Experimental Economics (2026), 1–25
doi:10.1017/eec.2026.10057

economic science association

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL PAPER

Recovering subjective probability distributions: a Bayesian approach

Glenn W. Harrison[1], Brian Albert Monroe[1] and Eric R. Ulm[2]

[1]Maurice R. Greenberg School of Risk Science, J. Mack Robinson College of Business, Georgia State University, Atlanta, US
[2]School of Economics and Finance, Victoria University of Wellington, Wellington, New Zealand
Corresponding author: Glenn W. Harrison; Email: gharrison@gsu.edu

(Received 15 July 2024 UTC; revised 9 January 2026 UTC; accepted 10 July 2026 UTC)

Abstract

An individual reports subjective beliefs over continuous events using a proper scoring rule, such as the Quadratic Scoring Rule. Under mild additional assumptions, 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 "competitiveness" of individuals and the addictiveness of smoking. These results significantly widen the domain of applicability of proper scoring rules for eliciting latent subjective belief distributions.

Keywords: subjective belief distributions; scoring rules; experimental methods; Bayesian inference

JEL codes: D81; D83; C11; C81; C91

An individual reports subjective beliefs over continuous events using a proper scoring rule, such as the popular Quadratic Scoring Rule. Under some mild additional assumptions, 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 et al. (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. All of these results extend to subjective belief distributions defined over discrete events, including the special case of binary events.

We extend these theoretical results, show how they can be operationalized by using Bayesian estimates of the risk preferences of individuals and incentivized reports about beliefs, and illustrate with two applications.

We extend the 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," under some conditions, 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

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

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. Third, we generalize these results to the complete class of proper scoring rules, of which the Quadratic Scoring Rule (QSR) is 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 for statistical imprecision in the estimates.[1] To account for this statistical imprecision when recovering beliefs one approach is to utilize Bayesian econometric methods. One reason this approach is attractive is that the posterior distribution from Bayesian inference invariably takes the form of a simulated set of samples using Markov Chain Monte Carlo (MCMC) algorithms,[2] and our theoretical results apply *exactly* to each of those 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 which we provide. Another reason it is attractive to use Bayesian methods is that we want to recover beliefs at the level of the individual, to respect the heterogeneity of individual subjective beliefs and individual risk preferences, and Bayesian Hierarchical Models (BHM) provide an attractive way to estimate risk preferences at the level of the individual.

We demonstrate the application of these theoretical results by recovering the latent subjective belief distributions from observed reports by individuals in two incentivized experiments, and for whom we also have individual estimates of their risk preferences using a BHM. One experiment is drawn from a typical example from experimental economics, where subjective beliefs play a conditioning role in understanding behavior. The other experiment is drawn from health economics, where we have direct interest in the subjective beliefs of individuals over the addictiveness of tobacco. 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. Our methods help make inference about belief distributions operational by recovering latent beliefs when one cannot be certain that the individual is risk neutral.

Section 1 reviews different approaches to eliciting subjective belief distributions, to be able to see how our approach complements alternative approaches. Section 2 provides the new theoretical results for belief recovery under RDU, extending our own previous theoretical results for belief recovery under EUT. Section 3 explains how one can generate a posterior distribution for the risk preferences of individuals, and how that applies directly to our theoretical results, and allows them to be applied. Section 4 describes our experimental procedures. Section 5 provides two applications of our approach. One is recovering belief distributions about the competitiveness of men and women, a

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

[2]We estimated all models using *Stan*, which employs an efficient sampler derived from the Hamiltonian Markov Chain algorithm: see Stan Development Team (2025), Kruschke (2015) and McElreath (2020).

[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 et al. (2012) use random coefficient methods to simulate risk preferences for *pooled* samples. The Bayesian approach is appropriate 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 et al. (2023). Here the application is *descriptive*, closer to the joint estimation of risk preferences and subjective probabilities for binary events of Andersen et al. (2014) using classical econometric methods.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

familiar application for experimental economists. The other is to directly elicit beliefs about the addictiveness of smoking, a familiar context for health economists. Online appendices provide additional proofs of formal results, instructions for the experiments, documentation of computer software we provide to undertake these calculations, and documentation of all data.

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.[5] There are three ways to elicit beliefs in economics, each with strengths and weaknesses for different applications.

Following Manski (2004), 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, easy to explain to subjects, easy 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 et al. (2011).

The disadvantage of this approach is that the results could exhibit hypothetical bias, and it is easy to document that this bias can be significant (Harrison (2024)). 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 advantage of hypothetical elicitation 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, for beliefs about certain non-verifiable events, 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 *directly* report their beliefs only 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

[5]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 et al. (2009) and Andersen et al. (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.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

Malouf (1979).[6] The first statements of this mechanism, joining the QSR and the BLP, appear to be Allen (1987) and McKelvey and Page (1990).[7]

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. Under EUT the theoretical basis of the BLP is non-controversial, and under RDU it only requires that the probability weighting function should increase. The disadvantage is that it adds an additional layer of understanding for subjects when it comes to translating earnings in the QSR into cash. In our experience, documented in Harrison et al. (2014a) (2014b) and Harrison et al. (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. 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.

Although there are some studies showing the apparent failure of the BLP in other settings, that evidence is not as compelling as many assume: see Harrison et al. (2013) for a critical literature review. The claim by Danz et al. (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 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 actually hold, and they are not just testing the latter hypothesis.

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.[8] 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. This approach was used by Andersen et al. (2014) for the recovery of beliefs over binary events by jointly estimating risk preferences and beliefs, allowing risk preferences to be characterized by EUT or RDU. We extend that approach in Section 2 and 3 to beliefs about non-binary events and to the level of the individual subject.

[6]Berg et al. (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.

[7]Schlag and van der Weele (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 et al. (2015) evaluated the QSR with the BLP in terms of the elicitation of subjective belief distributions. Harrison and Phillips (2014b) applied this method to elicit belief distributions about financial risk by Chief Risk Officers of major companies, all of whom had advanced training in statistical and actuarial methods such that the compound risks of adding the BLP would not likely be confusing.

[8]Or one could assume explicitly that the subjects are approximately risk neutral over small stakes, as in Rutström and Wilcox (2009) and Eil and Rao (2011), but then one locks in the use of stakes which might not provide salient motivation.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

Overall, we see the second and third approaches as the most attractive, and complementary, and have used both in different settings. The evidence of hypothetical bias across a wide range of elicitation methods in experimental economics is clear to us, and usually easy to avoid. The second and third approaches 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. We show how experimenters can meet that burden.

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)). Partition the domain into K intervals, and denote as r_k the report of the likelihood that the event falls in interval $k = 1, \ldots, K$. 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, $\{ r_1, r_2, \ldots, r_k, \ldots, r_K \}$ such that $r_k \geq 0 \; \forall k$ and $\sum_{i=1 \ldots K}(r_i) = 1$.

If k is the interval in which the actual value lies, then the payoff score is defined by Matheson and Winkler (1976): $S = (2 \times r_k) - \sum_{i=1 \ldots K}(r_i)^2$. The reward in the score is a doubling of the report allocated to the true interval, and the penalty depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval the punishment is severe. The punishment includes all possible reports, including the correct one.

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

$$\alpha + \beta \left[(2 \times r_k) - \sum\nolimits_{i=1 \ldots K} (r_i)^2 \right], \tag{1}$$

where we initially assumed $\alpha = 0$ and $\beta = 1$. We can assume $\alpha > 0$ and $\beta > 0$ to get the payoffs to any positive level and units we want.[9] Let p_k represent the underlying, true, latent subjective probability of an individual for an outcome that falls into interval k.

Harrison et al. (2017) characterized several general properties of the QSR. Let the reports consist of an allocation of some number of tokens, typically 100, to each possible event. First, if the subject places zero subjective belief on any event, she should not allocate any tokens to those events. Second, if the subject has the same subjective belief for two or more events, she should allocate the same number of tokens to each of those events. Both of these results apply under EUT and RDU, and whether or not the subject is risk neutral. Third, if the agent is risk averse, under SEU she should report tokens that are "flatter" than her latent beliefs over the events she assigned positive subjective probability, in effect reducing the variance of the subjective risk of different payoffs for the same average payoff. How much flatter this allocation of tokens is depends on the extent of risk aversion.

We extend this characterization to agents whose risk preferences are characterized by RDU rather than EUT. We state parametric versions of EUT and RDU decision making over objective probabilities, to introduce notation and basic concepts. Nothing hinges on the parametric assumptions, although the parametric forms assumed are standard in the literature.

