Evaluating the Expected Welfare Gain From Insurance

Glenn W. Harrison
Jia Min Ng

Abstract

Economic theory tells us how to evaluate the expected welfare gain from insurance products on offer to individuals. If we know the risk preferences of the individual, and subjective beliefs about loss contingencies and likelihood of payout, there is a certainty equivalent of the risky insurance policy that can be compared to the certain insurance premium. This simple logic extends to nonstandard models of risk preferences, such as those in which individuals exhibit "optimism" or "pessimism" about loss contingencies in their evaluation of the risky insurance policy. We illustrate the application of these basic ideas about the welfare evaluation of insurance policies in a controlled laboratory experiment. We estimate the risk preferences of individuals from one task, and separately present the individual with a number of insurance policies in which loss contingencies are objective. We then estimate the expected consumer surplus gained or foregone from observed take-up decisions. There is striking evidence of foregone expected consumer surplus from incorrect take-up decisions. Indeed, the metric of take-up itself, widely used in welfare evaluations of insurance products, provides a qualitatively incorrect guide to the expected welfare effects of insurance.

Consider the humble question of the welfare valuation of some new insurance product, such as the "micro-insurance" products being offered and promoted in developing countries. In general, these policies currently are evaluated by the metric

of product take-up.[1] Although take-up is easy to measure, it does not automatically reflect the existence or size of the welfare gain of the insurance product to the insured. An insurance product usually involves the individual[2] giving up a certain amount of money *ex ante* some event in the expectation of being given some money in the future if something unfortunate occurs. Welfare evaluation, therefore, generally requires that one knows risk and time preferences of the individual, since the benefits of the product are risky, and in the future, while the costs are normally[3] certain and up front. We must also know the subjective beliefs that the individual used to evaluate possible losses.[4]

Of course, there is a "revealed preference" argument that if the product is (not) taken up it was perceived to be a positive (negative) net benefit. But that is only the starting point of any serious welfare evaluation, particularly if one wants to *quantify* the size of the welfare effect. What if the subjective beliefs were biased, in the sense that the individual would revise them if given certain information? What if the evaluation of the product used some criteria other that Expected Utility Theory (EUT)? What if the individual simply made a mistaken decision, given beliefs and risk preferences? Invoking this revealed preference argument implies that one could *never* find a negative welfare from *any* insurance decision!

Instead of making *a priori* assumptions about those preferences that are likely to be wrong, we can use controlled experiments to estimate individual preferences,

---

[1]Many recent field experiments that evaluate alternative insurance products focus exclusively on take-up as a proxy for welfare: for example, see Giné, Townsend, and Vickrey (2007, 2008), Cole et al. (2013), Dercon et al. (2014), Cole, Stein, and Tobacman (2014), Banerjee, Duflo, and Hornbeck (2014), and Cai et al. (2015). Virtually no attempt is made to design products that reflect the risk preferences of individuals. One example of the casual nature of judgments in this area comes from Giné, Townsend, and Vickrey (2008, p. 544), describing how the premium was set: "The policy premium was initially benchmarked on projected payouts using historical rainfall data (at least 25 years of data for each rain gauge were used). The premium was calculated as the sum of the expected payout, 25 percent of its standard deviation, 1 percent of the maximum sum insured in a year, plus a 25 percent administrative charge and 10.2 percent government service tax. In some cases the premium dictated by this formula was then reduced, because it was believed to exceed farmers' willingness to pay." After all of the formal actuarial arithmetic, we scratch our heads and just change things based on some hunch. To justify being puzzled by low take-up, Banerjee, Duflo, and Hornbeck (2014, p. 292) refer to "evidence of a strong need for health insurance," but by this all they mean is evidence that *average* health expenditures exceed the typical premium by a factor of $8.9 = 4670/525$. Such expenditures always exhibit significant skewness, with many at zero, affecting the welfare evaluation of the insurance product. An online appendix details the various welfare metrics used in applied evaluation of insurance products.

[2]One could extend this approach to consider the social welfare evaluation of insurance products for groups of individuals, such as households, villages, or even nations.

[3]Some insurance products in developing countries spread the premium payments over the life of the contract.

[4]There is an unfortunate tendency in many academic evaluations of insurance purchase to assume that individuals somehow know the probabilities that are estimated or guessed at by actuaries.

valuations, and beliefs, and use those estimates in the welfare evaluation of insurance policies. Laboratory experiments provide the ideal environment to set out all of the information and behavior we need to observe in order to draw inferences about welfare.[5] Once we move to the field and consider naturally occurring data, we will then immediately realize what information is missing if we want to make interesting welfare evaluations. In this sense, laboratory and field experiments are complements (Harrison and List, 2004).

The "Theory" section presents the basic theory of insurance demand to be tested. The "Design Implications" section draws implications for the design of experimental tasks to evaluate welfare. The "Experiment" section presents the laboratory design needed to minimally address the question, and the "Results" section presents the results. The "Implications, Extensions, and Limitations" section reviews limitations of the exercise and immediate extensions, and the "Conclusion" section concludes.

THEORY

Decision-Making Models

Assume that agents behave as if using an EUT or rank dependent utility (RDU) model. These two variants are sufficient to illustrate the issues of interest here, and can be extended to other models.

Assume that utility of income is constant relative risk aversion (CRRA), defined by

$$U(x) = x^{(1-r)}/(1-r), \qquad (1)$$

where x is a lottery prize and $r \neq 1$ is a parameter to be estimated. Thus r is the coefficient of CRRA for an EUT individual: $r = 0$ corresponds to risk neutrality, $r < 0$ to risk loving, and $r > 0$ to risk aversion.

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 induced by the experimenter, so expected utility (EU) is simply the probability weighted utility of each outcome in each lottery i:

$$EU_i = \Sigma_{j=1,J} \left[p(x_j) \times U(x_j) \right]. \qquad (2)$$

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 Equation (1) considered for EUT.[6] To calculate decision weights under RDU one replaces expected utility defined by Equation (2) with RDU

---

[5]An online appendix reviews previous laboratory experiments on insurance demand.

