumption that the BLP works as advertized by theory. The disadvantage is that it adds an additional layer of understanding for subjects when it comes to translating earnings in the QSR into cash.[6] In our experience, documented in Harrison et al.

[5] Berg, Daley, Dickhaut and O'Brien [1986] extended the BLP to allow for a non-linear exchange rate between the "points" that define the binary probability lottery chances of a bigger prize. This extension allows the experimenter to induce risk averse or risk-loving preferences with concave or convex exchange rate functions, albeit with added complexity for subjects.

[6] Although there are some studies showing the apparent failure of the BLP in other settings, that evidence is not as compelling as many claim: see Harrison et al. [2013] for a critical literature review. The claim by Danz, Vesterlund and Wilson [2022] that the QSR using the BLP fails metrics of "behavioral incentive compatibility" is premature: they explicitly assume that subjects correctly understand the objective likelihoods the experimenter induces. They correctly note (p. 2853) that the real "challenge for examining whether information on the mechanism's quantitative incentives encourages truth telling is that we do not know participants' true beliefs." But they then take the extreme metric of "reports of the objective prior as truthful [subjective beliefs]. This true/false terminology is chosen for clarity, and does not imply that all participants are assumed to understand that the objective prior is the true likelihood." So they are testing some variant of a rational expectations model of subjective beliefs *jointly with* the incentive compatibility of the QSR using the BLP applied to the innate subjective beliefs subjects might hold, and they are not just testing the later hypothesis.

[2013][2014][2015], this is relatively easy to overcome: one simply defines all earnings from the QSR in terms of points rather than a natural currency, and then add a paragraph at the end explaining how points get turned into cash using the BLP.[7] Nonetheless, there are many settings in which one does not want to have to assume understanding of the BLP mechanism, particularly for less formally literate field populations who, in our experience, can be wary of "sophisticated", indirect payment protocols.

The third approach is to just implement some scoring rule, such as the QSR, and pay the subjects the stated rewards in a natural currency. The strength of this approach is that it allows the subject to see the monetary bets that her reports are generating, just as if she was placing bets with a series of bookies at some sporting event. This framing of the reports as bets is made more transparent with the use of real-time interfaces of the kind we use, developed by Harrison et al. [2017]. Early implementations of the QSR relied on subject's understanding algorithms or squinting at long numerical tabulations of potential payoffs, but those have not been used for decades.

The weakness of this approach is that one must account for the effect of risk preferences on stated reports. This requires one additional task be conducted to elicit risk preferences, some econometrics to infer risk preferences at the individual level, and the recovery of beliefs once one has those estimated risk preferences in hand. As demonstrated by Gao, Harrison and Tchernis [2022], the use of a BHM allows one to reduce the number of choice tasks required of each subject significantly, and still maintain reliable inferences using informative priors from the choices of other individuals. This speeds up the overall task of eliciting risk preferences, as well as providing estimates at the individual level. And our present results address the final issue, the difficulty of actually inferring latent

[7] We avoid reference to "experimental currencies" and a stated exchange rate between those currencies and some natural currency. This device is often used to just scale up rewards in some framed experimental currency, with the hope that this motivates subjects better. This procedure risks the confound of money illusion: if the subjects do not suffer from money illusion, the procedure adds nothing to incentives in terms of the natural currency, but if they do suffer from money illusion the experimenter has lost control of incentives.

beliefs from observed reports.

In summary, we see the second and third approaches as the most attractive, and complementary, and have used both in different settings. They trade off the "extra work" needed to rigorously identify subjective beliefs. The second approach effectively puts that burden on the subject, and the third approach effectively puts that burden on the experimenter.

2. Theoretical Results

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.

Let the decision maker report her subjective beliefs in a discrete version of a QSR for continuous distributions (Matheson and Winkler [1976]).[8] 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 $\sum_{i=1...K} (r_i) = 1$.

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)]: $S = (2 \times r_k) - \sum_{i=1...K} (r_i)^2$. So the reward in the score is a doubling of the report allocated to the true interval, and the penalty depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval the punishment is severe. The punishment includes all possible reports, including the

[8] 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.

correct one.[9]

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 α=0 and β=1. We can assume α>0 and β>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 visual representation of the QSR, which we will use in experiments, for α = β = 25 and K=10.

We restate Lemma 1 from Harrison, Martínez-Correa, Swarthout and Ulm [2017]:

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

[9] 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 = $(2 \times ¼) - ((¼)^2 + (¼)^2 + (¼)^2 + (¼)^2) = ½ - ¼ = ¼ < 1$. So the tradeoff from the last case is that one can always ensure a score of ¼, but there is an incentive to provide less diffuse reports, and that incentive is the possibility of a score of 1.

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.[10] 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

[10] 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.

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]:

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

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 $\gamma>1$.

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 $\varphi>0$.[11] When $\varphi=1$ this function collapses to the Power function $\omega(p) = p^{\eta}$.

One important reason to be able to recover beliefs that accommodate these different specifications is that there is clear evidence that there is considerable individual heterogeneity in the *type* of probability weighting that individuals exhibit (Wilcox [2022]). It is simply not true that many, or even

[11] 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.

most, individuals exhibit behavior consistent with "inverse-S" probability weighting functions. Nor is much gained by further debating that point, when we can span these three specifications.

We generalize Lemma 1 to include individuals that distort probabilities, with all proofs in Appendix A (online):

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

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

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.[12] 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 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 $\partial u/\partial\theta = \theta^{-s}$.[13] For RDU individuals further assume the inverse-S probability

---

[12] While these equations can be solved using standard linear algebra techniques, it can be shown that the exact solution is $w(p_k) = r_k \; [\partial\theta/\partial u \,|_{\theta=\theta(k)}] \,/\, \sum_{j=1 \ldots K} r_j \; [\partial\theta/\partial u \,|_{\theta=\theta(j)}]$. In the CRRA case we are considering here, this becomes $w(p_k) = r_k \; \theta(k)^{-s} \,/\, \sum_{j=1 \ldots K} r_j \; \theta(j)^{-s}$.

[13] 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 $\exp\{-\theta^{(1-s)}\alpha\}\ (1-s)\ \theta^{-s}$.

weighting function (7) with $\gamma = 0.5$ and numerically invert (7).[14]

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

$$0.049591 \times w(p_3) - 0.018 \times w(p_4) - 0.02267 \times w(p_5) = 0 \quad (10a)$$
$$-0.03188 \times w(p_3) + 0.032997 \times w(p_4) - 0.034 \times w(p_5) = 0 \quad (10b)$$
$$w(p_3) + w(p_4) + w(p_5) = 1 \quad (10c)$$

The numbers in these 3 equations are direct applications of (9). For example, from (10a) we have

$$0.049591 = 0.7 \times 0.07084443 = (1\text{-}r_3) \times [25 + (50 \times 0.30) - (25 \times (0.30^2 + 0.45^2 + 0.25^2))]^{-0.77}$$
$$-0.018 = -0.3 \times 0.059994478 = (-r_3) \times [25 + (50 \times 0.45) - (25 \times (0.30^2 + 0.45^2 + 0.25^2))]^{-0.77}$$
$$-0.02267 = -0.3 \times 0.075562439 = (-r_3) \times [25 + (50 \times 0.25) - (25 \times (0.30^2 + 0.45^2 + 0.25^2))]^{-0.77}.$$

We solve (10a), (10b) and (10c) for decision weights $w(p_3) = (0.3/0.07084443)/15.04384315 = 0.281486$, $w(p_4) = (0.45/0.059994478)/15.04384315 = 0.498589$ and $w(p_5)$ $(0.25/0.075562439)/15.04384315 = 0.219925$. 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, Martínez-Correa, Swarthout and Ulm [2017], and the assumed value of s, the differences between these weights and the observed reports are small. The mean of the observed reports is 34.5 if the bins intervals are 0 to 10, 11 to 20, ..., 91 to 100, and the mean of the recovered beliefs 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

[14] 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 $\exp\{-[(\ln p)/(-\eta)]^{1/\varphi}\}$.

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 mean of the recovered beliefs in this case is 33.39445, noticeably different from the mean of 34.5 for observed reports.[15]

This example illustrates the computational steps involved. We have developed a *Stata* program to undertake these calculations for practical applications, and document this in Appendix B (online). The program handles CRRA, Power, CARA or Expo-Power utility functions, and Power, Inverse-S or Prelec probability weighting functions, where EUT is a special case.

Finally, we can also prove a generalization to all *proper* scoring rules. 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. Our generalization is then:

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

[15] 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.

This also means that Propositions 1 through 7 of Harrison, Martínez-Correa, Swarthout and Ulm [2017], characterizing the beliefs recovered for an SEU decision-maker, also generalize.

3. Beliefs About the Addictiveness of Smoking

To illustrate our theoretical results, we examine their application in an incentivized experiment examining beliefs about a basic informational input to the decision to smoke when young: beliefs about how addictive smoking is.[16] The "rational model of addiction" of Becker and Murphy [1988] has been criticized for the assumption that the addict has full information about the consequences of starting to consume cigarettes. Many daily adult smokers started smoking when they were teenagers, and were able to quit for periods relatively easily in their early years of smoking. Hence they could easily form beliefs that smoking is not very addictive, precisely due to them have good experience with their ability to quit when just starting to smoke and, rationally *in terms of that personal data*, forming beliefs that quitting is easy in general. Hence they could plausibly develop poor beliefs about the *longer-term* addictiveness of cigarettes after some initial period of consumption.

Orphanides and Zervos [1995; p.740] hypothesize

> ... that the bulk of objections concerning earlier rational models can be attributed *not* to rational decision making, but rather to the common implicit assumption of perfect foresight. The essential feature lacking from these models is the recognition that inexperienced individuals are initially uncertain of the exact potential harm associated with consuming an addictive good. Once uncertainty and a process of learning through experimentation are incorporated into the earlier rational framework, the process of rationally getting "hooked" into an addiction becomes evident, and our understanding of the determinants of addiction is substantially improved.

They present a model in which Bayesian updating, starting from optimism about the ease of quitting

[16] Declaration of Competing Interests: Harrison has served as a testifying expert witness for plaintiffs in litigation against tobacco companies in the United States and Canada for 25 years, and continues to do so. His testimony is primarily on health care costs to governments associated with tobacco-related diseases, and he has received compensation for the research and time supporting this testimony. The design, of the experiments reported here was undertaken independently of any litigation, and subject fees were funded by the Center for the Economic Analysis of Risk at Georgia State University.

smoking, could lead to young smokers become addicted as the result of past consumption. The validity of the "rational addiction" model is not the issue here, just the motivation for wanting to examine subjective beliefs of young people about the addictiveness of smoking cigarettes. A characterization of these beliefs is presumably of value to any model of smoking addiction, as well evaluations of informational public polices.