Assume that utility of income in an elicitation is defined by the Constant Relative Risk Aversion (CRRA) function,

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

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,

[9] To help visualize how this QSR works out in practice, Figures 1 and 2 illustrate one representation which we use in experiments, for $\alpha = \beta = 25$ and $K = 10$. We discuss the experimental tasks later.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

$s < 0$ 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. Under EUT the probabilities for each outcome x_j, $p(x_j)$, are those that are stated, so expected utility is simply the probability weighted utility of each outcome in each lottery i:

$$\text{EU}_i = \sum\nolimits_{j=1,J} [\, p(x_j) \times U(x_j) \,]. \tag{3}$$

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 EU defined by (3) with RDU

$$\text{RDU}_i = \sum\nolimits_{j=1,J} [\, w(p(x_j)) \times U(x_j) \,] = \sum\nolimits_{j=1,J} [\, w_j \times U(x_j) \,] \tag{4}$$

where

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

for $j = 1, \ldots, J-1$, and

$$w_j = \omega(p_j) \tag{6}$$

for $j = J$, with the subscript j ranking outcomes from worst to best, and $\omega(\cdot)$ is some probability weighting function. In this RDU specification of probability weighting, the arguments of the probability weighting function are the *decumulative* probabilities, not the probabilities for each event.[11]

We use a general functional form for probability weighting proposed by Prelec (1998) that exhibits considerable flexibility. This function is

$$\omega(p) = \exp\{-\eta(-\ln p)^{\varphi}\}, \tag{7}$$

and is defined for $0 < p \leq 1$, $\eta > 0$ and $\varphi > 0$. When $\varphi = 1$ this function collapses to the Power function $\omega(p) = p^{\eta}$. It also provides a good approximation to the Inverse-S function $\omega(p) = p^{\gamma}/(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma}$. One important reason to be able to recover beliefs that accommodate a flexible specification is that there is clear evidence that there is considerable individual heterogeneity in the *type* of probability weighting that individuals exhibit (Wilcox (2023)).

Our two theoretical results characterize how to recover beliefs for individuals that distort probabilities. The first result shows how to recover such beliefs from observed reports from a QSR:

Theorem 2.1 *Let p_k represent the underlying subjective probability of an individual for outcome k and let r_k represent the reported probability for outcome k in a given scoring rule. Let $\theta(k) = \alpha + \beta 2r_k - \beta \sum_{i=1\ldots K}(r_i)^2$ 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 (5) and (6). 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*

[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.

[11] For the event with the highest payoff, the decumulative probability *is* the probability of the event, since the decumulative probability is the probability of getting at least the highest payoff. But for the event with the second-highest payoff, the decumulative probability is the probability of getting at least the second-highest payoff, which is the sum of the probability of getting the highest payoff and the probability of getting the second-highest payoff.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

probabilities, the actual and reported probabilities must obey the following system of equations:

$$w(p_k) \times \partial u/\partial\theta\,\big|_{\theta=\theta(k)} - \sum\nolimits_{j=1,K} \{\, w(p_j) \times r_k \times \partial u/\partial\theta\,\big|_{\theta=\theta(j)} \,\} = 0,\ \forall k = 1,\ldots,K \tag{8}$$

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

$$\mathbf{Max}_{\{\,r\,\}}\ E_{w(p)}\,[\,u(\theta)\,] \text{ subject to } \sum\nolimits_{i=1\ldots K} (r_i) = 1 \tag{9}$$

where $E_{w(p)}\,[\,u(\theta)\,] = \sum_{j=1\ldots K} w(p_j) \times u[\,\alpha + \beta 2 r_j - \beta \sum_{i=1\ldots K} (r_i)^2\,]$. In some experimental configurations there may be K additional constraints: $r_i \geq 0$ for $i = 1,\ldots,K$. These constraints are not included in (9) because they are automatically satisfied by the solution (8) for both risk-averse and risk-loving individuals.

Problem (9) can be solved by maximizing the Lagrangian

$$\mathcal{L} = \sum\nolimits_{j=1\ldots K} w(p_j) \times u[\,\alpha + \beta 2 r_j - \beta \sum\nolimits_{i=1\ldots K} (r_i)^2\,] - \lambda\,[\,\sum\nolimits_{i=1\ldots K} (r_i) - 1\,] \tag{10}$$

The solution to the problem must satisfy $K+1$ conditions. The K first order conditions with respect to report r_k, $\forall k = 1,\ldots,K$, are[12]

$$\partial\mathcal{L}/\partial r_k = \sum\nolimits_{j=1\ldots K} (\, w(p_j) \times \partial u(\theta(j))/\partial r_k\,) - \lambda = 0,\ \forall k = 1,\ldots,K \tag{11}$$

where $\partial u(\theta(j))/\partial r_k = \partial u/\partial\theta\,\big|_{\theta=\theta(j)} \times (2\beta\delta_{jk} - 2\beta \times r_k)$ 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 (10) with respect to the Lagrangian constant:

$$\sum\nolimits_{i=1\ldots K} (r_i) - 1 = 0. \tag{12}$$

We can simplify the K equations in (11) as:

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

$$\text{or}\quad w(p_k) \times (\,\partial u/\partial\theta\,\big|_{\theta=\theta(k)}\,) - r_k\, E_{w(p)}\,[\,\partial u/\partial\theta\,] = \lambda/2\beta,\ \forall k = 1,\ldots,K. \tag{13}$$

Summing over the K first-order conditions we get

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

Notice that $\sum_{k=1\ldots K} r_k\, E_{w(p)}\,[\,\partial u/\partial\theta\,] = E_{w(p)}\,[\,\partial u/\partial\theta\,]$ because the expectation term is a constant and because of (12). Then (14) implies that $K\lambda/2\beta = 0$, which can only be satisfied if $\lambda = 0$ since $K > 0$ and $\beta > 0$. This result and (13) implies that the solution to problem (9) must satisfy the following K conditions:

$$w(p_k) \times \partial u/\partial\theta\,\big|_{\theta=\theta(k)} - r_k \times E_{w(p)}\,[\,\partial u/\partial\theta\,] = 0,\ \forall k = 1,\ldots,K. \tag{15}$$

The application of (8) 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

[12] 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.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

is a linear system of equations in the unknown decision weights.[13] 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.

It is notable that Theorem 2.1 does not specify particular parametric forms for utility functions or probability weighting functions.

Our second theoretical result is a generalization of Theorem 2.1 to all *proper* scoring rules. Define a scoring rule S where $S_1(r_1, \ldots, r_n)$, $S_2(r_1, \ldots, r_n)$,…, and $S_n(r_1, \ldots, r_n)$ represent the payoffs for each of the possible states of nature $1, \ldots, n$. S_k is the payoff if state k is realized after reports $r_1, \ldots, r_n$, where $r_n = 1 - \sum_{i=1,\ldots n-1} r_i$. Let $f(p_1, \ldots, p_n; r_1, \ldots, r_n) = \sum_{i=1,\ldots n} p_i S_i(r_1, \ldots, r_n)$. A scoring rule is said to be 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 provided by Theorem 2.1. Our generalization is then:

Theorem 2.2 *Theorem 2.1 generalizes to include all proper scoring rules.*

The proof of Theorem 2.2 is not as instructive or compact as the proof of Theorem 2.1, and is relegated to Appendix A (online) As a corollary, Theorem 2.2 also means that Propositions 1 through 7 of Harrison et al. (2017), characterizing the beliefs recovered for an SEU decision-maker, also generalize to include all proper scoring rules, not just the QSR.

3. Recovering beliefs by eliciting risk preferences

We now turn to how to apply these theoretical results. The first step is to see how the QSR works with simple numerical examples. The second step is then to see how Theorem 2.1 is applied with specific functional forms for the RDU risk preferences of an individual, with assumed parameter estimates. The third step is to review how we can exactly infer these parameters using Bayesian methods applied to a risky lottery task undertaken by the same subject that reported beliefs in a QSR.

3.1. Numerical examples

To understand the basic logic of the QSR in (1), consider some numerical examples assuming $K = 4$, $\alpha = 0$ and $\beta = 1$. 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, and that is what $\alpha > 0$ does. Assuming that the subject has very diffuse subjective beliefs and allocates 25% of the weight to each interval, the score is less than 1: $S = (2 \times 1/4) - ((1/4)^2 + (1/4)^2 + (1/4)^2 + (1/4)^2) = ½ - 1/4 = 1/4 < 1$. So the tradeoff from the last case is that one can always ensure a score of $1/4$, but there is an incentive to provide less diffuse reports, and that incentive is the possibility of a score of 1.