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

$$\text{RDU}_i = \Sigma_{j=1,J} \left[w\big(p(x_j)\big) \times U(x_j) \right] = \Sigma_{j=1,J} \left[w_j \times U(x_j) \right], \tag{3}$$

where

$$w_j = \omega\Big(p_j + \ \dots \ + p_J\Big) - \omega\Big(p_{j+1} + \ \dots \ + p_J\Big) \tag{4a}$$

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

$$w_j = \omega\Big(p_j\Big) \tag{4b}$$

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

We consider three popular probability-weighting functions. The first is the "power" probability-weighting function considered by Quiggin (1982), with curvature parameter γ

$$\omega(p) = p^\gamma. \tag{5}$$

So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation. Convexity of the probability-weighting function when $\gamma > 1$ is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p$ for all 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)

$$\omega(p) = p^\gamma / \big(p^\gamma + (1 - p)^\gamma\big)^{1/\gamma}. \tag{6}$$

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

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

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

Welfare Gain Evaluation

If the subject is assumed to be an EUT type, the consumer surplus (CS) of the insurance decision is calculated as the difference between the certainty equivalent

(CE) of the EU *with* insurance and the CE of the EU *without* insurance. CS is calculated the same way using the RDU instead of EU if the subject was a RDU type.[7]

Assume a simple indemnity insurance product, which provides full coverage in the event of a loss. We assume an initial endowment of $20, with a 10 percent chance of a $15 one-time loss occurring. If an individual purchased the insurance, she could avoid the loss with certainty by paying the insurance premium up front. There are four possible payoff outcomes. If no insurance is purchased, the individual keeps her $20 if no loss occurs, but is only left with $5 if there is a loss. If insurance is purchased, the individual keeps $20 less the premium if no loss occurs, and still keeps $20 less the premium if the loss does occur.

Using the decision-making models discussed above, the EU or RDU across the two possible states, loss or no loss, can be calculated for each choice, to purchase or not to purchase insurance. The CE from the EU or RDU of each choice can be derived, and the difference between the CE from choosing insurance and the CE from not choosing insurance would then be the expected welfare gain of purchasing insurance for that individual.

Figure 1 shows how this CS from purchasing insurance would vary for an EUT individual following the above example, for premiums ranging from $0.01 to $4.50. Each bar shows the CS for a CRRA coefficient ranging from 0.3 to 0.7, typical values expected for a risk-averse EUT individual in an experiment. We see that the CS is larger if the individual is more risk averse, which follows from the fact that more risk-averse individuals are willing to pay more for insurance. As premiums increase, CS becomes negative, showing that there is a threshold premium above which the subject would experience negative expected welfare from purchasing the insurance product.

Figures 2 and 3 show the same graph, but for RDU using an inverse-S probability weighting function and a power weighting function, respectively. For both models γ

---

[7]Mossin (1968) and Smith (1968) show that if a decision maker is a risk-averse EU maximizer and insurance is actuarially fair, then full insurance is optimal. Since the loading is normally positive in practice, this means that only partial coverage should be observed in practice. Mossin (1968, p. 558) was aware that people choose full coverage quite often, and offered several explanations for this behavior: people behave irrationally and do not bother to calculate the optimal coverage, there can be uncertainty around the value of the insured asset, and people might overestimate the probability of the loss. Additionally, this theoretical result may or may not hold depending on the subjects' preference representation. For instance, Razin (1976) analyze insurance demand using the minimax regret criterion suggested by Savage (1951), and show that if a subject follows this criterion it is optimal to buy full coverage even if the insurance contract is not actuarially fair. Briys and Loubergé (1985) demonstrate that if an individual follows the Hurwicz (1951) criterion then the individual does not insure at all or buys full insurance even if the insurance loading is positive. On the other hand, Lee and Pinches (1988) show that by introducing risk aversion to the Hurwicz criterion, partial insurance can also be optimal. Machina (1995, 2000, p. 56) shows that Mossin's full-coverage result does hold in a very general non-EU framework that assumes probabilistic sophistication. He also recognizes that the robustness of this under his generalized non-EU framework depends heavily on the insured's subjective probability of a loss coinciding with the probabilities used to price the insurance in actuarially fair terms.

Figure 1

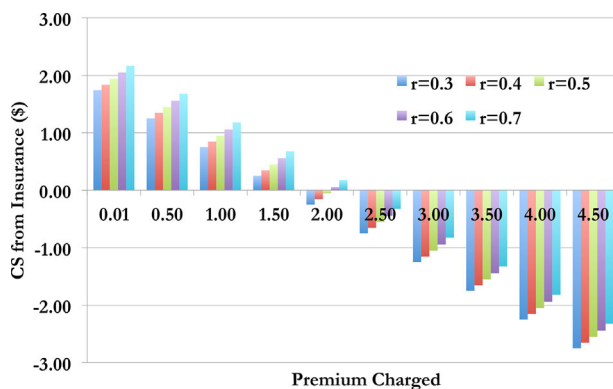Consumer Surplus Across EUT CRRA Coefficients

Figure 2

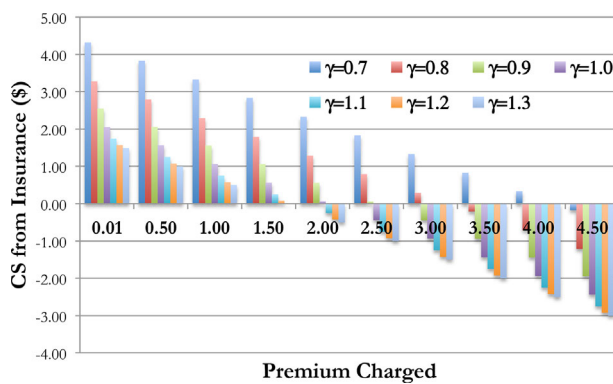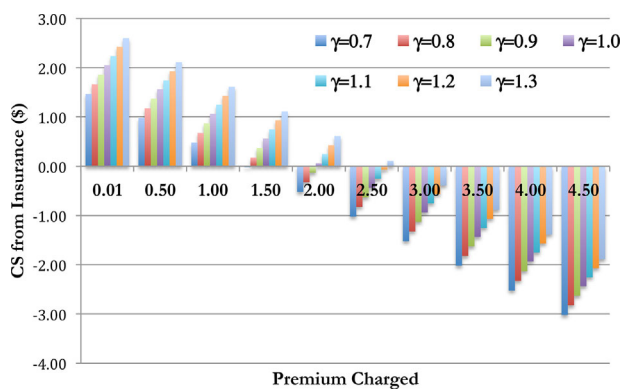Consumer Surplus Across Inverse-S Probability-Weighting Parameter (r = 0.6)

Figure 3

Consumer Surplus Across Power Probability-Weighting Parameter (r = 0.6)

ranges from 0.7 to 1.3, and the CRRA coefficient is held constant at 0.6. As γ increases, CS decreases when the inverse-S probability-weighting function is used, but CS increases when the power function is used. Assigning the right decision-making model, even from this basic set of EUT and RDU specifications, is important for measuring individual welfare evaluation. In general, estimating the right risk parameters for the individual, conditional on the decision-making model, will also affect the identification of the correct decision as well as the opportunity cost of that decision.