The two belief questions we asked about addictiveness[17] were as follows, with response intervals in square brackets and the correct answers used to reward subjects in curly brackets:

1. Over a 5-year period, 687 out of 1000 **current adult smokers** reported quitting for **at least a month**. How many of the 687 relapsed, and **started smoking again**? [None, 1 to 99, 100 to 199, 200 to 299, 300 to 399, 400 to 499, 500 to 599, 600 to 687] {Correct answer is 532: see Borland et al. [2011; p.677]}.
2. On average, how many times does a **current adult smoker** try to quit before successfully quitting for **1 year or more**? [1, 2 or 3, 4 or 5, 6 or 7, 8 or 9, 10 to 14, 15 to 19, 20 to 24, 25 to 35, 36 or more] {Correct answer is 29.6: see Chaiton [2016; p.5]}.

Question #1 used 8 bins for responses, and questions #2 used 10 bins. All questions allowed 100 tokens, and QSR parameters $\alpha = \beta = 25$ resulting in a $50 payment if all 100 tokens were allocated to the correct answer.

Figures 1 and 2 illustrate the interface used to elicit beliefs from subjects, developed by Harrison, Martínez-Correa, Swarthout and Ulm [2017]. Appendix C (online) contains the instructions presented to subjects. Each subject allocates 100 token across 10 or 8 bins, and the displays change in real-time so the subject sees the payoffs from the bets being placed about the correct answer. As

[17] Two prior questions about the mortality effects of smoking were asked. The first was "Men and women over 55 who **currently smoke** are how many more times as likely to die of **lung cancer** as men and women over 55 who have **never smoked**?" and the second was "Men and women over 55 who are **former smokers** are how many more times as likely to die of **lung cancer** as men and women over 55 who have **never smoked**?". From data in Thun et al. [2013; pages 356 and 358], the correct answer for the first (second) question was 25.66 (6.70) for women and 24.97 (6.75) for men, so about 25 (6.7) times. These questions were incentivized elicitations of mortality risk questions widely used by Professor W. Kip Viscusi in defense testimony for tobacco companies in the United States and Canada (e.g., Viscusi [2002; chapter 7]). One additional question was asked at the end, "Based on survey data from 2012/13, what is the likelihood of somebody being **clinically dependent on nicotine** if they are a **current daily smoker** and started smoking up to 40 years earlier? From updated calculations using the statistical model of Harrison [2017], the correct answer is 85%.

stressed by Savage [1971], scoring rules can be viewed as formalizations of asking individuals to stake an endowment with an array of "bookies" offering different odds. After responses for all belief questions are collected, one question is selected at random for payment and the subject receives the cash payment indicated on their report for the bin corresponding to the correct answer.

Subjects were recruited in 2019 and early 2020 from the population of undergraduate students from a variety of majors at Georgia State University. All subjects had also completed 100 incentivized, binary choices over risky lotteries, in experimental sessions documented by Harrison, Morsink and Schneider [2020]. The BHM of Gao, Harrison and Tchernis [2022] was applied to these data, assuming an RDU model of risk preferences with a CRRA utility function and the Prelec probability weighting function. The result of that estimation is a Bayesian posterior distribution for each subject with M = 25,000 samples for each of the parameters of an RDU model of their risk preferences.

In our sample of 383 the average age was 20, 64% were self-reported as female, 62% were black, 27% had a business major, 36% were Juniors, Seniors or Graduate students, 46% were working part-time or full-time, 18% had parents with a joint annual income between $65,000 and $100,000, 15% had parents with a joint annual income over $100,000, 12% reported a non-Christian religion, and 19% reported being an atheist, agnostic or non-religious. Very few students report being current smokers, and the sample was not constructed to identify smokers and non-smokers.

We report results by examining the recovered beliefs using the posterior predictive distribution.[18] We compare these to the raw reports. Our belief recovery is at the individual subject and question level, but of course we can pool results over subjects. One could then easily stratify results by demographic or treatment characteristics.

[18] Each of the 25,000 samples of estimates of the set $\{s, \eta$ and $\varrho\}$ generate beliefs, by definition, that sum to 1, and we confirmed this. In principle the mean of the beliefs for each bin need not sum exactly to 1, but in fact they do. Some intuition for this result comes from viewing the estimate for the parameters as unimodal around the means, so that there are many individual sets of estimates of the set $\{s, \eta$ and $\varrho\}$ from the sample of 25,000 that are arbitrarily close to these means.

Individual results for two subjects are presented in Figures 3 and 4, focusing on the comparison of observed reports and recovered beliefs using the posterior mean belief. In each case we display the reports and beliefs in the left panels, and the risk preferences for this subject in the right panels. We also report a simple measure of the difference between reports and recovered beliefs, Ω, given by the absolute value of the differences in reports, expressed as a percent. If beliefs were the same as reports Ω would be 0%, and if the two were completely different Ω would be 100%. However, we know that recovered beliefs cannot arise when there are no tokens allocated to a bin, so the upper bound for Ω is strictly less than 100%. In Figure 3 the subject has the same type of report for each question: 50 tokens allocated to one modal response, and 25 tokens allocated either side of that modal response, so Ω is the same value, 29%, in both cases.

Figure 3 shows a subject that is optimistic with respect to probabilities, leading to decision weights that favor higher value prizes by enhancing their weight above the actual probability. The subject also has a relatively concave utility function, with s = 0.74. The way to understand behavior here is to examine the recovered beliefs, then consider the risk preferences and infer why the subject would then have been induced, given those risk preferences, to make the observed reports. For both questions the subject has beliefs that lead to greater weight being placed on the report for the modal bin, and that weight is enhanced by the decision weight for that prize being view optimistically. The two extreme reports are *not* affected differently by the optimism of the subject, since they have the same weight and hence are not distinct when viewed as monetary rewards if those outcomes are realized. For this reason we display the decisions weights for 2 prizes as well as for 3 prizes: the decisions weights for 2 prizes actually apply here. Then there is an effect from the concave utility function, leading to the individual wanting, *ceteris paribus* any effect of probability weighting, to spread the tokens more equally over all three outcomes that are assigned any weight at all.

In this instance the effect of optimistic probability weighting for the modal belief dominates the effect of utility concavity, so the observed reports are significantly less uniform than the beliefs, placing less weight on the extremes, and greater weight on the modal response. This example demonstrates the "first-order" effect that probability weighting can have on the *distribution* of recovered beliefs, even if it does not change the *average* belief of the subject. Since it is the distribution of this probability mass function that matters for the evaluation of the lottery, under EUT or RDU, it is critical for inference that we correctly recover the distribution and not just the average.

Figure 4 displays the results from a relatively rare case, in which the individual has "inverse-S" probability weighting, leading to decision weights in which the lowest ranked and highest ranked prize are given extra weight. This type of probability weighting occurs, but is not nearly as common as claimed by some; again, see Wilcox [2022]. In any event, this is an interesting case for another reason, since the effect of probability weighting on reports is not particularly pronounced. The subject has a relatively concave utility function, with s = 0.9, and that plays much more of a factor in explaining observed reports. Again, focus on the beliefs, and then knowing these risk preferences, explain the observed reports. In both questions the observed reports are tending to provide less variability in returns across the outcomes that the individual thinks has any chance of occurring.

Figure 5 displays a characterization of the posterior probability of beliefs for the individual in relation to the true outcome of each question. Apart from being interested in how different recovered beliefs can be from observed reports, our ultimate interest in recovering beliefs is to be able to evaluate hypotheses about the accuracy of beliefs. This is what Figure 5 delivers. In each case the vertical axis shows the cumulative posterior probability of the individual having a belief that is equal to or less than the outcome indicated on the horizontal axis. The cumulative posterior mean is displayed by a solid line, and then the 90% credible interval around that cumulative posterior mean is displayed. This credible interval is generated directly from the 25,000 simulated values for the belief for each outcome by that subject. For reference, in each panel of Figure 5 we also display in red the true outcome, allowing an easy evaluation of the accuracy of the beliefs of each individual. For further reference, Figure 5 also displays a dashed line at cumulative posterior values of 0.5, which is one common Bayesian metric for "the weight of the evidence." Of course, these displays allow the use of any such metric.

Consider the subjects in the top row of Figure 5, who are the subjects whose beliefs are reported in detail in Figures 3 and 4. In both cases no report assigned positive density to the true outcomes, and all reports were below those true outcomes: hence the same is true for recovered beliefs, as a matter of theory, irrespective of the risk preferences of the subjects. Subject #4 had beliefs that were close to being uniform over the outcomes that had any positive density. It follows that any imprecision of the risk preference estimates for this subject would have little effect on the imprecision of the recovered beliefs of the subject, and that is what we observe for this subject in Figure 5, with very "thin" credible intervals around the cumulative posterior mean values. By contrast, subject #17 exhibited some variation in the density of beliefs, and we observe much "thicker" credible intervals around the cumulative posterior mean values. Both subjects, then, exhibit clear bias with respect to the true addictiveness of smoking, underestimating the risks of addiction for both questions.

The next two subjects, in the second line of Figure 5, exhibit reasonably accurate beliefs for one or other of the two questions about the addictiveness of smoking, but not both questions. Subject #51 clearly underestimated the number of smokers that relapsed, but was relatively accurate with respect to the number of quit attempts, and *vice versa* for subject #151. Finally, in the third line of Figure 5 we display results for two subjects that had reasonably accurate beliefs for *both* questions: we can count the number of subjects that reasonably met this criteria on one hand.

The information in Figure 5 also allows the construction of more complete Bayesian tests of the hypothesis of biased beliefs, using null hypotheses stated in the form of intervals around some point-null outcome. Armed with the full posterior distribution, one can then directly calculate the fraction of the distribution that falls in or out of that interval. In our two cases, the null of no bias falls on one side of the range of outcomes, so inspection of the cumulative posterior is sufficient. But these methods generalize immediately to more complex null hypotheses.[19]

Detailed results for each of the 383 subjects, such as those shown in Figures 3, 4 and 5, are available to the researcher for use and evaluation.[20] Figure 6 collates all results over all subjects for the two questions. Perhaps remarkably, there is not a major difference between the raw reports and the recovered beliefs *at this level of aggregation*, as shown by the two top panels of Figure 6. What is happening is that differences between reports and recovered beliefs, which can be *severe at the level of the individual*, are tending to cancel out. So if one was just wanting to recover beliefs from individuals to make some statement about average beliefs of the sample, not much is gained from the recovery exercise in these cases, as long as one does not then casually claim that all individuals have the same beliefs as their reports. The bottom panels show that there is considerable heterogeneity in the recovered beliefs at the level of the individual. These panels show the density of deviations Ω from the reports, and for many individuals these deviations are quite large.