A numerical example can also illustrate the basic ideas of Theorem 2.1. Assume $K = 10$ bins. An individual reports 30, 45 and 25, out of 100 tokens, in bins 3, 4 and 5, respectively, 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 (1) with $\alpha = \beta = 25$, consistent with the experiments reported later. Let the CRRA utility function be given by (2) with

[13] 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 $w(p_k) = r_k\, \theta(k)^{-s} / \sum_{j=1\ldots K} r_j\, \theta(j)^{-s}$.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

$s = 0.77$, broadly consistent with evidence from a wide array of experiments, so that $\partial u/\partial\theta = \theta^{-s}$.[14] For RDU individuals further assume the inverse-S probability weighting function with $\gamma = 0.5$ and numerically invert it.[15]

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

$$0.049591 \times w(p_3) - 0.018 \times w(p_4) - 0.02267 \times w(p_5) = 0 \tag{16a}$$

$$-0.03188 \times w(p_3) + 0.032997 \times w(p_4) - 0.034 \times w(p_5) = 0 \tag{16b}$$

$$w(p_3) + w(p_4) + w(p_5) = 1 \tag{16c}$$

The numbers in these 3 equations are direct applications of (8). For example, from (16a) we have

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

We solve (16a), (16b) and (16c) 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 $w(p_5)$ $(0.25/0.075562439)/15.04384315 = 0.219925$, and again these decision weights sum to 1. 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 et al. (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 had risk preferences characterized by RDU. 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. We sort so that $w(p_5) = 0.219925$, $w(p_3) = 0.281486$ and $w(p_4) = 0.498589$. We then apply a decumulative process to extract the cumulative distribution function for the probabilities. For example, p_4 produces the largest decision weight, since bin 4 was allocated the most tokens, and the relevant probability of being in bin 4 is then $\omega^{-1}(0.498589) = 0.802313$. The probability of being in bin 3 is then $\omega^{-1}(0.281486 + 0.498589) - \omega^{-1}(0.498589) = 0.179121$, since bin 3 was allocated the second-highest number of tokens, and the residual probability of being in bin 5 is then $\omega^{-1}(0.219925 + 0.281486 + 0.498589) - \omega^{-1}(0.281486 + 0.498589) = 0.018566$. These probabilities must finally be "de-sorted" to connect with the appropriate bin, so $p_3 = 0.179121$, $p_4 = 0.802313$ and $p_5 = 0.018566$. These are significant, first-order, differences from the observed reports $r_3 = 0.30$,

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

[15] Since the Inverse-S probability weighting function is not monotonic for $\gamma < 0.278$, as noted by Rieger and Wang (2006), 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 the inverse function is $p^{(1/\gamma)}$, and for the Prelec probability weighting function the inverse function is $\exp\{-[(\ln p)/(-\eta)]^{1/\varphi}\}$.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

$r_4 = 0.45$ and $r_5 = 0.25$, relative to the second-order effect of risk aversion. The mean of the recovered beliefs in this case is 33.39445, different from the mean of 34.5 for observed reports.[16]

3.2. Applying our theorems with Bayesian estimates of risk preference parameters

The numerical examples picked empirically reasonable values for risk preference parameters. When we turn to implementing our Theorem 2.1 to specific subjects, we need to obtain estimates of the risk preference parameters for those subjects. To do so we have the same subjects that have completed a subjective belief elicitation task with a QSR also complete a task in which they make choices over risky objective lotteries. We then estimate risk preference parameters for that subject from their choices in this risky lottery task. Harrison and Rutström (2008) review a variety of elicitation interfaces for risk preferences, and the classical econometrics that can be used to then infer risk preferences under EUT and RDU by structural estimation.

In our case we will use Bayesian econometric methods, for reasons noted earlier. The exact specification is explained below. The result of using these methods is a posterior distribution for each of the risk preference parameters of an individual. If we assume an RDU model, with a CRRA utility function and a Prelec probability weighting function, we have parameters s, η and φ for each subject. In modern Bayesian econometrics, these posteriors are approximated by MCMC samples. Typically there will be thousands of samples, in our case 100,000. Each sample consists of one value of s, η and φ. Those 3 numbers fold *directly* into our Theorem 2.1, as the numerical examples illustrated, and generate a vector of K subjective probabilities *exactly*. By applying Theorem 2.1 for each subject and each belief elicitation report, we end up with 100,000 recovered beliefs for that subject and belief question. Each recovered belief, then, amounts to a $100{,}000 \times K$ matrix. The 100,000 values for each of the K bins represents the posterior predictive distribution for that subjective belief of each subject and bin. Theorem 2.1 ensures that these 100,000 recovered beliefs each sum to 1, as probabilities must. We can then examine the average probability for each bin for a given subject, the median, or the complete distribution.

What makes our approach Bayesian is not Theorem 2.1, which tells us how to infer latent beliefs from observed reports conditional on us knowing the parameters defining the risk preferences of the subject. The use of a Bayesian posterior distribution for these risk preferences is what renders the conditional inferences from our Theorem 2.1 unconditional and "Bayesian."

3.3. A Bayesian hierarchical model for risk preference parameters of individuals

The novelty with all Bayesian models is how the prior beliefs about parameter values are defined, and then used with the likelihood evaluators to infer posterior beliefs about parameter values. The novelty of a BHM is that hyper-priors for the individual-specific parameters are estimated with pooled data over the sample, and the resulting hyper-posterior then used as an *informative* prior for the estimation of the individual-specific posterior for that parameter. The BHM is now a standard type of Bayesian model. Excellent expositions include Johnson et al. (2012), Kruschke (2015) and Kruschke and Vanpaemel (2015). Applications to the modeling of risk preferences using a BHM include Nilsson et al. (2011), Gao et al. (2023), and Balcombe and Fraser (2025).

The data-generating process revolves around core parameters s_i, η_i, φ_i and μ_i, with all but the last parameter introduced earlier. Define the latent index for choice t by subject i as the difference between the RDU of the left and right lottery subject to a Fechner noise parameter μ_i and a random noise term ε_{it}:

[16]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 stem from subjects with extreme, *a priori* unreasonable preferences.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

$$\begin{aligned} y_{it}^{*} &= \nabla \text{RDU}_{it}(s_i, \eta_i, \varphi_i, \mu_i) + \varepsilon_{it} \\ &= \left\{ \left[(\text{RDU}_{it}^{R}(s_i, \eta_i, \varphi_i) - \text{RDU}_{it}^{L}(s_i, \eta_i, \varphi_i)) / \nu_{it} \right] / \mu_i \right\} + \varepsilon_{it} \,, \end{aligned} \tag{17}$$

where ν_{it} is a "contextual utility" term specific to choice t to normalize utilities of prizes between 0 and 1, and s_i, η_i, φ_i and μ_i are the parameters for subject i we want to estimate (see Wilcox [2011]). We assume ε_{it} follows a logistic distribution and is independently and identically distributed across individuals and decisions. If that subject i selects the right lottery in lottery pair t whenever the latent index y_{it}^{*} is greater than or equal to 0, then

$$\text{Prob}(y_{it} = 1) = \text{Prob}(\nabla \text{EU}_{it}(r_i, \alpha_i, \mu_i) \geq 0) = \Lambda(\nabla \text{EU}_{it}(r_i, \alpha_i, \mu_i)) \,, \tag{18}$$

where $\Lambda(x) = 1/(1 + e^{-x})$ is the logistic function. The specifications (17) and (18), connecting the core model of risk preferences to the likelihood of observing the actual choice, are common in the literature on the structural estimation of risk preferences reviewed by Harrison and Rutström [2008].

We posit *two* hyper-parameters that describe the distribution that characterizes *each* of

- s_i, the curvature of the utility function of individual i;
- η_i, one of the parameters of the probability weighting function of individual i;
- φ_i, the other parameter of the probability weighting function of individual i; and
- μ_i, the Fechner noise parameter of individual i.

Hence we estimate 8 hyper-parameters in all, based on the pooled data across all N subjects. In addition, we estimate s_i, η_i, φ_i and μ_i for each individual $i = 1, \ldots, N$. In all, therefore, we jointly estimate $8 + (4 \times N)$ parameters for the full RDU hierarchical model. Hence if $N = 80$, for example, we jointly estimate 328 parameters for the RDU model.

Turning to the specific prior distributions assumed, we generally adopt diffuse or weakly informative[17] hyper-priors, and then of course allow the estimated hyper-posterior be the informative prior for the parameters of each individual. The models were solved using 100,000 MCMC simulations over 4 chains. The summary Gelman-Rubin statistic for convergence is below 1.1, a widely used threshold. The priors and convergence properties of the BHM models are documented further in Appendix B (online).

3.4. Software

We have developed a *Stata* program to undertake these belief recovery calculations for practical applications, and document this in Appendix C (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. We also have *Stata* programs to estimate a BHM for the EUT or RDU risk preferences of individuals, documented in Appendix B (online).