Design Implications

In order to evaluate the welfare gain of an individual from insurance conditional on their risk preferences, we need information to determine the following: the classification of the best descriptive decision-making model for the individual, the risk attitudes of the individual given the decision-making model, the type of insurance contract the individual is considering, as well as details of the potential loss the individual is insuring against. This information set includes the loss probabilities the individual is using to access the value of the insurance contract.

In this illustration, we only consider RDU as a sole alternative to EUT. The individual is classified with the model that best fits their risk preferences. This approach can easily be extended to include other decision-making models, as discussed in the "Implications, Extensions, and Limitations" section. For each decision-making model we are considering, the risk parameters specific to that model are estimated using maximum likelihood methods described by Harrison and Rutström (2008).[8]

To keep the theory and design in this experiment transparent we use a plain vanilla indemnity insurance contract. An *ex ante* premium is paid to fully insure a loss should it occur, when it occurs. This approach can be readily extended to investigate the welfare gain from other insurance contracts.

Lastly, to run this experiment we will also need to know about the potential loss the insurance protects against. The wealth amounts for each possible state of nature in this experiment are given, assuming no asset integration with wealth outside the laboratory setting. While the loss probability is clearly given to the subjects in this laboratory experiment, if this experiment was conducted in the field the loss probability might not be objectively clear to the subject. Of course, each subject in the laboratory could still interpret the objective probability subjectively, but that is accounted for in the probability-weighting function of the RDU model. It is also possible that in the field an actuarial probability estimate of loss might exist, but is unknown to the subject.

---

[8]These maximum likelihood methods have a long tradition in experimental economics, such as Camerer and Ho (1994, § 6.1), Hey and Orme (1994), and Holt and Laury (2002, § III). Any econometric method that can produce consistent estimates of the structural parameters of the model, along with standard errors, will suffice. Alternative approaches are reviewed in Harrison and Rutström (2008, § 2).

Our laboratory experiment allows us to control for time preferences by resolving all uncertainty about the effects of insurance purchase in the same session. In general, one would need to add some task to measure time preferences along with risk preferences, following Andersen et al. (2008, 2014). Indeed, Hansen, Jacobsen and Lau (2013) use data from these experiments to evaluate the *present value* of willingness to pay for actual insurance products in Denmark, explicitly allowing for time preferences. They "map" estimated risk and time preferences to certain demographic characteristics for the individuals purchasing actual insurance products, and allow for RDU risk preferences as well as EUT risk preferences. They are forced to assume that individuals know and use the true loss rate for auto, home, and house insurance claims.[9]

EXPERIMENT

The theoretical framework we have employed implies that we need two experimental tasks in order to evaluate welfare from observed take-up. One task is to elicit risk preferences defined over lottery choices, and the other task is to elicit insurance choices of the individual.

We recruited 111 subjects from a database of students across several undergraduate colleges from Georgia State University. All sessions were conducted in 2014 at the ExCEN experimental lab of Georgia State University. Each subject received a show-up fee of $5, and no specific information about the task or expected earnings prior to the session. Each subject undertook both tasks, making 80 binary choices across lotteries in the risk task, and 24 binary choices in the insurance task. The insurance task always came first, and the subject was paid out for it prior to the risk task choices.[10] Each subject was told that one of the 80 risk choices and one of the 24 insurance choices would be selected at random for payment. A standard survey was used to collect individual demographic characteristics. An online appendix contains all instructions.

Each subject is asked to make 80 binary choices between lotteries with objective probabilities, using a standard "pie display" interface. After all decisions were made,

---

[9]They infer that the present value of willingness to pay is marginally higher than the actuarial expected value under EUT, and significantly higher under RDU. Of course, if subjective loss beliefs differ from the actuarial loss probabilities assumed, these results could imply any expected welfare gains or losses.

[10]We adopted this ordering and payment of tasks so that one could say that the estimated risk preferences were elicited independently of the insurance task. One could certainly study the empirical effect of order effects in the obvious manner, although our prior is that this is likely to be empirically unimportant. Paying subjects for the first task prior to the second task ensured that any "wealth effect" was from a known datum, rather than a subjective estimate based on choices. Using data from 63 subjects each making 60 binary choices in the gain domain from Harrison and Rutström (2009), one can directly test if a known initial wealth increment affects (pooled) risk preferences. In their design each subject received a randomly generated wealth increment, between $1 and $10 in $1 increments. Using their statistical specifications, there is no significant effect of these wealth increments on estimated risk preferences using EUT or RDU models.

one of the 80 choices was chosen at random to be played out in accordance with the choices of the subject. Under EUT, this experimental payment protocol provides incentives for truthful binary choices.[11]

The battery of lottery pairs is carefully selected for our purpose of identifying if the individual subject behaves consistently with EUT or some probability-weighting alternative such as RDU.[12] Loomes and Sugden (1998) pose an important design feature for common ratio tests of EUT: variation in the "gradient" of the EUT-consistent indifference curves within a Marschak–Machina (MM) triangle. The reason for this is to generate some choice patterns that are more powerful tests of EUT for any given risk attitude. Under EUT the slope of the indifference curve within an MM triangle is a measure of risk aversion. So there always exists some risk attitude such that the subject is indifferent, and evidence of common ratio violations has virtually zero power (EUT does not, then, predict 50:50 choices, as some claim). All of the lottery pairs implied by their battery have one or both lotteries on the "border" of the MM triangle.