Finally, Figure 7 shows the cumulative posterior distribution of beliefs for all subjects, including both 90% and 80% credible intervals. Considering the cumulative posterior mean values, the evidence shows that 80% of the beliefs are less than the true addictiveness for the relapse question, and 93% of the beliefs are less than the true addictiveness for the quit attempts question.[21] The lower bound of the 90% credible intervals indicate that 55% or 75% of the beliefs are less than the true addictiveness, respectively. We stress, again, that the results in Figure 5 vividly illustrate that stronger inferences are

---

[19] This interval is often referred to as a "region of practical equivalence" (ROPE) from a Bayesian perspective: see Kruschke [2017] for an exposition. In our case we could defined a ROPE as any outcome bin either side of the bin containing the true outcome.

[20] The software we provide generates all of these results for every individual, and displays them automatically in a single PDF document for review.

[21] These are exact values, which can be proximately inferred from Figure 7 as well.

available for individuals, and in this qualitative respect the selected individuals whose beliefs are displayed in Figure 5 are representative of the sample.

The results in Figures 5, 6 and 7 do allow us to make some substantive inferences about the complete sample in terms of beliefs about the addictiveness of smoking. The vast majority of individual subjects have beliefs that understate the addictiveness of cigarettes, and we can be confident about having the sign of that bias correct.

4. Conclusion

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 Scoring Rule, but are proven to generalize to the complete 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 theoretical results on the recovery of beliefs map directly and exactly to the estimates of risk preferences obtained from Bayesian methods, which are also ideally suited to estimating risk preferences at the individual level. We have developed user-friendly software to allow these results to be applied with ease and speed. And we provide an application by evaluating two subjective belief questions of substantive importance with respect to the addictiveness of cigarette smoking by teenagers and young adults. Our results allow the easy 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