4. Experimental procedures

Figures 1 and 2 illustrate the interface used to elicit beliefs from subjects, developed by Harrison et al. (2017). Appendix D (online) contains the instructions presented to subjects. In these experiments we used a generic question on unemployment to *explain* the interface, rather than the specific questions that subjects were then asked to respond to with incentives. This approach avoids our worked examples in the instructions having any "priming" effect on subjects for the questions we were paying

[17] Our weakly informative priors only restrict certain parameters to be positive, for statistical or theoretical reasons. For example, parameters for standard deviations must be non-negative by definition, and η, φ and μ are defined by the economics of the model to be positive.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

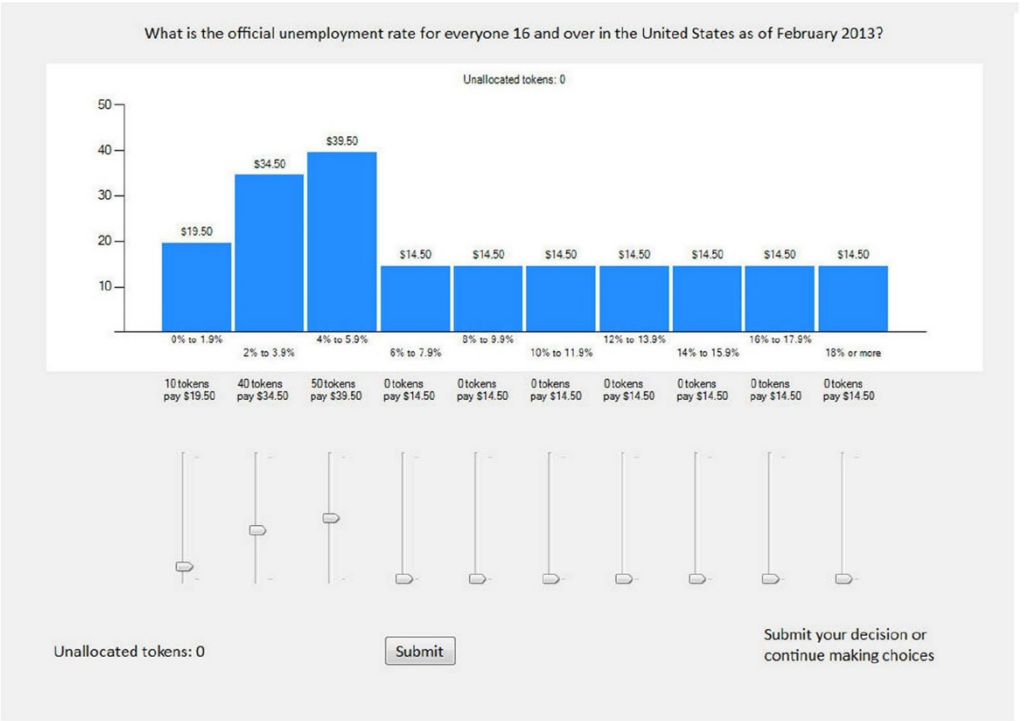


Figure 1. Belief elicitation interface

them to answer. It also allows us to then have batteries of incentivized belief questions across several domains.

Each subject allocates 100 token across 10, 8 or 4 bins, and the displays change in real-time so the subject sees the payoffs from the bets being placed about the correct answer. As stressed by Savage (1971; 1972), 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 from the population of undergraduate students from a variety of majors at Georgia State University. In the first application reported in Section 5, all subjects had completed 30 incentivized, binary choices over risky lotteries, in experimental sessions documented by Harrison et al. (2026) using lottery pairs selected at random from a larger battery of 67 gain-frame lottery choices from Harrison and Swarthout (2023). In the second application reported in Section 5, all subjects had also completed 100 incentivized, binary choices over risky lotteries, in experimental sessions documented by Harrison et al. (2022), using the full battery of 100 gain-frame, mixed-frame and loss-frame lottery choices from Harrison and Swarthout (2023). The lotteries were designed, following a neglected design of Loomes and Sugden (1998), to provide stress tests of the independence axiom of EUT as well as to allow for identification of a wide range of risk attitudes. The BHM of Gao et al. (2023) was applied to these data, assuming an EUT or RDU model of risk preferences with a CRRA utility function and the Prelec probability weighting function.[18] The result of that estimation

[18]Any loss-frame or mixed-frame lotteries were accompanied by a house endowment to cover losses. Under EUT and RDU these payouts were assumed to be netted out of the endowment, so that all lotteries involved non-negative payoffs.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

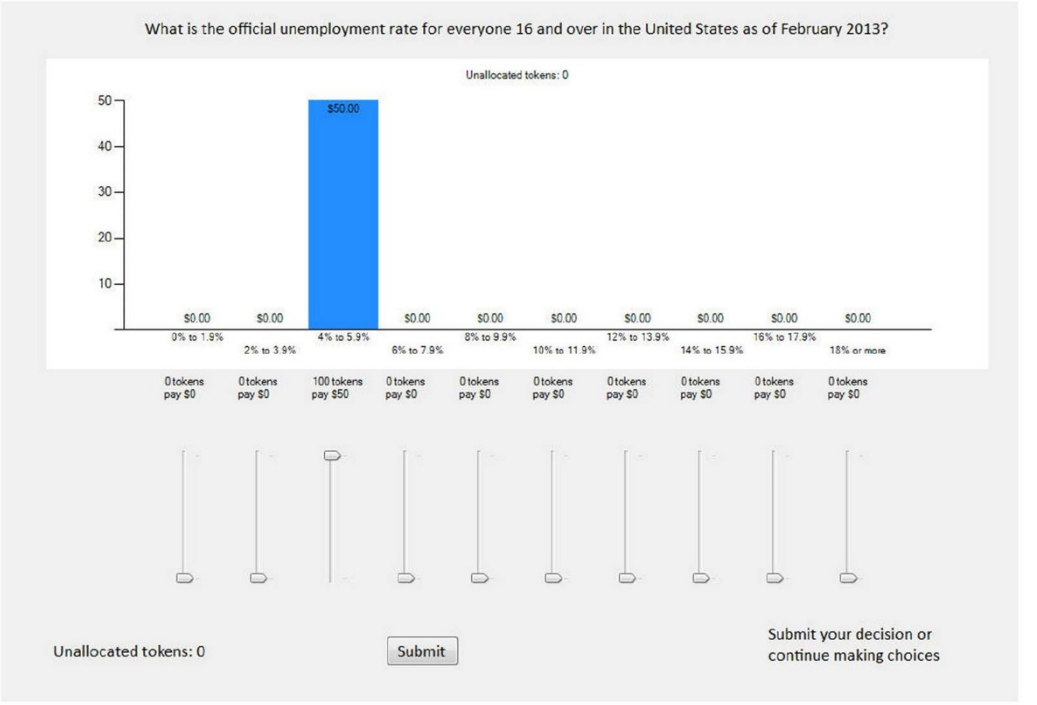


Figure 2. One possible belief elicitation response

is a Bayesian posterior distribution for each subject with 100,000 samples approximating each of the parameters of an RDU model of their risk preferences.

5. Applications

We examine two applications. The first is from a well-known design in experimental economics, examining the hypothesis that women are less competitive than men. Subjective beliefs about winning a tournament play a central role in evaluating that hypothesis, along with risk preferences of course. The second is from a direct elicitation of beliefs about the addictiveness of smoking. These beliefs play a critical role in public health policy and tobacco litigation.

5.1. Beliefs about gender and competitiveness

Niederle and Vesterlund (2007) consider the effect of more or less competitive compensation schemes on the incentives that lead women to participate in a workforce characterized by those schemes. At one extreme is a piece rate system, in which a fixed payoff is provided for every unit produced in a given period of time. At the other extreme is a tournament, in which only one in four participants will receive a positive payoff, based on having the greatest number of units produced in a given time period. The setting is one in which the production process requires real cognitive effort over some period of time. They observe from laboratory experiments that women tend to avoid tournaments in favor of piece rates, when asked to choose one or the other to be compensated, even when the production task has been selected to be one where men and women have roughly the same ability. This tendency is regarded as a bad thing, as reflected by the first part of the title "Do Women Shy Away from Competition?" and myriad references in a large, subsequent literature to women not being willing to compete.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

We examine subjective beliefs elicited as part of an experiment by Harrison et al. (2026) that mimics the excellent design of Niederle and Vesterlund (2007). There are six parts to this design. In **part 1** the subject is asked to solve up to 12 questions evaluating fluid intelligence in 10 minutes, and are rewarded by points. Each QSR allowed subjects to earn up to 200 points in each question if all tokens were allocated to the correct response. If this part was selected for payment, the reward would be based on their total earnings of points, multiplied by \$0.01, so each point was worth a penny, and subjects could earn up to \$2 per question or \$24 for the 12 questions. This is the piece rate compensation treatment. In **part 2** the underlying cognitive production task was the same, with a different 12 questions in 10 minutes. However, the compensation scheme was based on a tournament in which the point earnings of the subject would be compared to the point earnings of 3 other subjects selected at random. Only the subject with the best score would receive compensation[19] if this part was selected for payment, but the potential payoff was 4 times the payoff from the piece rate part 1, so \$8 per question or \$96 for the 12 questions.