"Border effects" arise in tests of EUT when one nudges the lottery pairs in common ratio tests and common consequence tests into the interior of the MM triangle, or moves them significantly into the interior. The striking finding is that EUT often performs better when one does this. Actually, the evidence is mixed in interesting ways.[13] Camerer (1992) generated a remarkable series of experiments in which EUT did very well for interior lottery choices, but his data were unfortunately from hypothetical choices. These lotteries were well off the border. These lotteries can be contrasted with those in Camerer (1989) that were on the border, and where there were significant EUT violations. But Harless (1992) finds that *just* nudging the lotteries off the boundary did not improve behavior under EUT for real stakes. So one natural question is whether the common ratio tests lead to EUT *not* being rejected when we are in the interior triangle, and to EUT being rejected when we are have choices on the boundary. This seems to be the conclusion from Camerer (1989, 1992), but it is not as clean as one would like. We, therefore, generated 40 "boundary" pairs and 40 "interior" pairs with exactly the same common ratios, making 80 pairs in all.

---

[11]Harrison and Swarthout (2014) discuss the evidence for this experimental payment protocol, particularly when drawing inferences about RDU models. Their findings just make our classifications of subjects as EUT or RDU more conservative with respect to EUT (i.e., we are more likely with this payment protocol to classify subjects as RDU than if the protocol had no effect).

[12]Given that the insurance choices only used a probability of 0.1 for the risky outcome, one might have used a more restrictive battery that just focused on that probability. However, we wanted to have a "standard battery" where properties for estimating RDU models had been demonstrated, and that battery employed a wide range of probabilities.

[13]"Event-splitting" tasks arose out of test of boundary effects, but are different. They arise when separating an event which yields a particular (positive) outcome into two (or more) sub-events increases the attractiveness of the lottery offering that outcome, despite the fact that the likelihood of the outcome occurring remains unchanged. They were first noted by Starmer and Sugden (1993) and Humphrey (1995).

The specific lottery parameters employed are documented in an online appendix. Prizes were $10, $30, and $50 in both lotteries. Each subject received the chosen lotteries in random order, to mitigate any possible order effects. Left-right presentation order was also randomized, as well as the ascending and descending order of presentation of prizes in terms of the dollar value of the prizes.

Every random event determining payouts was generated by the rolling of one or more dice. These dice were illustrated visually during the reading of the instructions, and each subject rolled their own dice.
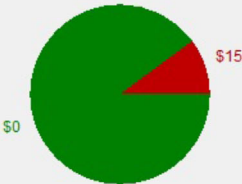
In the insurance task, subjects were asked to express their binary willingness to pay for insurance against a potential loss. The setup of each decision was similar to Laury, McInnes, and Swarthout (2009): each subject is given an initial endowment, the amount of a potential loss, and the probability of the loss occurring. In our experiment these amounts were set at $20, $15, and 10 percent, respectively, for all decisions. Subjects were asked if they would like to purchase insurance for full indemnity for a premium. The premium offered across decisions ranged from $0.20 to $4.80 in 20 cent increments. These premia were offered to each subject in a random order.

Figure 4 shows the interface the subjects used for the insurance choices. The experiment was programmed and conducted with the *z-Tree* software developed by Fischbacher (2007). The choices were described to the subjects as follows:

Figure 4
Insurance Purchase Choice Interface

> In this lottery, there is a 10 percent chance you will experience a loss of fifteen dollars ($15) that corresponds with the red portion of the pie, and a 90 percent chance you will experience no loss ($0) that corresponds with the green portion of the pie. Since you start out with $20, this means there is a 90 percent chance your earnings remain at $20, but there is a 10 percent chance you will lose $15, which would leave you with $5.

> You are given the option to buy insurance to protect yourself against the potential loss in this lottery. You should decide if you want the insurance before you know if a loss will occur. In this example, the insurance will cost you $2.20. This is full insurance, meaning if you purchase the insurance and a loss should occur, the insurance will cover the full loss, and your net earnings will be your initial earnings of $20 less the price paid for the insurance ($2.20), which is $17.80. If you choose to purchase insurance and there was no loss you would still need to pay for the $2.20 insurance, and your net earnings will be $17.80.

One choice out of the 24 insurance choices made by the subject was selected at random to be played out to calculate the subject's actual payoff. In this experiment, only the premium amount was varied across the choices: initial endowment, loss probability, and loss amount were kept constant. These choices could all be varied to include different features of insurance products, such as deductibles and coinsurance.

In addition to the information identified above, we need to make a theoretical assumption that we can indeed identify risk preferences independently of the insurance choice under evaluation. The reason is deceptively simple: in our setting there is almost always some assumption about risk preferences that can rationalize any insurance decision as generating a positive expected welfare gain.[14] It could be that the only models of risk preferences that can rationalize certain decisions require some normative departure from EUT, as in Hansen, Jacobsen, and Lau (2013) and Barseghyan et al. (2013) who stress the role of "probability distortions" akin to the

---

[14]There are settings where this is not true, such as where observed insurance behavior appear to violate elementary requirements of all of the models of risk preferences we consider here, such as first-order stochastic dominance. Or they imply *a priori* implausible levels of risk aversion, as in Sydnor (2010). We have concerns that many studies of naturally occurring insurance choices do not know the *subjective* loss probabilities that guided purchase decisions, and of course that is one good reason to start the welfare evaluation of insurance in a controlled experiment in which these can be induced. Some have attempted to measure risk preferences by observing naturally occurring insurance choices when there are controlled variations in contract features such as deductibles. But one must invariably make strong assumptions about the perception of losses; for instance, Cohen and Einav (2007, p. 746) note, "Two key assumptions—that claims are generated by a Poisson process at the individual level, and that individuals have perfect information about their Poisson claim rates—allows us to use data on (*ex post*) realized claims to estimate the distribution of (*ex ante*) claim rates." Again, we view our controlled laboratory experiments as consistent with some of the conclusions of this behavioral insurance literature, but with far greater control of potential confounds.

RDU models we consider. But in order to identify an expected welfare *loss*, one must conceptually have some independent measure of risk preferences.

We make the simplest possible assumption here that the risk task identifies these risk preferences for the individual, and then we use those estimated risk preferences to evaluate expected welfare gains or losses of that individual's insurance choices. An alternative assumption, of course, is that risk preferences for the same individual differ between our risk task and our insurance task, for whatever "framing" reason one might think of. This assumption might be descriptively correct, and indeed would be implied conceptually if we find, as we do, that risk preferences in the risk task do not explain every insurance choice. But note how our assumption is logically required if we are *ever* to declare some insurance purchase a mistake—we need to have some separate metric for declaring what is and is not a mistake than the choice itself. We consider ways to slightly weaken this assumption later to evaluate the robustness of our analysis of behavior.