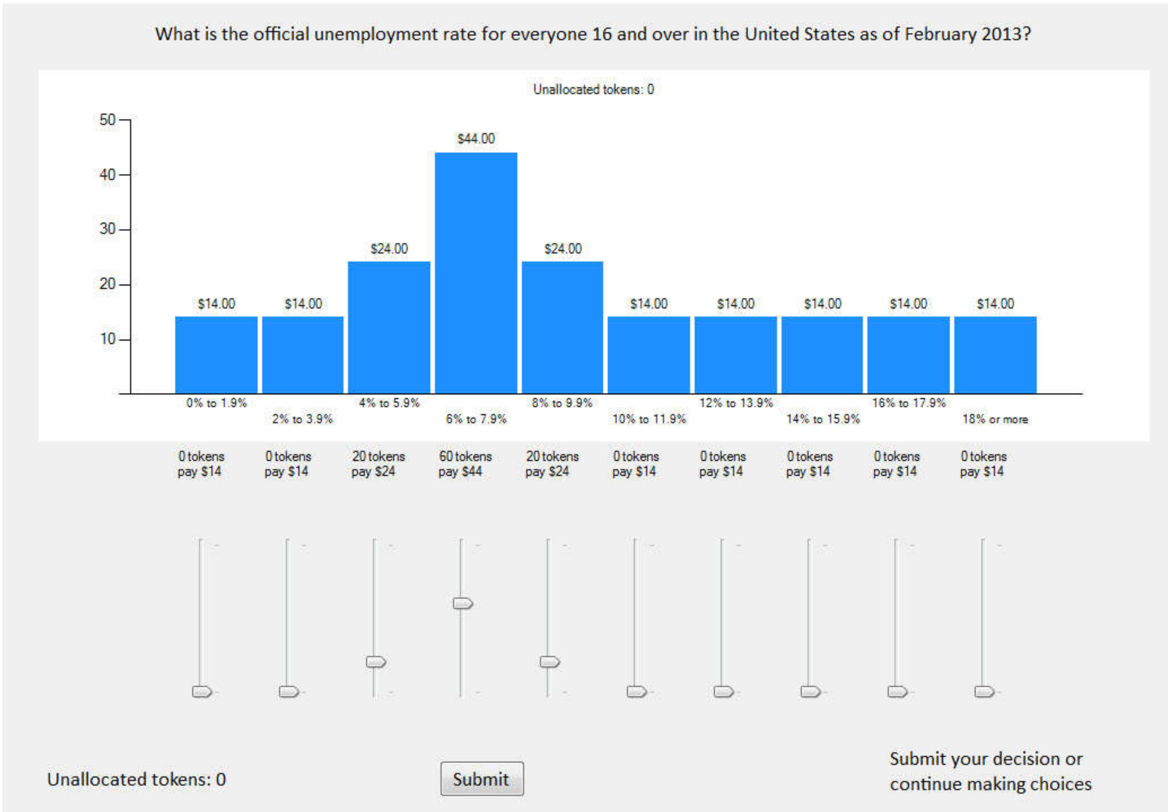


Figure 2: Possible Belief Elicitation Response

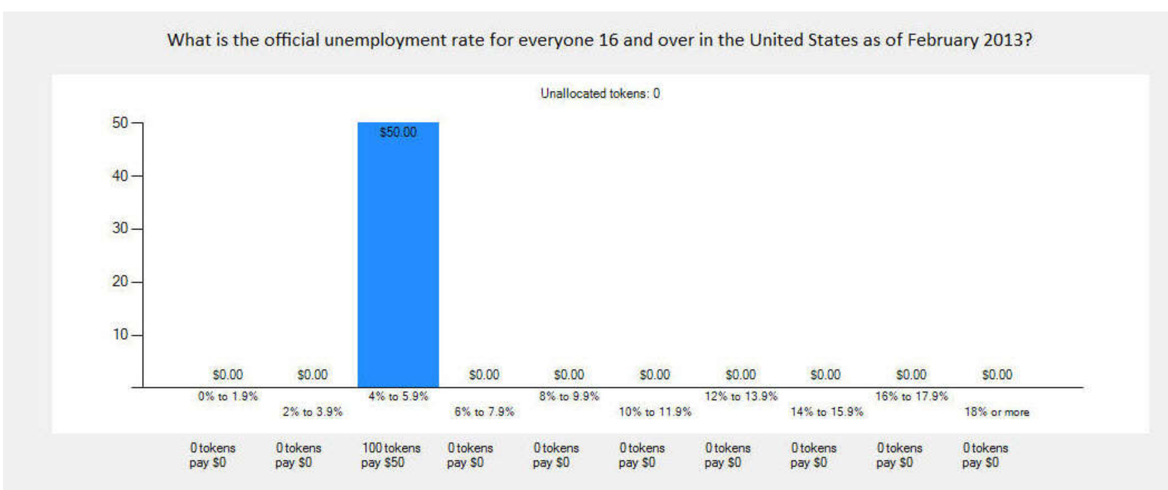


Figure 3: Reports, Beliefs and Risk Preferences for Subject #4

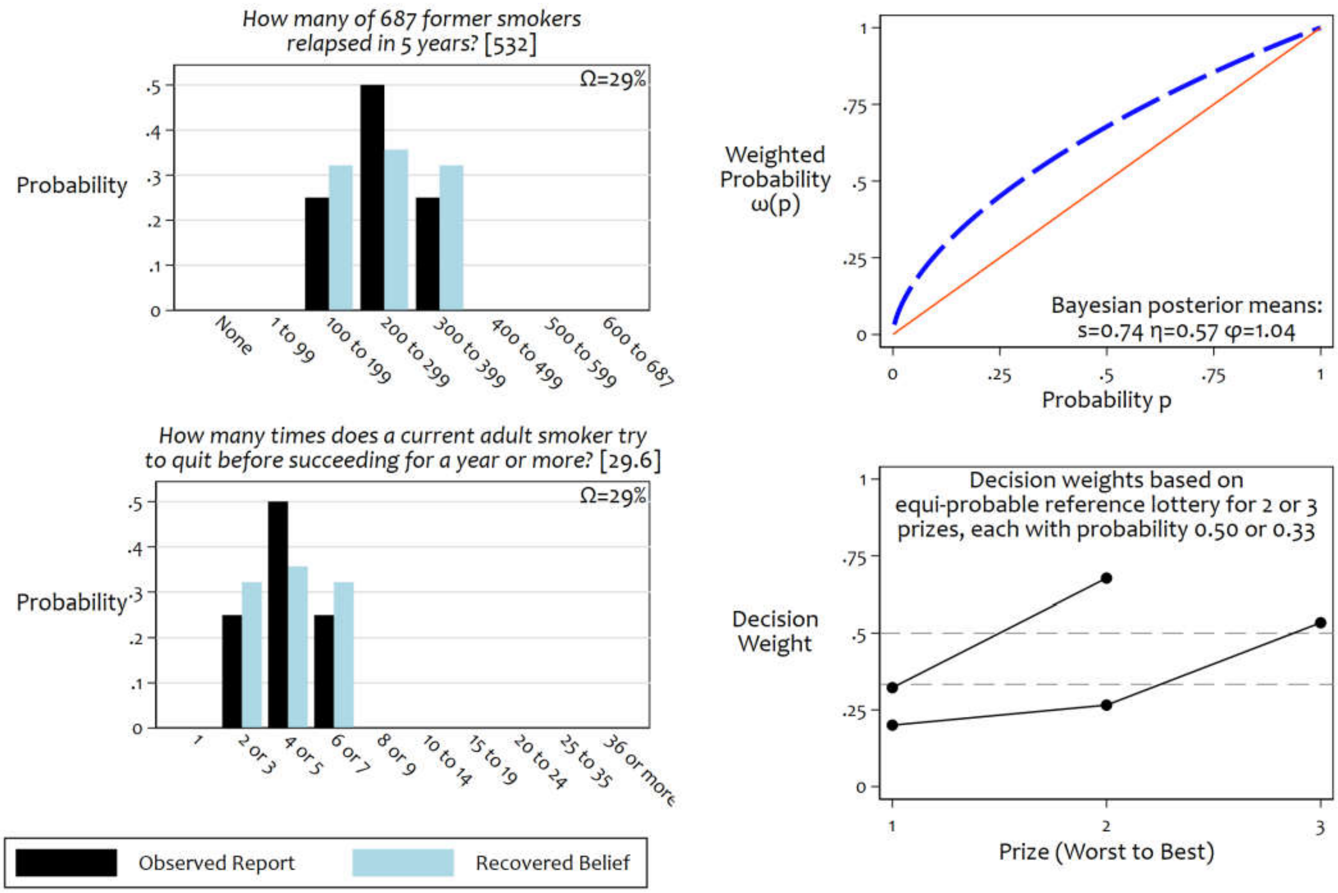


Figure 4: Reports, Beliefs and Risk Preferences for Subject #17

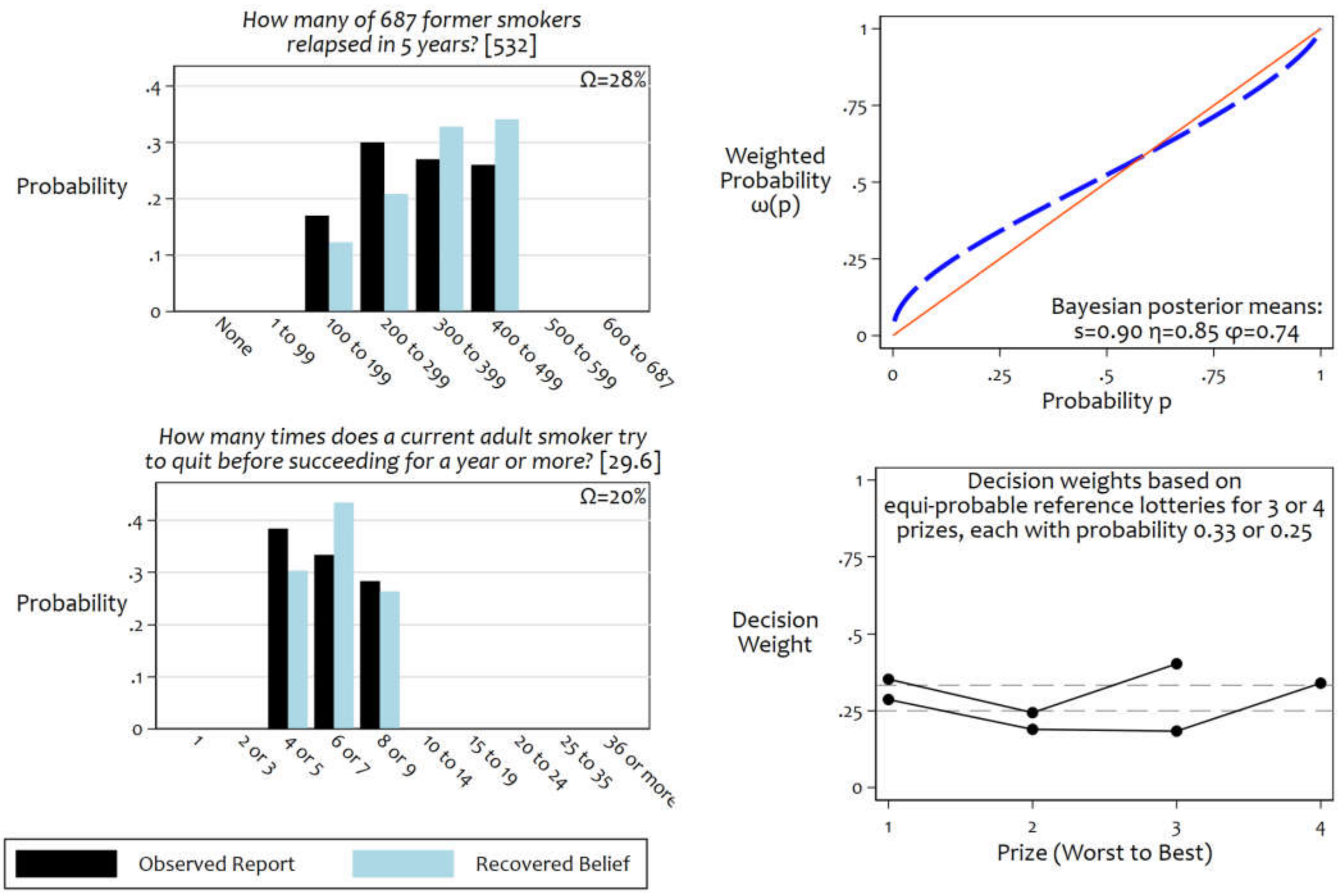


Figure 5: Cumulative Posterior Probability of Beliefs in Relation to True Outcome for Select Subjects

Cumulative Posterior Mean and 90% Posterior Credible Interval

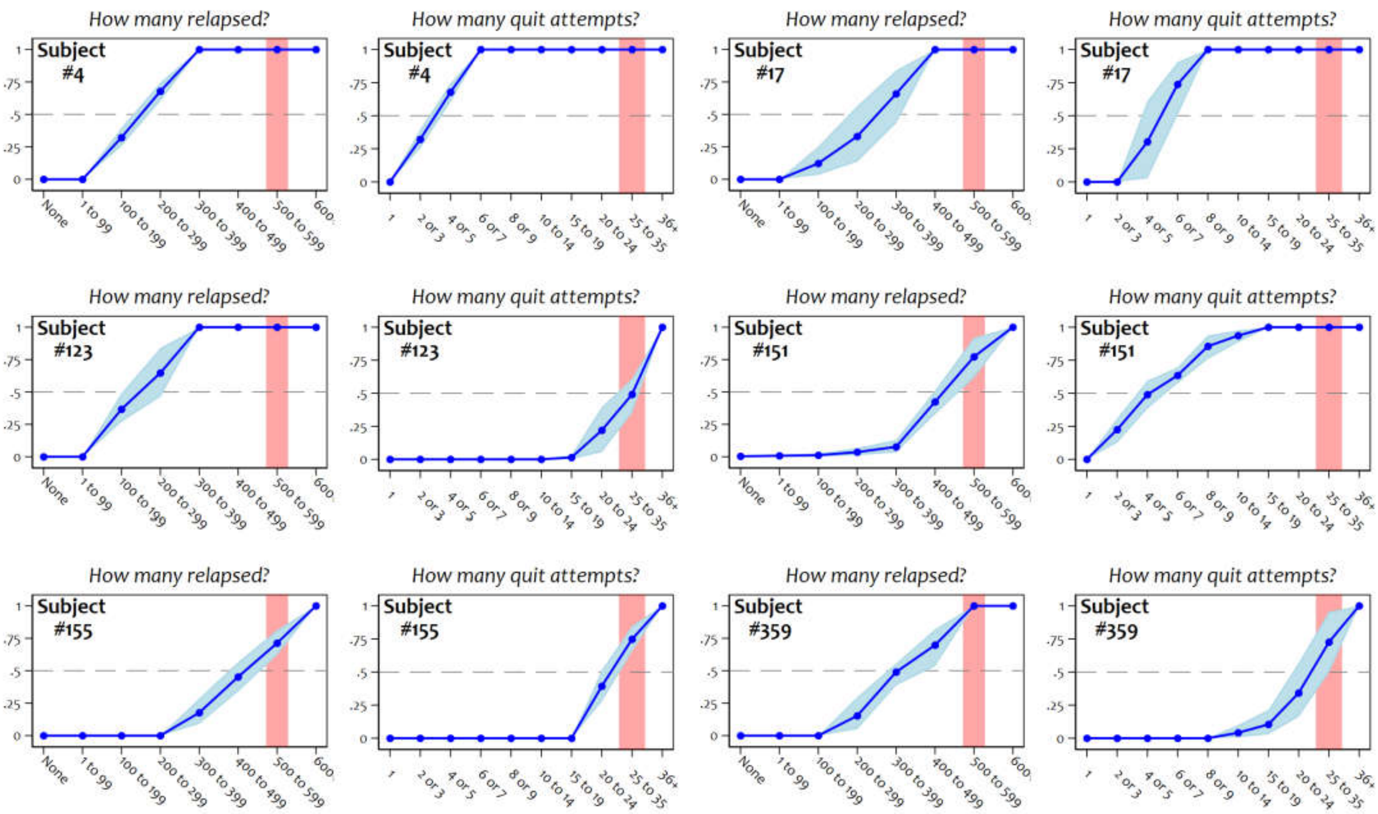


Figure 6: Pooled Reports and Beliefs, and Distribution of Individual Ω Deviations from Reports