In **part 3** the focus is on the choice of compensation scheme. The subject is told that they will face yet another 12 questions in 10 minutes. Without knowing their earnings, or the earnings of anyone else from parts 1 and 2, the subject must decide whether to have their earnings in part 3 determined by the piece rate scheme of part 1 or the tournament scheme of part 2. If the tournament is selected, the new point earnings of the subject will be compared to the point earnings from part 2 of the other 3 members of her group. This is a key point of the design.

There are three extra steps to the experimental design, and they play a surprising role in helping evaluate observed behavior. In **part 4** the focus is again on the choice of compensation scheme, but here the subject decides if she wants her earnings from part 1, the initial piece rate task, to be evaluated with the piece rate compensation scheme or the tournament compensation scheme.[20] This choice of compensation schemes differs from part 3, since it uses previously generated earnings by the subject, rather than adding the "stress" of competing afresh. Hence part 4 elegantly teases apart two attributes of competition: being asked to *perform* some task knowing that the earnings will depend on doing it better than others, and being *compared* to others to generate a binary payoff for the best performance. Part 4 controls for any stress from having to perform, but leaves in place the risky lottery involved if the subject selects to be evaluated by a tournament.

The final steps of the experimental design elicit subjective beliefs that drive the choices in parts 3 and 4. In **part 5** the subject reports beliefs about the rank of her piece rate earnings in part 1 compared to the others in her tournament group of 4. We used a QSR with four possible responses, and provide 100 tokens for the subject to allocate. The subject could earn up to \$30 by allocating all 100 tokens to the correct response. Finally, in **part 6** the subject again reports beliefs about her point earnings rank, in this case with respect to the tournament earnings in part 2. The QSR setting and incentives are the same as in part 5. Of course, the response from part 6 provides exactly the individual subjective probability of winning the tournament that is needed to evaluate the choice in part 3 descriptively and normatively; we do not need to assume a diffuse probability of $1/4$ of winning.[21]

As flagged above, the subjective beliefs elicited in part 6 play a crucial role in evaluating behavior. Niederle and Vesterlund (2007), p.134 summarize their data as follows:

> Accounting for ties, at most 30 percent of men and women should guess that they are the best in their group of four. We find that 75 percent of men compared to 43 percent of women

[19]Subjects were told that ties would be resolved randomly.

[20]In part 3 and part 4 subjects do not yet know their realized earnings from parts 1 or 2.

[21]The belief elicitation step is not in the written instructions for the experiment of Niederle and Vesterlund (2007), but is embedded in *Ztree* code available from Lise Vesterlund (private communication, May 2021). It asked the subject to "guess" their rank in part 2 or part 1, and pays \$1 if the subject is correct and \$0 if incorrect. Of course, this procedure elicits the modal belief over the 4 possible ranks; it does not elicit the subjective probability of winning the tournament. And the subjective probability of the modal belief can be anywhere between 0.26 and 1.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

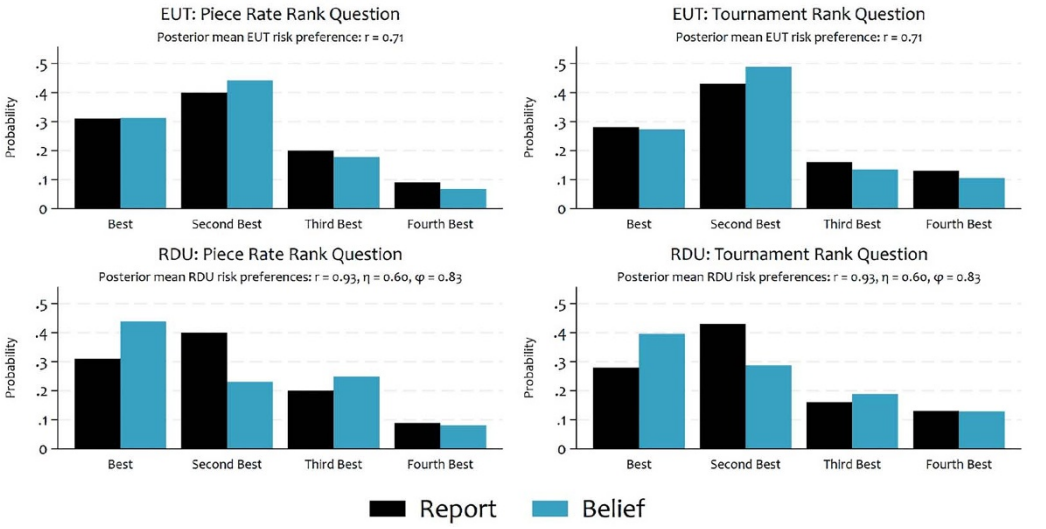


Figure 3. Recovered beliefs of subject #22 with respect to competitiveness

guessed that they were the best. While both men and women are overconfident, men are more overconfident than women.

Recognizing that the word "overconfident" is used here to mean optimistic, this finding is crucial. The full title of Niederle and Vesterlund (2007) is "Do Women Shy Away From Competition? Do Men Compete Too Much?" It appears that their aggregate data say clearly that it is the latter, not the former.

We had 88 subjects undertake this task: 45 with 80 tokens to allocate in parts 1, 2 and 3, and 43 with only 1 token to allocate. Every subject completed a risk battery of 30 binary choices, selected at random from a larger battery of 67 gain-frame lottery choices.

If we just focus on *reports* about winning the tournament, we find that men and women are slightly pessimistic in comparison to the diffuse prior of 1/4 when there was only 1 token. On average men hold beliefs that are right at the diffuse prior of 1/4 when there were 80 tokens, and women were modestly optimistic on average when there are 80 tokens. However, this modest optimism by women was actually a marked optimism in terms of the *recovered beliefs*.[22]

With this background, we can use these data to illustrate the effect of the recovery of beliefs. Figure 3 displays beliefs by subject #22. The top two panels assume EUT for the risk preferences, and the bottom two panels assume RDU for the risk preferences; the mean posterior estimates are reported in each panel. The left two panels show beliefs that the subject's piece rate earnings would be ranked first, second, third or fourth in comparison with 3 other subjects, and the right two panels show comparable beliefs about ranks for the subject's tournament earnings. The first thing to see is how close the recovered EUT beliefs are to the reports, as expected from theory. The second thing to see is how much the recovered RDU beliefs differ from the reports. The third thing to see is the substantive import of these differences: using the reports or just the EUT recovered beliefs, we would infer that subject #22 thought that she was most likely *not* to win a tournament with either her piece rate or prior tournament performance, but this is exactly the *opposite* of what we infer if we use the RDU recovered beliefs. The final substantive thing to see is that this elicitation procedure does more than just inform us about the modal belief, it tells us the *value* of the modal belief probability. This

[22]It is optimism relative to the default, diffuse prior of 1/4.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

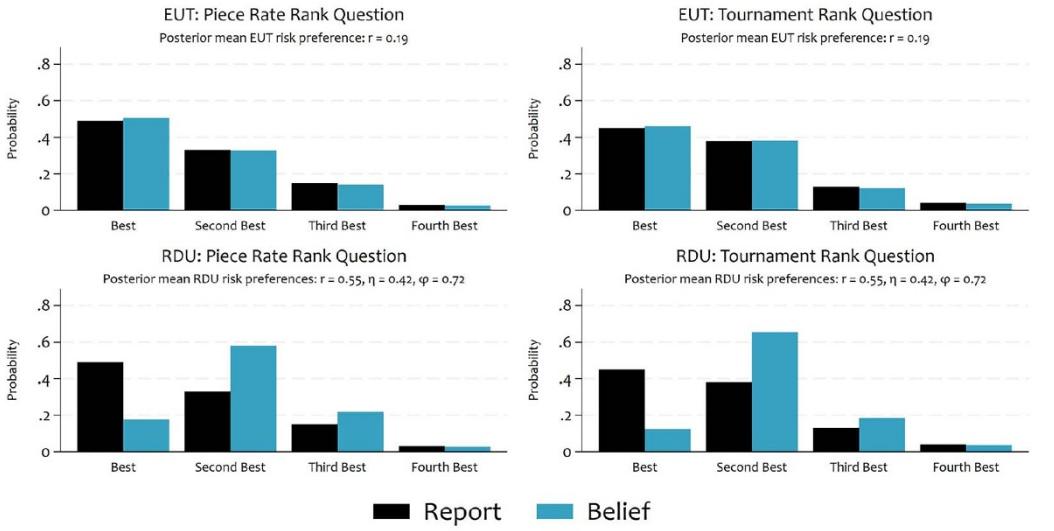


Figure 4. Recovered beliefs of subject #25 with respect to competitiveness

value is needed to calculate whether subject #22 thinks she has a better chance of winning a tournament with her piece rate performance or her tournament performance. In this case she has a slightly better chance with her piece rate performance.[23]

Figure 4 shows the reports and recovered beliefs for subject #25, with comparable general results: recovered EUT beliefs are again close to the reports compared to the recovered RDU beliefs. In this case we observe the reverse of the qualitative conclusion that we had for subject #22. Here the recovered RDU beliefs show a subject that believes that she is *less* likely to win a tournament with either piece rate or tournament performance. These low values for the chance of being the winner also suggest that the opportunity cost of making an error in this decision to compete cannot be too large, since she believes it would happen with very low probability in either case. Hence we are getting information on the descriptive strength of the incentive to report the correct choice from knowing the values of these non-modal ranks.