RESULTS

Risk Preferences

To evaluate RDU preferences, we estimate an RDU model for each individual. We consider the CRRA utility function as given by Equation (1) and one of three possible probability-weighting functions defined earlier by Equations (5)–(7). For our purposes of classifying subjects as EUT or RDU, it does not matter which of these probability-weighting functions characterize behavior: the only issue here is at what statistical confidence level we can reject the EUT hypothesis that $\omega(p) = p$.

Of course, if the sole metric for deciding if a subject were better characterized by EUT and RDU was the log-likelihood of the estimated model, then there were be virtually no subjects classified as EUT since RDU nests EUT. But if we use metrics of a 10, 5, or 1 percent significance level on the test of the EUT hypothesis that $\omega(p) = p$, then we classify 39, 49, or 68 percent, respectively, of the 102 subjects with valid estimates as being EUT consistent. Figure 5 displays these results using the 5 percent significance level. The left panel shows a kernel density of the *p*-values estimated for each individual and the EUT hypothesis test that $\omega(p) = p$; we use the best-fitting RDU variant for each subject. The vertical lines show the 1, 5, and 10 percent *p*-values, so that one can see that subjects to the right of these lines would be classified as being EUT consistent. The right panel shows the specific allocation using the representative 5 percent threshold. So 5 percent of the density in the left panel of Figure 1 corresponds to the right of the middle vertical line at 5 percent.

Sample Subject Analysis

Risk.   We use the results from a single subject, subject 8, to illustrate the impact of risk preferences on the expected welfare gain from insurance. We first have to determine if subject 8 should be classified as an EUT or RDU decision maker. The log-likelihood value calculated for RDU (-52.3) is better than the log-likelihood of the EUT model (-53.0), so the subject would be classified as RDU with Prelec probability-weighting function by this metric. The difference in log-likelihoods, however, is numerically

Figure 5
Classifying Subjects as EUT or RDU
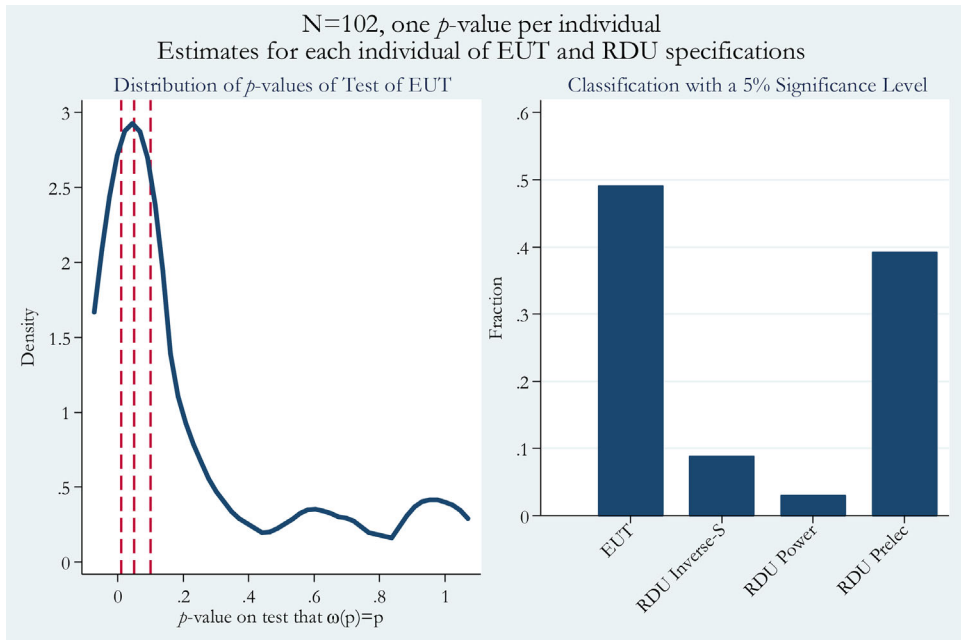
quite small. Once we test for the subject being EUT, the null hypothesis cannot be rejected at the 10, 5, or even 1 percent significance level, since the *p*-value is 0.5.

Figure 6 graphically displays the estimates of the risk parameters of each model for EUT and RDU. The utility curves show that subject 8 had a modestly concave utility function under both EUT and RDU specifications. The utility function is more concave under RDU implying, *ceteris paribus* the effect of probability weighting, greater risk aversion under RDU. Of course, under RDU risk aversion does not just depend on the curvature of the utility function.

The bottom right graph of Figure 6 shows the Prelec probability-weighting function, the best-fit model based only on log-likelihood criteria. Comparing the weighted probability based on the estimates for η and φ against the objective probabilities implies that the subject would overestimate the probabilities of the highest and lowest ranked outcomes, particularly the highest ranked outcome. This overweighting can be clearly seen in the right panel of Figure 7, which shows the impact of subject 8's probability weights on the decision weights for equi-probable outcomes. The subject would overestimate the loss probability, and would be willing to pay a higher premium to purchase the insurance. The estimated parameters for the RDU inverse-S model also reflect the same qualitative finding. The probability-weighting estimate for the RDU Power model, in the top right-hand corner of Figure 6, however, indicates that the subject would be optimistic, which would lower how much she should be willing to pay for insurance *ceteris paribus* the curvature of the utility function.