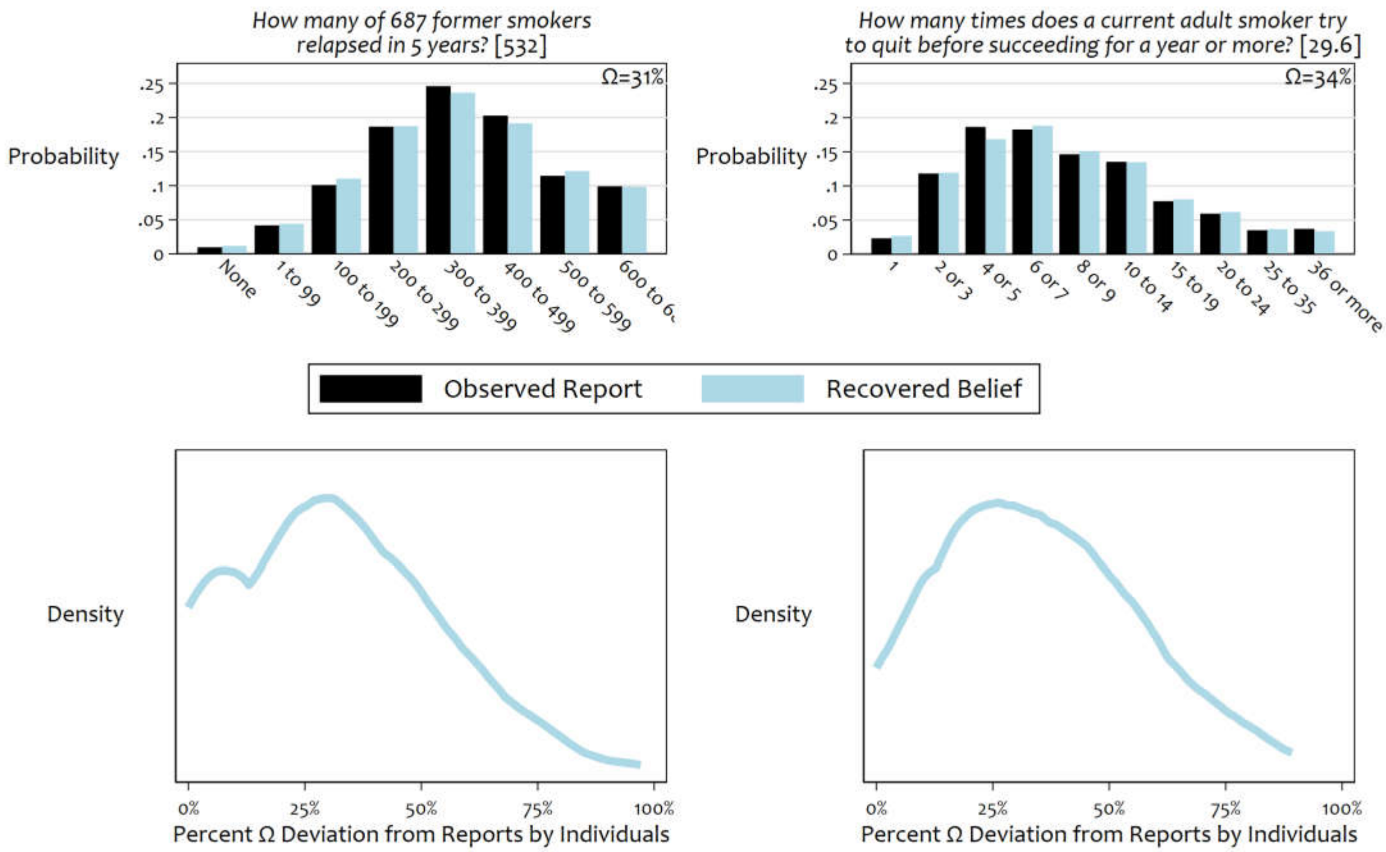
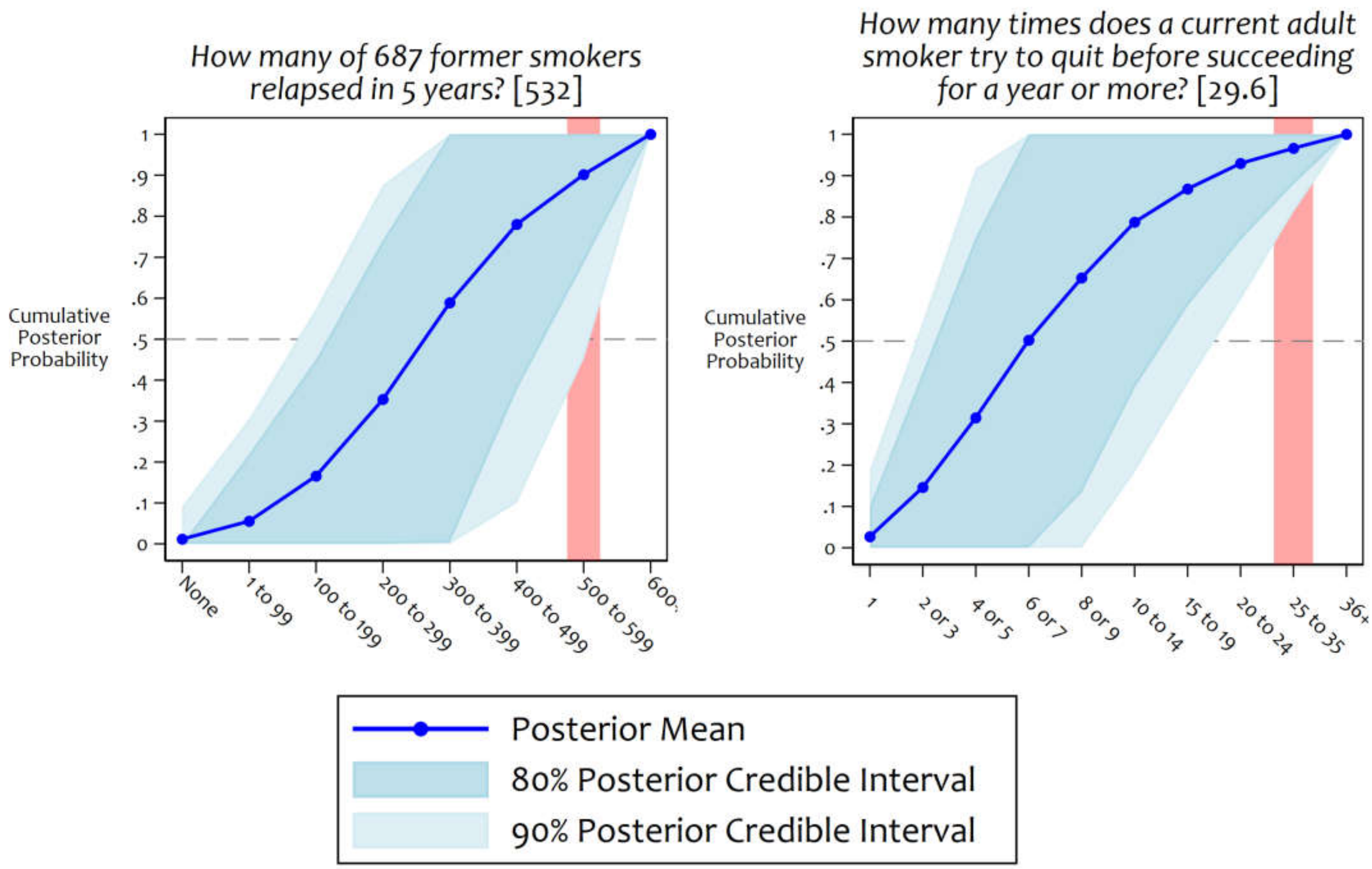


Figure 7: Cumulative Posterior Probability of Beliefs in Relation to True Outcome for All Subjects



References

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

Andersen, Steffen; Hole, Arne Risa; Harrison, Glenn W., Lau, Morten, and Rutström, E. Elisabet, "Non-Linear Mixed Logit," *Theory and Decision*, 73, 2012, 77-96.

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

Becker, Gary S., and Murphy, Kevin M., "A Theory of Rational Addiction," *Journal of Political Economy*, 96, 1988, 675-700.

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

Borland, Ron; Partos, Timea R.; Yong, Hua-Hie; Cummings, K. Michael, and Hyland, Andrew, "How Much Unsuccessful Quitting Activity is Going on Among Adult Smokers? Data from the International Tobacco Control Four Country Cohort Survey," *Addiction*, 107, 2011, 673-682.

Chaiton, Michael; Diemert, Lori; Cohen, Joanna E.; Bondy, Susan J.; Selby, Peter; Philipneri, Anne, and Schwartz, Robert, "Estimating the Number of Quit Attempts it Takes to Quit Successfully in a Longitudinal Cohort of Smokers," *BMJ Open*, 2016;6: e011045. doi:10.1136/bmjopen-2016-011045.

Danz, David; Vesterlund, Lise, and Wilson, Alistair J., "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, 112(9), 2022, 2851-2883.

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

Gao, Xiaoxue Sherry; Harrison, Glenn W., and Tchernis, Rusty, "Behavioral Welfare Economics and Risk Preferences: A Bayesian Approach," *Experimental Economics*, https://doi.org/10.1007/s10683-022-09751-0, forthcoming 2022.

Harrison, Glenn W., "Behavioral Responses to Surveys About Nicotine Dependence," *Health Economics*, 26, 2017, 114-123.

Harrison, Glenn W.; Martínez-Correa, Jimmy, and Swarthout, J. Todd, "Eliciting Subjective Probabilities with Binary Lotteries," *Journal of Economic Behavior and Organization*, 101, May 2014, 128-140.

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.; Martínez-Correa, Jimmy; Swarthout, J. Todd, and Ulm, Eric "Scoring Rules for Subjective Probability Distributions," *Journal of Economic Behavior & Organization*, 134, 2017, 430-448.

Harrison, Glenn W.; Morsink, Karlijn, and Schneider, Mark, "Do No Harm? The Welfare Consequences of Behavioural Interventions," *CEAR Working Paper 2020-12*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2020.

Harrison, Glenn W., and Ng, Jia Min, "Evaluating the Expected Welfare Gain from Insurance," *Journal of Risk and Insurance*, 83(1), 2016, 91-120.

Harrison, Glenn W., and Phillips, Richard D., "Subjective Beliefs and the Statistical Forecasts of Financial Risks: the Chief Risk Officer Project," in T.J. Andersen (ed.) *Contemporary Challenges in Risk Management* (New York, Palgrave Macmillan, 2014).

Hossain, Tanjim and Okui, Ryo, "The Binarized Scoring Rule," *Review of Economic Studies*, 2013, 80, 984-991.

Kruschke, John K., "Rejecting or Accepting Parameter Values in Bayesian Estimation," *Advances in Methods and Practices in Psychological Science*, 2018, 1, :270-280.

Manski, Charles F., "Measuring Expectations," *Econometrica*, 72(5), September 2004, 1329-1376.

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

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

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

Orphanides, Athanasios, and Zervos, David, "Rational Addiction with Learning and Regret," *Journal of Political Economy*, 103, 1995, 739-758

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.

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

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).

Schlag, Karl H., and van der Weele, Joël, "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk-Neutrality," *Theoretical Economics Letters*, 3, 2013, 38-42.

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

Thun, Michael J.; Carter, Brian D.; Feskanich, Diane; Freedman, Neal D.; Prentice, Ross; Lopez, Alan D.; Harge, Patricia, and Gapstur, Susan, "50-Year Trends in Smoking-Related Mortality in the United States," *The New England Journal of Medicine*, 368(4), January 24, 2013, 351-364.

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

Viscusi, W. Kip, *Smoke-Filled Rooms: A Postmortem on the Tobacco Deal* (Chicago: University of Chicago Press, 2002).

Wilcox, Nathaniel T., "Unusual Estimates of Probability Weighting Functions," in G.W. Harrison and D. Ross (eds.), *Models of Risk Preferences: Descriptive and Normative Challenges* (Bingley, UK: Emerald, Research in Experimental Economics, 2022).

Appendix A: Proofs (Online Working Paper)

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:

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

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:

$$\mathbf{Max}_{\{r\}}\; E_{w(p)} [u(\theta)] \text{ subject to } \sum_{i=1...K} (r_i) = 1 \tag{A10}$$

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 (A10) because they are automatically satisfied by the solution (9) for both risk-averse and risk-loving individuals.

Problem (A10) can be solved by maximizing the Lagrangian

$$\mathfrak{L} = \sum_{j=1...K} w(p_j) \times u[\alpha + \beta 2r_j - \beta \sum_{i=1...K} (r_i)^2] - \lambda [\sum_{i=1...K} (r_i) - 1] \tag{A11}$$

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[22]

$$\partial\mathfrak{L}/\partial r_k = \sum_{j=1...K} (w(p_j) \times \partial u(\theta(j))/\partial r_k) - \lambda = 0, \forall\, k = 1,..., K \tag{A12}$$

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 (A11) with respect to the Lagrangian constant:

$$\sum_{i=1...K} (r_i) - 1 = 0. \tag{A13}$$

We can simplify the K equations in (A12) 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

$$w(p_k) \times (\partial u/\partial\theta\,|_{\theta=\theta(k)}) - r_k E_{w(p)} [\partial u/\partial\theta] = \lambda/2\beta, \forall\, k = 1,..., K. \tag{A12$'$}$$

Summing over the K first-order conditions we get

$$E_{w(p)} [\partial u/\partial\theta\,|_{\theta=\theta(k)}] - \sum_{k=1...K} r_k E_{w(p)} [\partial u/\partial\theta] = K\lambda/2\beta . \tag{A14}$$

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 (A13). Then (A14) implies that $K\lambda/2\beta = 0$, which can only be satisfied if $\lambda = 0$ since K>0 and β>0. This result and (A12$'$) implies that the solution to problem (A10) must satisfy the following K conditions:

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

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

[22] 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.