Figures 5 and 6 mimic the qualitative patterns seen in Figures 3 and 4, respectively.

The substantive question here is how men and women balance their subjective beliefs about winning the tournament, if they opt for it in parts 3 and 4, with their risk preferences. Even with optimistic beliefs about winning the tournament, sufficiently risk averse agents should still not select it. Harrison et al. (2026) show that women balance these risk perceptions and preferences just as well as men, when properly evaluated in terms of the expected subjective welfare of observed choices, rather than the welfare-irrelevant metric of choosing to compete. For that evaluation one needs the subjective beliefs recovered here.

5.2. Beliefs about the addictiveness of smoking

We now consider an application in an incentivized experiment using GSU undergraduate subjects examining beliefs about a basic informational input to the decision to smoke when young: beliefs about how addictive smoking is.[24] The "rational model of addiction" of Becker and Murphy (1988)

[23]The value of the modal beliefs are also needed for the *normative* evaluation of the gained or foregone Certainty Equivalent of making the decision to compete, as calculated by Harrison et al. (2026).

[24]Declaration of Potential Competing Interests: Harrison served as a testifying expert witness for plaintiffs in litigation against tobacco companies in the United States and Canada for 30 years. His testimony was primarily on health care costs to

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

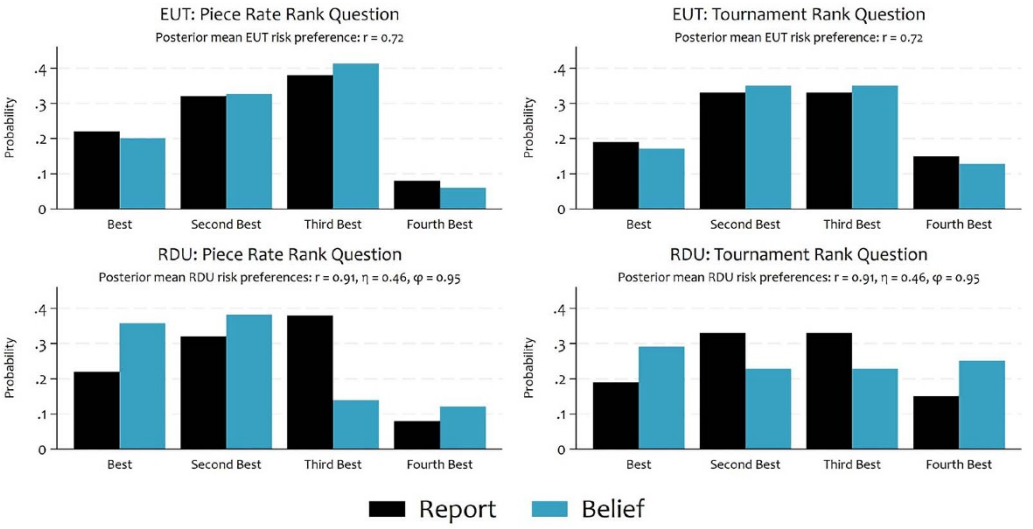


Figure 5. Recovered beliefs of subject #39 with respect to competitiveness

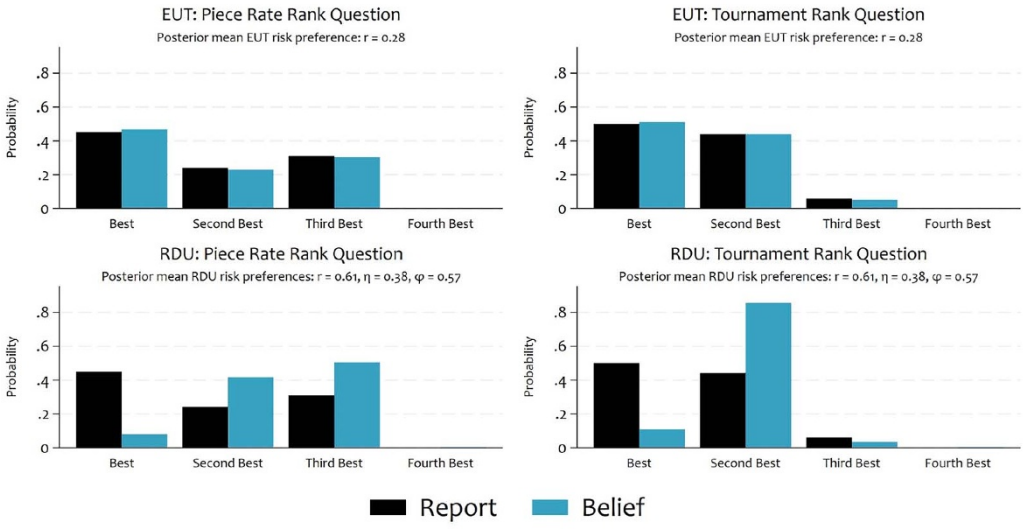


Figure 6. Recovered beliefs of subject #62 with respect to competitiveness

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.

governments associated with tobacco-related diseases, and he 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.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

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 smoking, could lead to young smokers becoming 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[25] 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 et al. (2016), p.5}.

Question #1 used 8 bins for responses, and question #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.

Individual results for two subjects are presented in Figures 7 and 8, focusing on the comparison of observed reports and recovered beliefs using the posterior mean belief.[26] 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

[25]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%.

[26]Each of the 100,000 samples of estimates of the set $\{s, \eta \text{and } \rho\}$ 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 estimates of the set $\{s, \eta \text{and } \rho\}$ from the sample of 100,000 that are arbitrarily close to these means.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

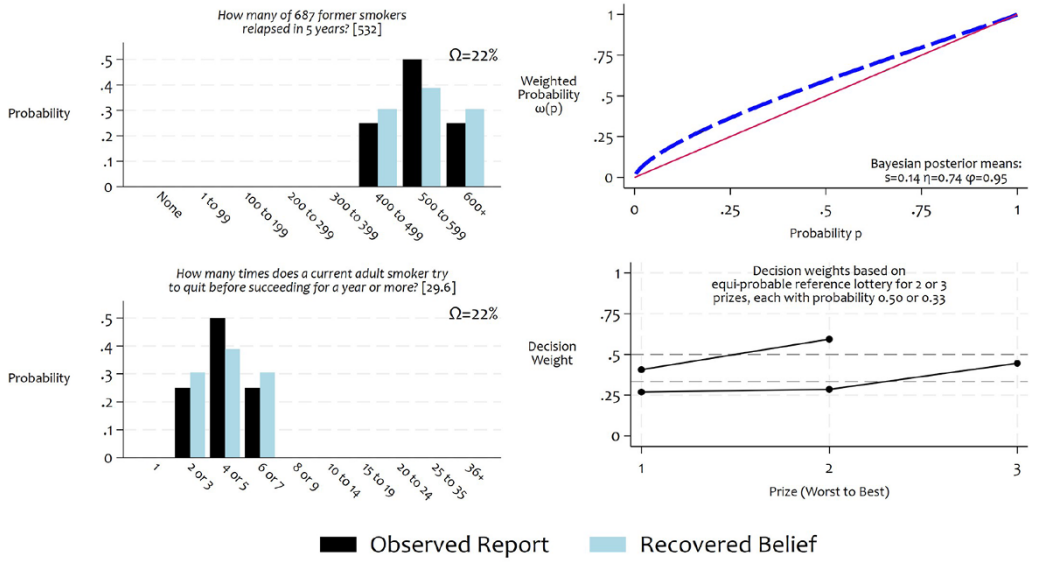


Figure 7. Reports, beliefs and risk preferences for subject #106

recovered beliefs cannot arise at all when there are no tokens allocated to a bin, so the upper bound for Ω is strictly less than 100%.

Figure 7 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 slightly concave utility function, with $s = 0.14$. 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 (largest) QSR payoff being viewed 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 payoffs if those outcomes were to be 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, because 2 of the 3 bins with positive reports have the same report. 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 "lotteries" the subject faces when deciding to smoke or not, it is critical for inference that we correctly recover the distribution and not just the average.