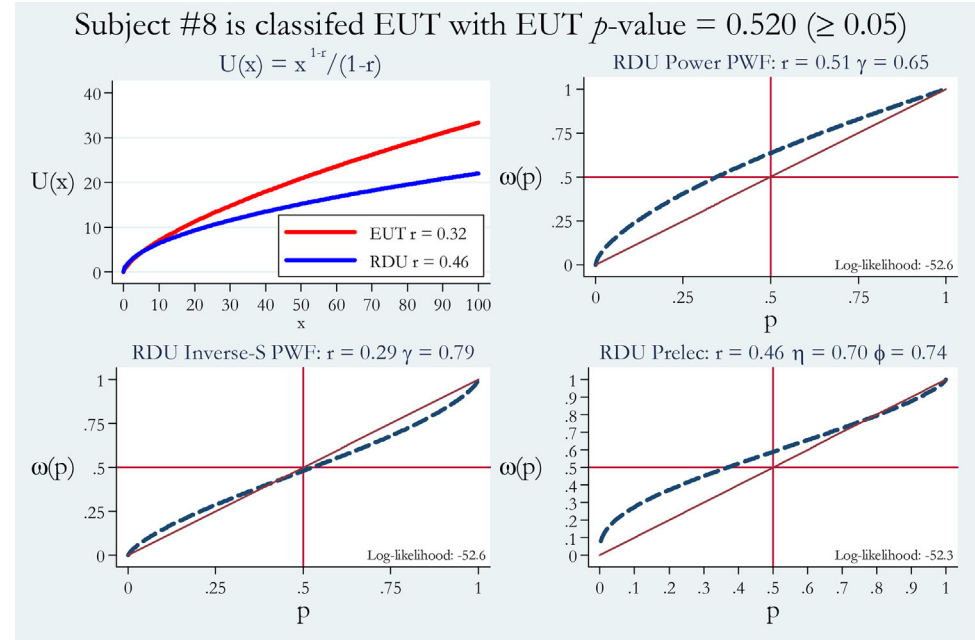**FIGURE 6**

Estimated Risk Parameters for Subject 8

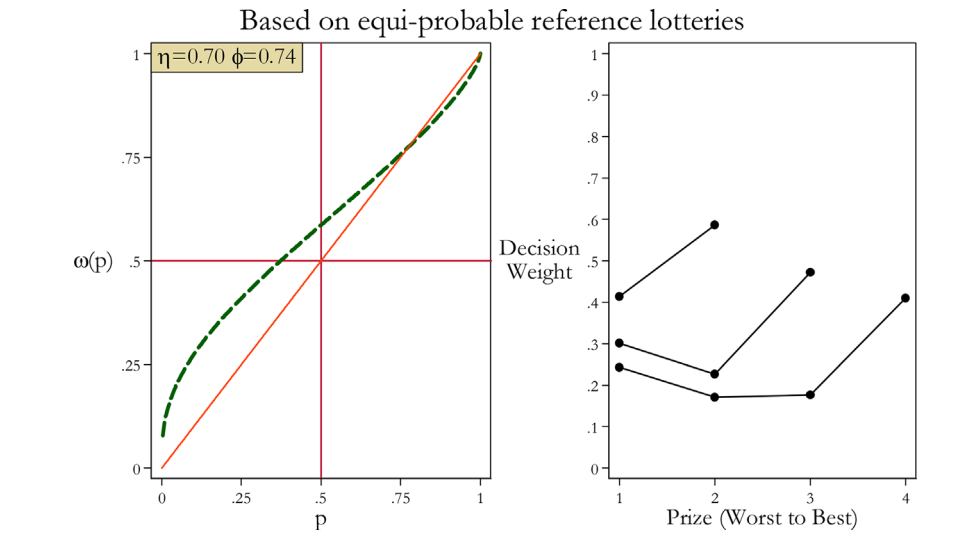**FIGURE 7**

Prelec Probability Weighting and Implied Decision Weights

TABLE 1
Ex Ante Consumer Surplus Across Classifications for Subject 8

Number	Premium ($)	Choice	EUT ($)	RDU Power ($)	RDU Inverse-S ($)	RDU Prelec ($)
1	0.2	1	1.57	1.11	2.72	2.13
2	0.4	1	1.37	0.91	2.52	1.93
3	0.6	1	1.17	0.71	2.32	1.73
4	0.8	0	0.97	0.51	2.12	1.53
5	1	1	0.77	0.31	1.92	1.33
6	1.2	1	0.57	0.11	1.72	1.13
7	1.4	1	0.38	−0.08	1.53	0.94
8	1.6	1	0.17	−0.29	1.32	0.73
9	1.8	1	−0.02	−0.48	1.13	0.54
10	2	1	−0.23	−0.69	0.92	0.33
11	2.2	1	−0.43	−0.89	0.72	0.13
12	2.4	1	−0.63	−1.09	0.52	−0.07
13	2.6	0	−0.82	−1.28	0.33	−0.26
14	2.8	1	−1.02	−1.48	0.13	−0.46
15	3	0	−1.22	−1.68	−0.07	−0.66
16	3.2	0	−1.43	−1.89	−0.28	−0.87
17	3.4	0	−1.63	−2.09	−0.48	−1.07
18	3.6	1	−1.82	−2.28	−0.67	−1.26
19	3.8	0	−2.02	−2.48	−0.87	−1.46
20	4	1	−2.22	−2.68	−1.07	−1.66
21	4.2	0	−2.42	−2.88	−1.27	−1.86
22	4.4	0	−2.63	−3.09	−1.48	−2.07
23	4.6	0	−2.82	−3.28	−1.67	−2.26
24	4.8	1	−3.03	−3.49	−1.88	−2.47

Consumer Surplus. The expected welfare gain from purchasing insurance in this model depends on how the individual is classified, as well as the individual's risk preferences given their classification. The maximum premium to be paid, given a certain potential loss and a certain insurance product, is affected by one's risk preferences. The classification of EUT or RDU will not only affect the breakeven price of the insurance for the individual, it will also affect the opportunity cost of the insurance decision. Table 1 shows the CS calculated for varying premia across the different behavioral models, should subject 8 choose to purchase insurance. For each decision, the subject started with an initial endowment of $20, a 10 percent chance of losing $15, and any insurance purchase would fully cover the loss. The choices are binary, where 1 indicates that the subject chose to purchase insurance and 0 indicates that the subject chose not to.

If the CS is positive in columns 5–8 of Table 1, the correct choice would be for the subject to purchase insurance, assuming the subject's risk behavior adheres to that model. For instance, since the sample subject 8 is classified as EUT at the 5 percent significance level, based on the results in the EUT column, the optimal choice would be for the subject to purchase insurance if the premium was below $1.80, but not to

purchase insurance if the premium was $1.80 or higher. We refer to this as the switch point of an individual's decision.