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 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 $r_n = 1 - \sum_{i=1...n-1} r_1$. 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).

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

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

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

This gives us

$$\begin{aligned}&- \sum_{i=1...n-1} q_i\ \sum_{j=1...n-1} r_j\, \partial^2 g/\partial r_j\,\partial r_k + \sum_{i=1...n-1} q_i\, \partial^2 g/\partial r_i\,\partial r_k - \sum_{j=1...n-1} r_j\, \partial^2 g/\partial r_j\,\partial r_k \\ &+ \sum_{i=1...n-1} q_i\ \sum_{j=1...n-1} r_j\, \partial^2 g/\partial r_j\,\partial r_k = 0.\end{aligned}$$

Cancelling terms, we obtain

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

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

$$\sum_{i=1...n-1} (q_i - r_i)\, \partial^2 g/\partial r_i\,\partial r_k = 0. \qquad \text{(A16)}$$

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.

Proof: 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

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

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

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

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

$$\begin{aligned}\partial V/\partial r_k &= \sum_{j=1...n} w(p_j)\, \partial u/\partial W|_{s_j}\, \partial S_n/\partial r_k + \sum_{j=1...n-1} w(p_j)\, \partial u/\partial W|_{s_j}\, \partial^2 g/\partial r_j\, \partial r_k = 0\\ &= \partial S_n/\partial r_k \sum_{j=1...n} w(p_j)\, \partial u/\partial W|_{s_j} + \sum_{j=1...n-1} w(p_j)\, \partial u/\partial W|_{s_j}\, \partial^2 g/\partial r_j\, \partial r_k = 0\\ &= \partial S_n/\partial r_k\, E_{w(p)}[\partial u/\partial W] + \sum_{j=1...n-1} w(p_j)\, \partial u/\partial W|_{s_j}\, \partial^2 g/\partial r_j\, \partial r_k = 0\end{aligned}$$

where $E_{w(p)}[\,.\,]$ denotes the expectations operator under probability measure $w(p) = \{w(p_1), ... , w(p_n)\}$.

Now, since $S_n = g - \sum_{j=1...n-1} r_j\, \partial g/\partial r_j$, we get

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

so

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

Therefore, we obtain

$$\sum_{j=1...n-1} [\, w(p_j)\, \partial u/\partial W|_{s_j} - r_j\, E_{w(p)}[\partial u/\partial W]]\, \partial^2 g/\partial r_j\, \partial r_k = 0. \qquad \text{(A17)}$$

Equation (A17) looks just like equation (A16) 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. This also means that Propositions 1 through 7 of Harrison, Martínez-Correa, Swarthout and Ulm [2017], that characterize the beliefs recovered for an SEU decision-maker, also generalize.

Appendix B: Software (Online Working Paper)

The procedures for recovering subjective belief distributions described in the text have been implemented in a flexible *Stata* program that is easy to use. The data inputs consist of two data files.

- A data file containing the M samples from a Bayesian posterior distribution of estimates from an EUT or RDU model. The data file must have an ID for the subject, and estimates of parameters for the specifications described below.
- A data file containing the reports that each subject ID provided for a belief question defined over K bins.

The program that undertakes the calculations is called **beliefs_recovery.do**, and is called from a shell program that constructs the data file for each subject and question, writes those data files to a sub-directory (created by the program if it did not already exit) called **recovery_room**, processes all files in that sub-directory and/or just one specified file, writes the results to a sub-directory (also created by the program if it did not already exist) called **recovered**, and optionally deletes all of the intermediate files in the **recovery_room** directory.

To illustrate we show the shell program **Recover Numerical Example of Recovery.do** that replicates the numerical example in the text. This program has M=1 and K=10, and just one subject. The part of the code in bold is specific to this toy application in terms of reading in the data:

```
/* Recovers beliefs from numerical example in text */

* log file
capture log close _all
log using "Recover Numerical Example of Recovery.log", replace name(recover)

* configure Stata
capture version 16.1
capture: version 17
set processors 4
set more off
set scheme s1color
set seed 987654321
timer clear 1
timer on 1

* graphics font
graph set window fontface "Candara"

* tell us what version ran
about

* do all the files in the dta_dir? (yes or no)
global do_all "no"

* do only the dta_file? if so, next list the name of the file, to get it done (yes or no)
global do_specific "yes"

* name of specific input and output file, only relevant if $do_specific is "yes"
global use_specific   "example"
global save_specific  "example_b"

* erase recovery_room files at the beginning and end?
global clean_room "no"

* create directories needed
capture: mkdir figures
capture: mkdir recovery_room
capture: mkdir recovered

* clean up
```

```
if "$clean_room" == "yes" {
	cd recovery_room
	local dfiles : dir "$dta_dir" files "*dta"
	foreach file of local dfiles {
		erase "`file'"
	}
	cd ..
}

* load the code to recover beliefs
qui: do belief_recovery

* quiet, to speed up
qui {

* install the BHM estimates
set obs 1

	generate int sid = 1
	generate r = 0.77
	generate gamma = 0.5

	summ r
	local mcmc = r(N)

save tmp_bhm, replace
noi: di "Saved posterior with risk preferences ..."

* install the beliefs data
clear
set obs 1

	generate int sid = 1
	generate int period = 1

	generate int r_3 = 30
	generate int r_4 = 45
	generate int r_5 = 25
	foreach x in 1 2 6 7 8 9 10 {
		generate r_`x' = 0
	}

	generate int nbin = 10
	generate int ntokens = 100
	generate int alpha = 25
	generate int beta  = 25

	summ sid
	local Nsub = r(max)


* QSR parameters specified as a global
foreach x in nbin ntokens alpha beta {
	summ `x'
	global `x' = r(mean)
}
di $nbin

* save data
keep sid period r_*
compress
sort sid period
save tmp_sid, replace
noi: di "Saved beliefs data, and now generating recovery files for each subject ..."

* generate for each subject
summ period
local np = r(max)
forvalues s=1/`Nsub' {
	forvalues p=1/`np' {
		use tmp_sid, clear
		keep if period==`p' & sid==`s'
		save tmp_sid_p, replace
```

```
		use tmp_bhm, clear
		capture: drop _log*
		keep if sid == `s'
		capture: generate eta = .
		capture: generate phi = .
		capture: generate mu = .
		generate int period = `p'
		sort sid period
		merge m:1 sid period using tmp_sid_p
		drop _merge
		generate long mcmc = `mcmc'
		compress
		save recovery_room/example, replace
		des
		summ
	}
}

* recover beliefs
timer on 2

* CRRA, CARA, Power, and Expo-Power utility functions supported, the last one standing is the one
used
global ufunc "cara"
global ufunc "expo"
global ufunc "power"
global ufunc "crra"

* Four probability weighting functions are supported, the last one standing in the next few lines
is the one used if RDU is used
global pfunc "eut"
global pfunc "power"
global pfunc "prelec2"
global pfunc "inverse_s"

* control the names of the parameters for the utility function
local upars "r"
if "$ufunc" == "expo" {
	local upars "r alpha"
}

* control the names of the parameters for the probability weighting function
local ppars ""
if "$pfunc" == "power" | "$pfunc" == "inverse_s" {
	local ppars "gamma"
}
if "$pfunc" == "prelec2" {
	local ppars "phi eta"
}

* do the belief recovery
noi: di "Recovering beliefs ..."

	if "$do_all" == "yes" {

		* subdirectory with data files
		cd recovery_room

		* use extended functions to get all the files in some data folder
		local dfiles : dir "$dta_dir" files "*dta"
		foreach file of local dfiles {

			* load the file
			use "`file'", clear

			* belief recovery
			beliefs_recovery `upars' `ppars'

			* save the output file in the current directory
			cd ..
			save "recovered/`file'", replace
			cd recovery_room

		}
		cd ..
```

```
	}

	if "$do_specific" == "yes" {

		* subdirectory with data files
		cd recovery_room

		use "$use_specific", clear
		beliefs_recovery `upars' `ppars'
		summ
		cd ..
		save "recovered/$save_specific", replace
		noi: list, noobs

	}


* clean up
if "$clean_room" == "yes" {
	noi: di "Cleaning up the recovery_room files ..."
	cd recovery_room
	local dfiles : dir "$dta_dir" files "*dta"
	foreach file of local dfiles {
		erase "`file'"
	}
	cd ..
}
capture: erase tmp.dta
capture: erase tmp_bhm.dta
capture: erase tmp_sid.dta
capture: erase tmp_sid_p.dta

* end of quietly
}

* time taken overall
timer off 2
timer list
local secs = r(t2)
local mins = `secs'/60
local hrs = `mins'/60
local secs_ = string(`secs', "%10.0f")
local mins_ = string(`mins', "%4.1f")
local hrs_ = string(`hrs', "%4.2f")
di "Belief recovery calculations by themselves took `secs_' seconds, `mins_' minutes, or `hrs_'
hours."

* time taken overall
timer off 1
timer list
local secs = r(t1)
local mins = `secs'/60
local hrs = `mins'/60
local secs_ = string(`secs', "%10.0f")
local mins_ = string(`mins', "%4.1f")
local hrs_ = string(`hrs', "%4.2f")
di "Complete calculations took `secs_' seconds, `mins_' minutes, or `hrs_' hours."

log close recover
```

There are several options available:

- The program can process a large number of files if the global **$do_all** is set to "yes" and we illustrate this option next. It can also process a single file if the global **$do_selected** is set to "yes" and this option is used here for the toy example.
- If the program is to process a single file the name of the input and output file must be specified.
- Optionally the **recovery_room** files can be erased after being processed by means of the global **$clean_room**. For a large number of files this is usually set to "yes" since these are

intermediate files. For the toy example we want to preserve the input file we generate, so set this to “no”.

- Then the data must be read in, as shown in bold. For the toy example we specify the parameter values we have assumed. These are saved in the same format that we use when we read in Bayesian posterior estimates, as in the next example. In some cases we refer to each belief question as a **period** and sometimes as a **question**, but that is easy to modify.
- If demographic or treatment information is to be read in, it should be saved to the beliefs data file **tmp_sid.dta**, and will be carried along. This might be useful to add variables such as a string label to identify the belief question (we use **qid** for that) or demographics such as gender. These variables will then be saved to the output file. We prefer to merge these variables in later, and keep the **tmp_sid.dta** file as lean as possible.
- The globals **$ufunc** and **$pfunc** document which model of risk preferences is to be used. In each case there are four options, and the “last one standing” is used. So in the above example we used the **crra** utility function and the **inverse_s** probability weighting function, as explained in the main text. The command files ending in EUT with CARA, EUT with CRRA, EUT with EP and RDU with Power illustrate some obvious variants. For us the default with real examples is RDU with the CRRA utility functon and the Prelec probability weighting function, used in the next example.
- The rest of the program is reasonably automatic, although one should py attention to whether the question has been defined by the variable **period** or by the variable **question** (or some other numeric variable). The program assumed that the variables identifying the subject (**sid** here) and question (**period** here) are numeric integers defined sequentially with no skips. This allows the program to write the intermediate files to the **recovery_room** directly in loops.