Figure 8 displays the results from a subject that has pessimistic probability weighting, leading to decision weights in which the lowest ranked QSR payoffs are given extra weight and the highest ranked QSR payoffs are given less weight (bottom right panel). This is an interesting case because the recovered belief distributions are generally more precise, with less variance, than the observed

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

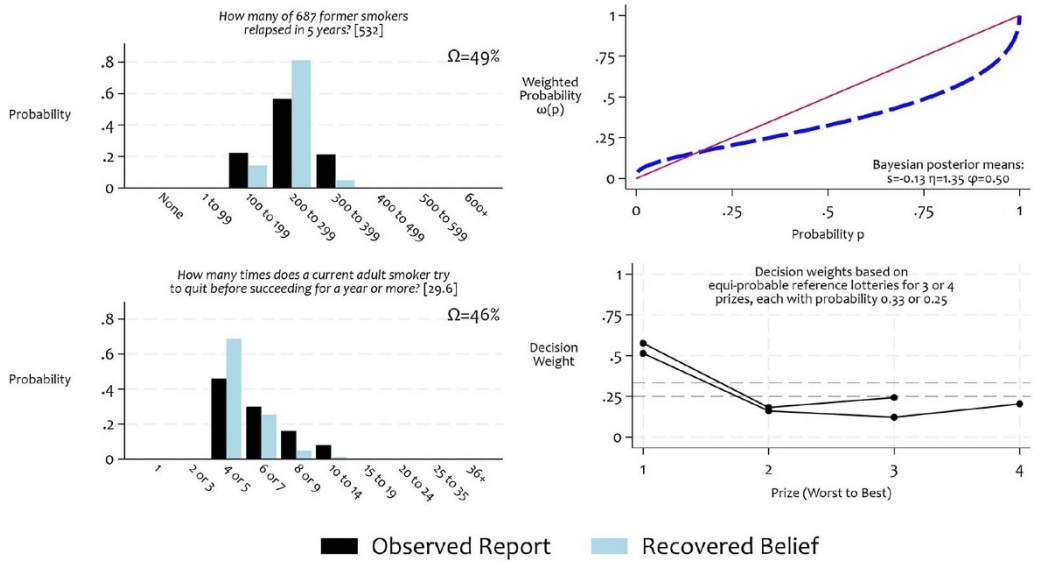


Figure 8. Reports, beliefs and risk preferences for subject #377

reports. The subject has another modestly concave utility function, with $s = 0.13$, so we would not expect that to plays much of a factor in explaining these observed reports. Again, focus on the beliefs, and then knowing these risk preferences, explain the observed reports. In both questions the observed *reports* are putting far less weight on the modal belief, and indeed all other beliefs except the "worst" belief in terms of QSR payoffs. So probability weighting here takes tokens away from the higher-probability beliefs, and towards the lower-probability beliefs, which is what we see in the token reports for the distributions as a whole.

Figure 9 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 belief distributions. This is what Figure 9 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.[27] For reference, in each panel of Figure 9 we also display a vertical red bar to the right showing the interval containing the true outcome, allowing an easy evaluation of the accuracy of the belief distributions of each individual. For further reference, Figure 9 also displays a dashed horizontal 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 9, who are the subjects whose beliefs are reported in detail in Figures 7 and 8. For all but one of the questions, 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. By contrast, the subjects in the remaining rows had beliefs for one or both questions that placed the weight of their beliefs around the true outcome. For subjects #123, #155 and #359 this was the case for the question about quit attempts, and for subject #151 this was the case for the question on relapses. In

[27]This credible interval is generated directly from the 100,000 simulated values for the belief for each outcome by that subject.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

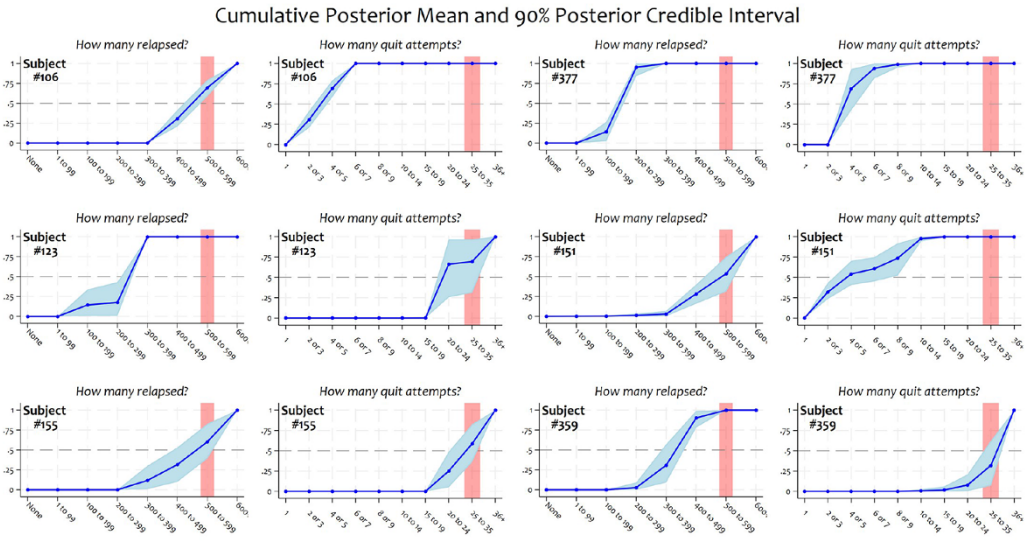


Figure 9. Cumulative posterior probability of beliefs in relation to true outcome for select subjects

all other cases, individual subjects exhibit clear bias with respect to the true addictiveness of smoking, underestimating the risks of addiction.

The information in Figure 9 uses individual subjects to illustrate how we can construct a more complete Bayesian tests of the hypothesis of biased beliefs for all subjects, 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.[28]

Detailed results for each of the 383 subjects, such as those shown in Figures 7, 8 and 9, are available to the researcher for use and evaluation.[29] Figure 10 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 10. 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 indeed 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. One corollary of our relentless focus on results at the level of the individual, and for belief distributions rather than belief scalars, is a need to pair displays such as those on the top panel of Figure 10 with those on the bottom panel of Figure 10.

[28] This interval is often referred to as a "region of practical equivalence" (ROPE) from a Bayesian perspective: see Kruschke (2018) for an exposition. In our case we could defined a ROPE as any outcome bin either side of the bin containing the true outcome.

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

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

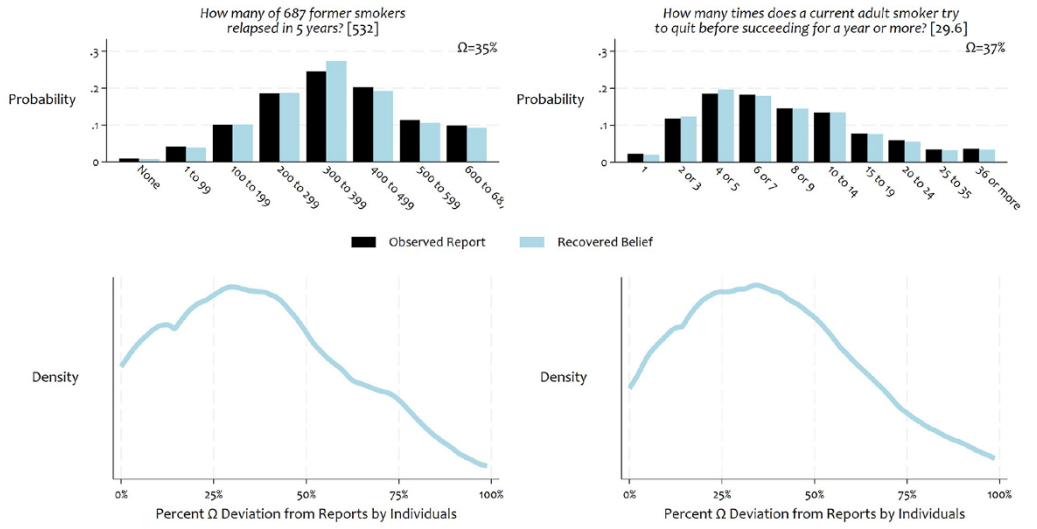


Figure 10. Pooled reports and beliefs, and distribution of individual Ω deviations from reports