The varying levels of risk aversion the subject displays can be seen in the premium level at which the subject should switch her insurance purchase decision according to her risk preferences. If the subject was classified as RDU Prelec or RDU inverse-S, the subject would be relatively more pessimistic by overweighting the probability of a loss, compared to the same person being classified as EUT. This is reflected in the larger expected welfare gain from insurance under these two models, seen in the higher switch points, $2.40 and $3.00, respectively.
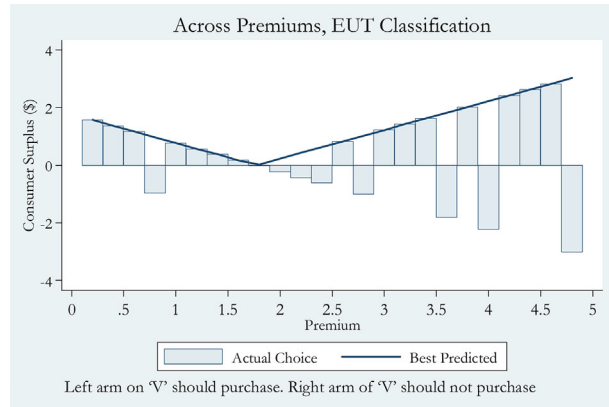
Using different models to model the subject's risk preferences could lead to different recommendations for insurance purchase. For subject 8, within the premium range of $1.40–$2.80, the different models recommend different insurance decisions. For purchase decision 10, with a premium of $2, basing the decision on the RDU inverse-S and RDU Prelec models, the correct decision would be for subject 8 to purchase insurance. However, the correct decision would be for her to not purchase insurance if the subject was classified as EUT or RDU Power. Even for a smaller premium, where the correct decision should unanimously be to purchase insurance, the expected CS amount would differ across the different models. If the subject makes an incorrect decision, we can calculate the size of the foregone CS according to each decision-making model; hence, this information from the smaller premia should not be ignored.

Actual Purchase Decisions

Individual Level.   Expected welfare gain is foregone if the subject chooses to purchase insurance when that decision has a *negative* CS, and similarly when the subject chooses not to purchase insurance when the decision has a *positive* CS. Figure 8 compares the expected welfare gain from each decision with the actual decisions made by subject 8, based on her EUT classification. It shows that the subject has foregone $10.37 out of a possible $31.36 of expected welfare gain from insurance. Subject 8's total expected welfare gain for all 24 decisions was $10.62; hence, the efficiency for subject 8, in the spirit of the traditional definition by Plott and Smith (1978), is 33.9 percent. In this experiment, the efficiency is the expected CS given the subject's actual choices and estimated risk preferences as a percent of total *possible* expected CS given her predicted choices and estimated risk preferences. The efficiency metric is defined at the level of the individual subject, whereas the expected welfare gain is defined at the level of each choice by each subject. In addition, efficiency provides a natural normalization of expected welfare gain on loss by comparing to the maximal expected welfare gain for that choice and subject. Both metrics are of interest, and are complementary.

Figure 9 shows the CS from the decisions that another sample subject 70 has made. Subject 70 is classified as RDU Prelec, and theory predicts that the subject's switch point, the point where she should change her decision from choosing to purchase insurance to choosing not to purchase insurance, should occur at the $1.20 premium. Instead the subject has delayed the switch point to when the premium is $2.20. This

FIGURE 8
Cosumer Surplus Based on Subject 8's Choices

Across Premiums, EUT Classification

Left arm on 'V' should purchase. Right arm of 'V' should not purchase

delay has contributed to the foregone expected welfare gain totaling $5.10 from the decisions with smaller premium amounts. Subject 70's total *possible* gain from all decisions was $40.86; hence, her total expected welfare gain was $30.66 and the efficiency of her choices was 75.0 percent.

Aggregate Level.   Expanding this analysis to look across all subjects, Figure 10 shows the kernel density of the expected CS of each decision made. We find that 49 percent of decisions made resulted in *negative* predicted CS. The distribution of expected CS from these results is similar to the distribution if the insurance purchase decision was randomized. If insurance was randomly purchased, 50 percent of decision made would result in negative predicted CS, and average expected welfare gain would not

FIGURE 9
Consumer Surplus Based on Subject 70's Choices

Across Premiums, RDU Prelec Classification

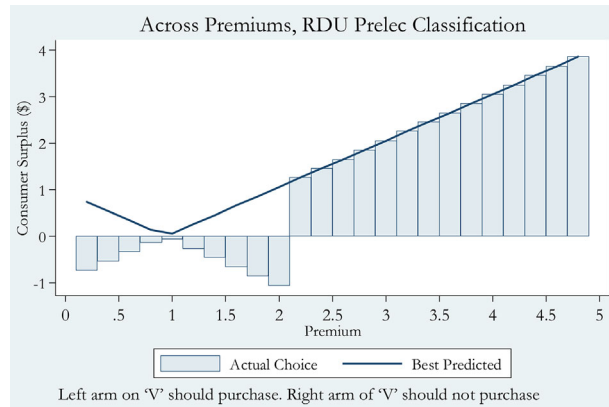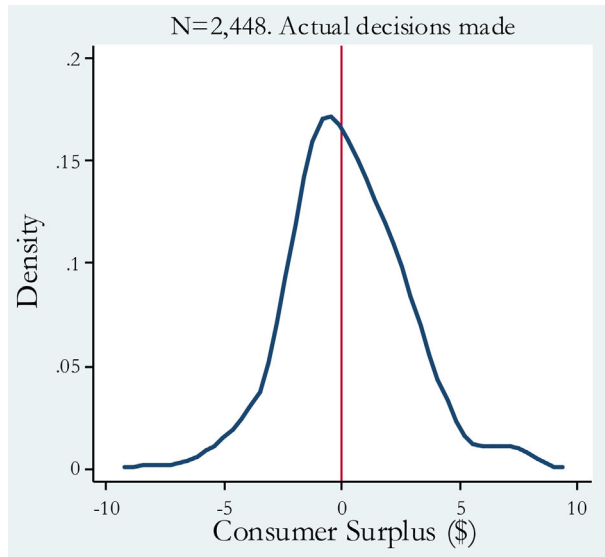Left arm on 'V' should purchase. Right arm of 'V' should not purchase

FIGURE 10
Distribution of Consumer Surplus

be significantly different from $0. Although the average expected welfare gain of $0.27 from actual decisions made is statistically greater than zero at a p-value of less than 0.001, there are still a large proportion of decisions where take-up is not reflecting the welfare benefit of the insurance product to the individual. The efficiency of all decisions made is only 14.0 percent.

Figure 11 shows the distribution of efficiency of decisions made by each individual. The modal efficiency is slightly less than 50 percent, and a significant proportion of individuals make decisions that result in negative efficiency. In other words, these subjects have made choices that resulted in a larger expected welfare loss than the choices that resulted in any expected welfare gain.