The program generates this input file for the beliefs recovery calculations, with observed reports from the subject in variables **r_1**, ..., **r10**:

sid	r	gamma	eta	phi	mu	period	r_3	r_4	r_5	r_1	r_2	r_6	r_7	r_8	r_9	r_10	alpha	beta	mcmc
1	.77	.5	.	.	.	1	30	45	25	0	0	0	0	0	0	0	25	25	1

When executed the commands are minimal, and then in this instance, since it is just one line, the output file is displayed:

```
Saved posterior with risk preferences ...
Saved beliefs data, and now generating recovery files for each subject ...
Recovering beliefs ...
```

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_10	sid	r	gamma	eta	phi	mu	period	alpha	beta	mcmc	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_10
0	0	.3	.45	.25	0	0	0	0	0	1	.77	.5	.	.	.	1	25	25	1	0	0	.1791209	.8023131	.01856601	0	0	0	0	0

Although not in the same order, the output file contains the same variables and data as the input file but just adds the recovered beliefs in variables **b_1**, ... , **b_10**.

The command file for a real example is very similar, apart from the steps in bold creating hte data. Here is the first part of the file **GSU Risk and Beliefs Data on Smoking 2019.do** which process the 10-bin questions on smoking addictiveness we report in the main text:

```
/* Recovers beliefs from GSU risk and beliefs data from the 2019 Smoking questions (DNH Sessions) */

* log file
capture log close _all
log using "GSU Risk and Beliefs Data on Smoking 2019.log", replace name(recover)

* configure Stata
capture version 16.1
capture: version 17
set processors 4
set more off
set scheme s1color
set seed 987654321
timer clear 1
timer on 1

* graphics font
graph set window fontface "Garamond"
graph set window fontface "Candara"

* tell us what version ran
about

* do all the files in the dta_dir? (yes or no)
global do_all "yes"

* do only the dta_file? if so, next list the name of the file, to get it done (yes or no)
global do_specific "yes"

* name of specific input and output file, only relevant if $do_specific is "yes"
global use_specific   "smoking_s1_q1"
global save_specific  "smoked_s1_q1"

* erase recovery_room files at the beginning and end?
global clean_room "yes"

* create directories needed
capture: mkdir figures
capture: mkdir recovery_room
capture: mkdir recovered

* clean up
if "$clean_room" == "yes" {
	cd recovery_room
	local dfiles : dir "$dta_dir" files "*dta"
	foreach file of local dfiles {
		erase "`file'"
	}
	cd ..
}

* load the code to recover beliefs
qui: do belief_recovery

* quiet, to speed up
qui {

* decide which risk model to use (eut or rdu)
local risk_model "rdu2019"

* collate the BHM estimates with beliefs
use bhm_`risk_model'_posterior
summ _loglikelihood if id==1
local mcmc = r(N)
rename _loglikelihood ll
egen int sid = group(id)
save tmp_bhm, replace
noi: di "Saved posterior with risk preferences ..."

* restart with main data containing the beliefs data
use ii_beliefs_recovery, clear
keep if belief_task==1
keep LinkID qid choice* v_* bquestion nbin alpha beta
keep if substr(qid,1,4) == "Smok"
```

```
tab qid
generate int question = 0
local xx = 0
generate int keep = 0
foreach x in 1_10 2_10 4_10 5_10 {
        summ choice* if qid == "Smoking_`x'bin"
        replace keep = 1 if qid == "Smoking_`x'bin"
}
keep if keep == 1
foreach x in 1_10 2_10 3_8 4_10 5_10 {
        local xx = `xx'+1
        replace question = `xx' if qid == "Smoking_`x'bin"
}
tab question

* all questions with 10 bins had 100 tokens to allocate
generate int ntokens = 100

egen int id = group(LinkID)
egen int sid = group(id)
summ sid
local Nsub = r(max)

* reports
forvalues x=1/10 {
        generate int r_`x' = ntokens*choice`x'
}

* QSR parameters specified as a global
foreach x in nbin ntokens alpha beta {
        summ `x'
        global `x' = r(mean)
}
di $nbin

* save data
keep sid question qid r_*
compress
sort sid qid
save tmp_sid, replace
noi: di "Saved beliefs data, and now generating recovery files for each subject ..."

* generate for each subject
summ question
local nq = r(max)
forvalues s=1/`Nsub' {
        forvalues q=1/`nq' {
                use tmp_sid, clear
                keep if question==`q' & sid==`s'
                save tmp_sid_p, replace
                use tmp_bhm, clear
                drop _log*
                keep if sid == `s'
                capture: generate eta = .
                capture: generate phi = .
                capture: generate mu = .
                generate int question = `q'
                sort sid question
                merge m:1 sid question using tmp_sid_p
                drop _merge
                generate long mcmc = `mcmc'
                compress
                save recovery_room/smoking_s`s'_q`q', replace
                des
                tab qid
                summ
        }
}
```

The particularly relevant code is again in bold:

- We specify with the global **$do_all** that all of the files generated in the **recovery_room** are to be processed. We have N = 383 subjects answering 4 questions in this case (we process the one 8-bin question separately). We prefer to generate 1532 data files, one for each subject and question, and then have all processed in sequence, since it makes it easier "downstream" for us to collate the results – we prefer lots of smaller data files, easily read with loops, and there is no significant computational time cost involved. Alternatively one could process these beliefs as one data file containing all 1532 records, for some computational time cost savings.
- The Bayesian posterior distribution is read in from a pre-existing data file. Invariably these are long, "skinny" data files. In our case there are M = 25,000 samples for each of N = 383 subjects. In fact, the sessions from which these belief questions were drawn had 614 subjects, only 383 of which we presented these belief questions to. So the data file contains 15,350,000 rows of data for all 614 subjects. These data files normally have a standard set of variables: the parameters estimated, a log-likelihood value, and a posterior density value. We retain the log-likelihood values so that we can make inferences about what the recovered beliefs would have been if one had employed the maximum-likelihood estimate; in general this is of no interest. Since MCMC samples often repeat, for efficiency these files are often saved with an integer variable keeping track of the frequency of the sample: we expand such data sets to include duplicates, so we do not have to re-weight estimates.
- The beliefs data is also from a pre-existing data file, matched in terms of an ID variable to the Bayesian posterior data of course. It contains the reports of each subject to each of the questions. In this case the format of the data may need some modest manipulations to only contain the information needed, and this type of specificity varies from application to application.
- Once the **tmp_bhm.dta** has been collated for the BHM estimates and the **tmp_sid.dta** collaed for the belief task reports, the program loops through each subject and question to generate data files for processing. This is the step that could be replaced by a single "merge" if one wanted to run all subjects and questions in one file. The lines in this subject-and-question-specific merge process that should be watched are in bold. In particular, the data files are given some name, in this case the text **smoking** and a reference to the subject and question, that does not overwrite other files. So the responses of subject 6 to question 5 would be in data file **smoking_s6_q5.dta**, and that same name would be in the **recovered** directory once beliefs are recovered.
- For quick debugging we turn off the **$do_all global** and just process one file with the **$do_specific** global, then turn the **$do_all** global back on. For the example illustrated here, with 383 subjects and 4 questions, and the RDU model with Prelec preferences, the belief recovery phase required 43 minutes on a laptop. The overall program required 110 minutes, due to the extra time needed to generate the 1532 data files and 1532 results files, since each file contains 25,000 records. For a calculation of this magnitude, that is fast enough. We know of ways to speed this up, but see no point.

Appendix C: Instructions (Online Working Paper)

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. You will be presented with a series of pairs of prospects where you will choose one of them. There are 100 pairs in the series. For each pair of prospects, you should choose the prospect you prefer to play. You will actually get the chance to play one of the prospects you choose, 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 such 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.

You will be told your cash endowment for each lottery at the top of the lottery. In this example it is $35, so any earnings would be added to or subtracted from this endowment. The endowment may change from choice to choice, so be sure to pay attention to it. The endowment you are shown only applies for that choice.

In the above example the left prospect pays twenty-five dollars ($25) if the number drawn is between 1 and 5, and pays negative five dollars ($-5) if the number is between 6 and 55, and pays negative thirty-five dollars ($-35) if the number is between 56 and 100. The blue color in the pie chart corresponds to 5% of the area and illustrates the chances that the number drawn will be between 1 and 5 and your prize will be $25. The orange area in the pie chart corresponds to 50% of the area and illustrates the chances that the number drawn will be between 6 and 55 and your prize will be $-5. The green area in the pie chart corresponds to 45% of the area and illustrates the chances that the number drawn will be between 56 and 100. When you select the lottery to be played out the computer will tell you what die throws translate into what prize.

Now look at the pie in the chart on the right. It pays twenty-five dollars ($25) if the number drawn is between 1 and 15, negative five dollars ($-5) if the number is between 16 and 25, and negative thirty-five dollars ($-35) if the number is between 26 and 100. As with the prospect on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $25 pie slice is 15% of the total pie.

Even though the screen says that you might win a negative amount, this is actually a loss to be deducted from your endowment. So if you "win" $-5, your earnings would be $30 = $35 - $5.

Each pair of prospects is shown on a separate screen on the computer. On each screen, you should indicate which prospect you prefer to play 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. You will then roll two ten-sided die to determine which pair of the 100 prospects you chose will be played out:. Since there is a chance that any of your 100 choices could be played out for real, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will roll the two ten-sided dice again to determine the outcome of the prospect you chose.

For instance, suppose you picked the prospect on the right in the above example. If the random number was 7, you would win $25 in addition to your endowment; if it was 93, you would lose $35 from your endowment. If you picked the prospect on the left and drew the number 7, you would lose $5 from your endowment; if it was 93, you would again lose $35 from your endowment.

Therefore, your payoff is determined by three things:

- by which prospect you selected, the left or the right, for each of these 100 pairs;

- by which prospect pair is chosen to be played out in the series of 100 such pairs using the two ten-sided die; and
- by the outcome of that prospect when you roll the two 10-sided dice.