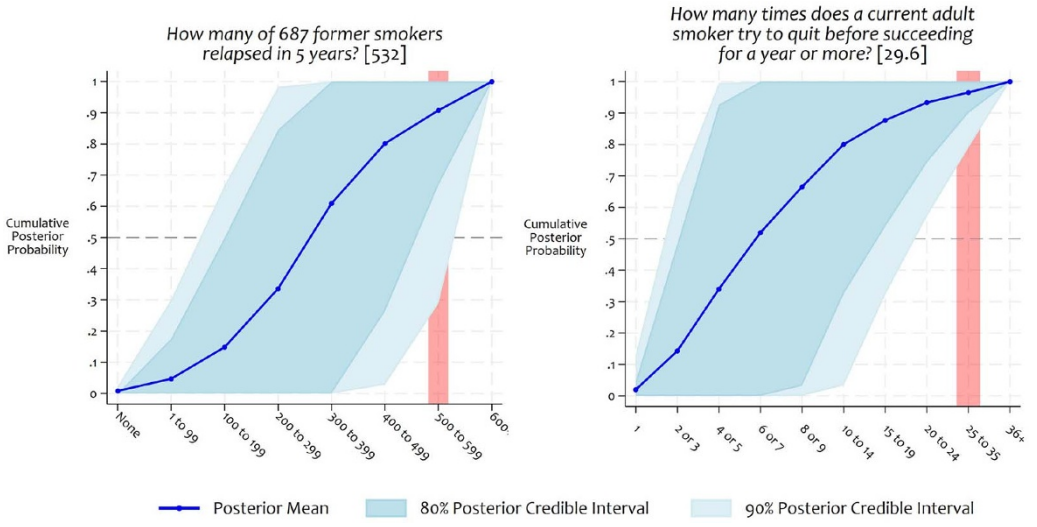


Figure 11. Cumulative posterior probability of beliefs in relation to the true outcome for all subjects

Finally, Figure 11 shows a more informative display[30] of 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.[31] We stress, again, that the results in Figure 9 vividly illustrate that stronger inferences are available for individuals.

[30] More informative than the top panels of Figure 10.

[31] These are exact values, which can be proximately inferred from Figure 11 as well.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

The results in Figures 9, 10 and 11 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.

6. 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 for 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 two applications by evaluating subjective belief questions about "competitiveness" in a canonical example from experimental economics, and by evaluating 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 wide class of risk preferences, enhancing the practicality of inferring subjective belief distributions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2026.10057.

Acknowledgements. We are grateful to Andre Hofmeyr, seminar participants, two referees, and the Editor for comments. All data and code to replicate our analyses in *Stata* are at https://osf.io/6kxhs.

References

Allen, F. (1987). Discovering personal probabilities when utility functions are unknown. *Management Science*, *33*(4), 452–454.

Andersen, S., Harrison, G. W., Hole, A. R., Lau, M., & Rutström, E. E. (2012). Non-linear mixed logit. *Theory and Decision*, *73*(1), 77–96.

Andersen, S., Fountain, J., Harrison, G. W., & Rutström, E. E. (2014). Estimating subjective probabilities. *Journal of Risk & Uncertainty*, *48*(3), 207–229.

Balcombe, K., & Fraser, I. (2025). Flexible estimation of parametric prospect models using hierarchical Bayesian methods. *Experimental Economics*, *28*(2), 354–378.

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, *96*(4), 675–700.

Berg, J. E., Daley, L. A., Dickhaut, J. W., & O'Brien, J. R. (1986). Controlling preferences for lotteries on units of experimental exchange. *Quarterly Journal of Economics*, *101*(2), 281–306.

Borland, R., Partos, T. R., Yong, H. -H., Cummings, K. M., & Hyland, A. (2011). How much unsuccessful quitting activity is going on among adult smokers? Data from the International Tobacco Control Four Country Cohort Survey. *Addiction*, *107*(3), 673–682.

Chaiton, M., Diemert, L., Cohen, J. E., Bondy, S. J., Selby, P., Philipneri, A., & Schwartz, R. (2016). Estimating the number of quit attempts it takes to quit successfully in a longitudinal cohort of smokers. *BMJ Open*, *6*(6), e011045. https://doi.org/10.1136/bmjopen-2016-011045

Danz, D., Vesterlund, L., & Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, *112*(9), 2851–2883.

Delavande, A., Giné, X., & McKenzie, D. (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, *94*(2), 151–163.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–138.

Gao, X. S., Harrison, G. W., & Tchernis, R. (2023). Behavioral welfare economics and risk preferences: A Bayesian approach. *Experimental Economics*, *26*(2), 273–303.

Harrison, G. W. (2017). Behavioral responses to surveys about nicotine dependence. *Health Economics*, *26*(63), 114–123.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

Harrison, G. W. (2024) Causal inferences over unobservables. CEAR Working Paper 2024-02. Robinson College of Business, Georgia State University; *Journal of Political Economy: Microeconomics*: forthcoming.
Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2013). Inducing risk neutral preferences with binary lotteries: A reconsideration. *Journal of Economic Behavior and Organization*, *94*, 145–159.
Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2014a). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior and Organization*, *101*, 128–140.
Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., & Ulm, E. R. (2015). Eliciting subjective probability distributions with binary lotteries. *Economics Letters*, *127*, 68–71.
Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., & Ulm, E. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior & Organization*, *134*, 430–448.
Harrison, G. W., Morsink, K., & Schneider, M. (2022). Literacy and the quality of index insurance decisions. *Geneva Risk & Insurance Review*, *47*(1), 66–97.
Harrison, G. W. and Phillips, R. D. (2014b). 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.
Harrison, G. W., Ross, D., & Swarthout, J. T. (2026). Gender, confidence, and the mismeasure of intelligence, competitiveness and literacy. *Journal of Political Economy*, *134*(2), 665–730.
Harrison, G. W. and Swarthout, J. T. (2023). Cumulative Prospect Theory in the laboratory: A reconsideration. In G.W. Harrison & D. Ross. (eds.) *Models of risk preferences: Descriptive and normative challenges*. Bingley, UK, Emerald, Research in Experimental Economics.
Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*(3), 984–1001.
Johnson, A. A., Ott, M. Q. and Dogucu, M. (2012). *Bayes rules! An introduction to applied Bayesian modeling*. Boca Raton, CRC Press.
Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.) Burlington, MA, Academic Press.
Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.
Kruschke, J. K., & Vanpaemel, W. (eds.). (2015). Bayesian estimation in hierarchical models. In Busemeyer, J.R., Townsend, J.T., Wang, Z.J., & Eidels, A. (eds.), *Oxford handbook of computational and mathematical psychology*. Oxford, UK, Oxford University Press.
Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, *65*(260), 581–598.
Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.
Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096.
McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and STAN*. Boca Raton, FL, CRC Press.
McKelvey, R. D., & Page, T. (1990). Public and private information: an experimental study of information pooling. *Econometrica*, *58*(6), 1321–1339.
Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much?. *Quarterly Journal of Economics*, *122*(3), 1067–1101.
Nilsson, H., Rieskamp, J., & Wagenmakers, E. -J. (2011). Hierarchical bayesian parameter estimation for Cumulative Prospect Theory. *Journal of Mathematical Psychology*, *55*(1), 84–93.
Offerman, T., Sonnemans, J., van de Kuilen, G., & Wakker, P. P. (2009). A truth-serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. *Review of Economic Studies*, *76*(4), 1461–1489.
Orphanides, A., & Zervos, D. (1995). Rational addiction with learning and regret. *Journal of Political Economy*, *103*(4), 739–758.
Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*(3), 497–527.
Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*(4), 323–343.
Rieger, M. O., & Wang, M. (2006). Cumulative Prospect Theory and the St. Petersburg paradox. *Economic Theory*, *28*(3), 665–679.
Roth, A. E., & Malouf, M. W. K. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review*, *86*(6), 574–594.
Rutström, E. E., & Wilcox, N. T. (2009). Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, *67*(2), 616–632.
Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of American Statistical Association*, *66*(336), 783–801.
Savage, L. J. (1972). *The foundations of statistics*. (2nd ed.) New York, Dover Publications.
Schlag, K. H. and van der Weele, & J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk-neutrality. *Theoretical Economics Letters*, *3*(1), 38–42.
Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society Series B*, *23*(1), 1–25.

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.

Stan Development Team (2025) Stan User's Guide, Version 2.37 Available at: https://mc-stan.org.
Thun, M. J., Carter, B. D., Feskanich, D., Freedman, N. D., Prentice, R., Lopez, A. D., Harge, P., & Gapstur, S. (2013). 50-year trends in smoking-related mortality in the United States. *The New England Journal of Medicine*, *368*(4), 351–364.
Viscusi, W. K. (2002). *Smoke-filled rooms: A postmortem on the tobacco deal*. Chicago, University of Chicago Press.
Wilcox, N.T. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, *162*(1), 89–104.
Wilcox, N. T. (2023). Unusual estimates of probability weighting functions. In G.W. Harrison & D. Ross (eds.), *Models of risk preferences: Descriptive and normative challenges*. Bingley, UK, Emerald, Research in Experimental Economics.

Cite this article: Harrison, G. W., Monroe, B. A., & Ulm, E. R. (2026). Recovering subjective probability distributions: a Bayesian approach. *Experimental Economics*, 1–25. https://doi.org/10.1017/eec.2026.10057

Downloaded from https://www.cambridge.org/core. 11 Nov 2026 at 20:28:41, subject to the Cambridge Core terms of use.