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

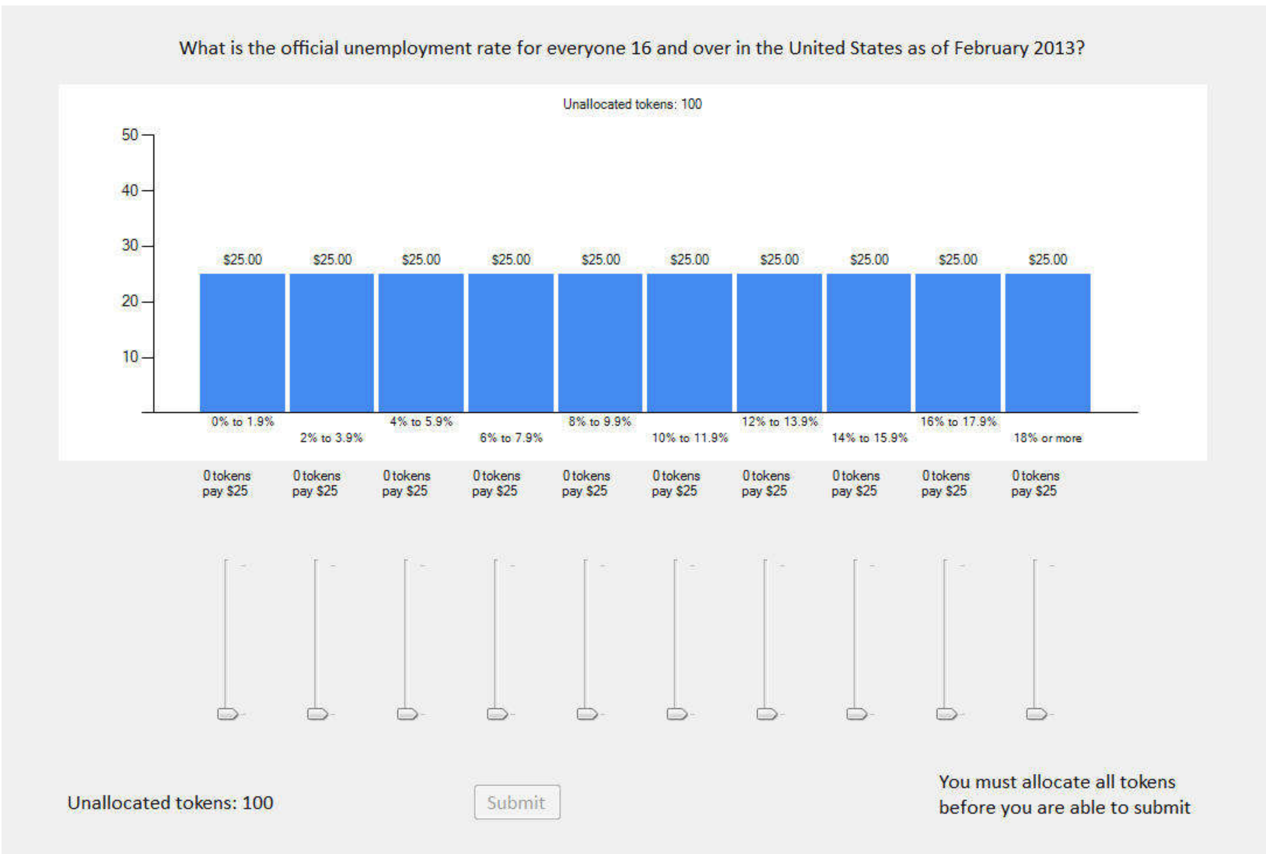
All payoffs are in cash, and are in addition to the show-up fee that you receive just for being here, as well as any other earnings in other tasks.

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.

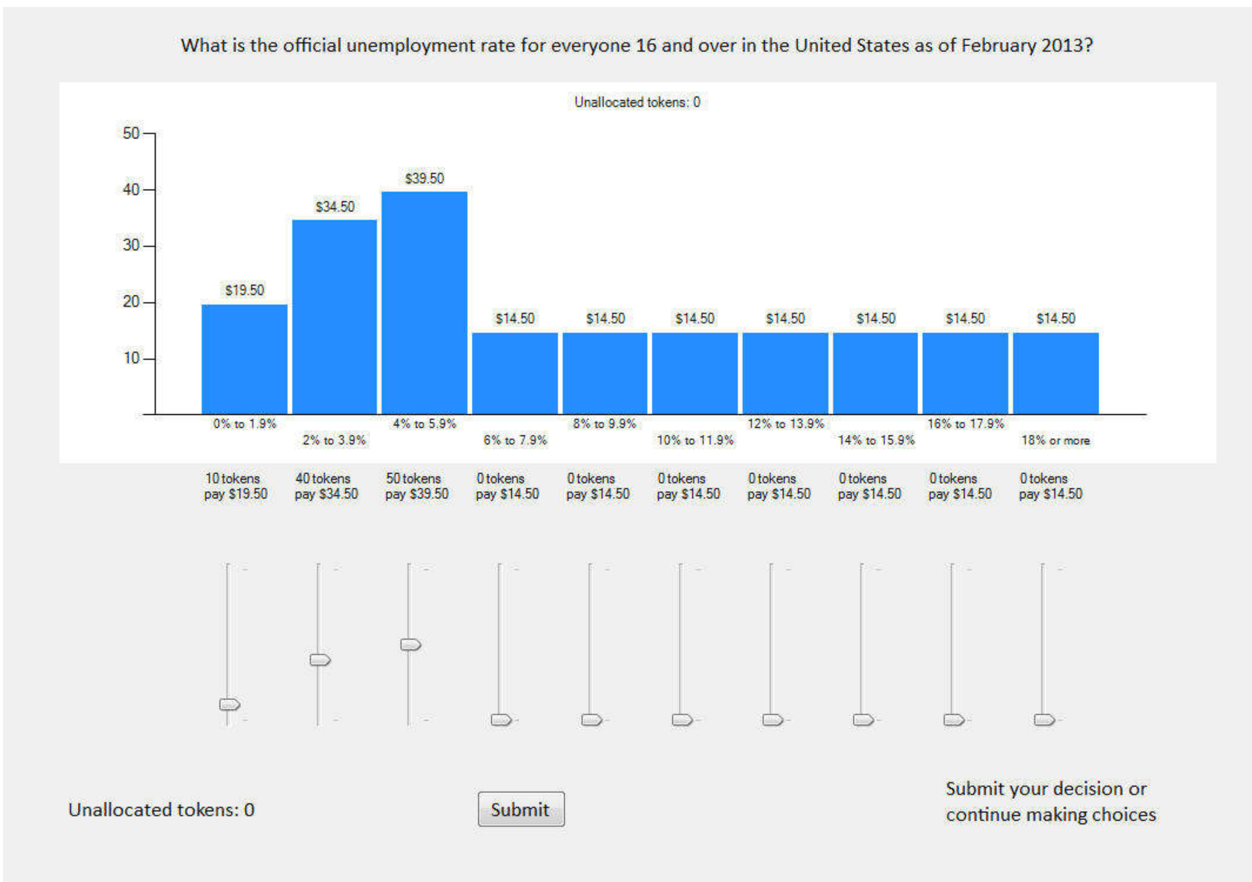


The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens, 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.

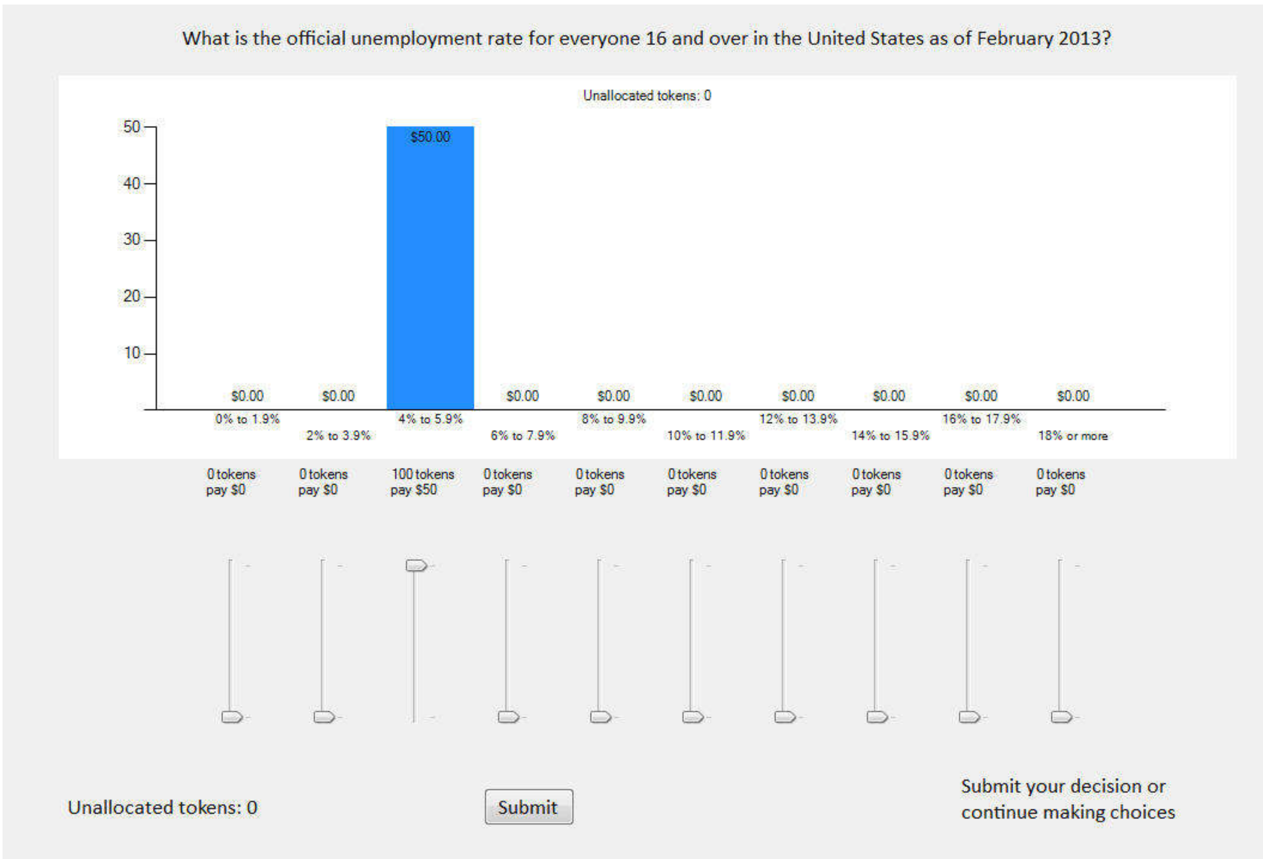


You can adjust the allocation as much as you want to best reflect your personal beliefs about the unemployment rate.

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?