How different are observed choices from those that one would expect if subjects chose to purchase insurance at random? We randomly decide 1,000 times whether each subject would purchase or not with a probability of ½ for each outcome, and not surprisingly find that the corresponding distributions of Figures 10 and 11 are symmetric around zero and bell shaped. For the efficiency measure, this immediately indicates that our welfare calculations provide different estimates than if subjects chose at random, since the observed distribution in Figure 11 is skewed, with a mean of 0.13 and a median of 0.24. For the welfare measures, we cannot say at this aggregate level if there is any difference from our welfare estimates. But it is easy to show that our welfare estimates allow one to identify systematic differences in efficiency for certain subjects. For instance, black subjects are 10.4 percentage points (pp) more likely to make an insurance take-up decision with a positive welfare gain, with a 95 percent confidence interval between 2.6 and 18.3 pp. As a result, their welfare gain from insurance decisions is $0.55 higher per choice, with a 95 percent confidence

FIGURE 11
Distribution of Efficiency of Choice
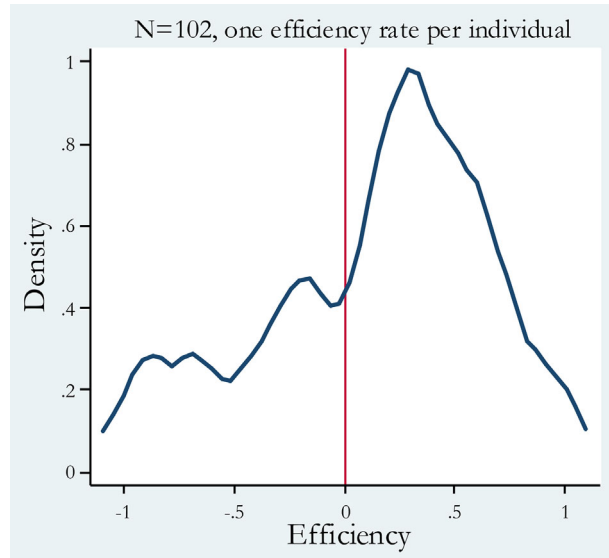

interval between $0.07 and $1.04. This insight is not obtained, by definition, if all decisions are made at random.

Figure 12 shows the breakdown of the number of insurance decisions that were predicted to lead to take-up or not, compared to the actual take-up. The predictions for each subject reflect the best model of risk preferences for that subject. Only 881 out of the 2,448 insurance purchase decisions, or a third of the decisions, were predicted to lead to take-up; however, 62 percent or 1,509 of the actual decisions made were to purchase insurance. More actual decisions were made to purchase insurance than to not purchase insurance regardless of predicted decisions. Fisher's Exact test, however, with a p-value of less than 0.001 for this two-sided test, shows that we can reject the null hypothesis of no significant difference in predicted and observed take-up.

The black bar in Figure 12 shows the predicted choices consistent with what was observed, and the gray bars show the predicted choices that were inconsistent with what was observed. If take-up is a reliable metric for evaluating welfare, even at the crude level of "getting the sign right" about purchase decisions, the black bars should account for the bulk of decisions. They do not. There are 909 decisions out of the 2,448 decisions made where subjects chose to purchase insurance that resulted in foregone expected welfare gain. From a policy making point of view, it is critical to look into the factors that contribute to this significant proportion of negative social welfare gain and what can be done to shrink it, either by reducing the size of the distribution in Figure 10 on the left, or by shifting the entire distribution toward the right.

Proportion of Actual Take-Up to Predicted Choices
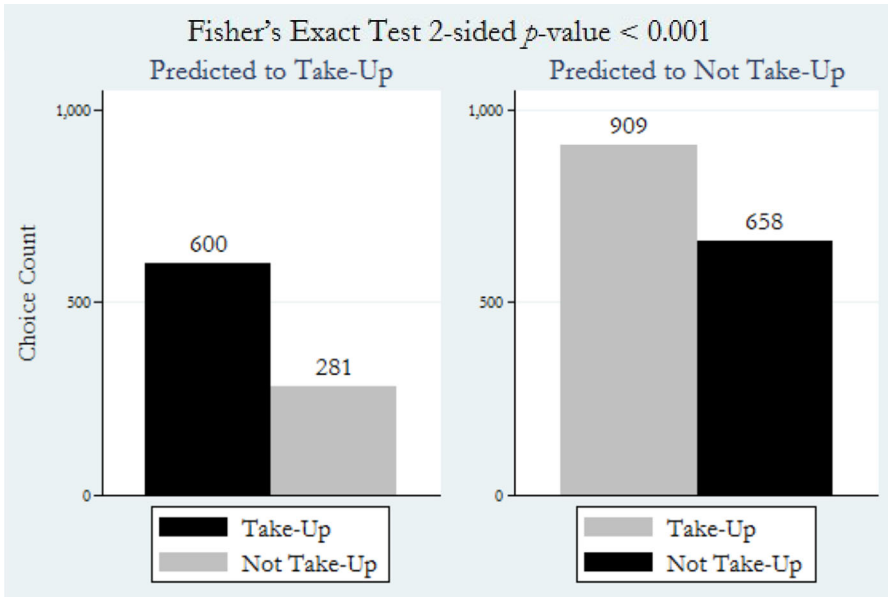
Our approach allows us to say something about what types of decisions led to welfare losses, and if we can identify certain demographics that are more likely to make those types of decisions. To illustrate, consider the welfare losses that arise from excess take-up: someone deciding to purchase insurance when our analysis implies a welfare loss from that decision. Out of all purchase decisions, 60 percent were associated with a welfare loss. Of those, women have a 9.8 pp higher chance than men of making such "excess purchase" errors, with a 95 percent confidence interval between 0 and 20 pp; when we consider the marginal effect of gender, controlling for other demographics, this estimated effect is 11.8 pp with a 95 percent confidence interval between 1 and 23 pp. In principle, this type of information allows one to structure interventions to improve decisions by targeting certain demographic groups and certain types of errors.

Bootstrap

These calculations of expected welfare are conditional on estimates of risk preference, which in turn have estimated standard errors. We allow for these errors in the estimates and bootstrap the effects on calculated welfare. Using the bootstrapped calculated welfare to predict the beneficial insurance choice, Figure 13 shows the comparison of the actual take-up to the bootstrapped predicted choices at a 99 percent significance level. There is a slight difference from the results when point estimates were used. Out of the 939 choices to not purchase insurance, 6 more decisions were estimated to result in foregone expected welfare gain, but 34 less decisions out of the

FIGURE 13
Proportion of Actual Take-Up to Bootstrapped Predicted Choices
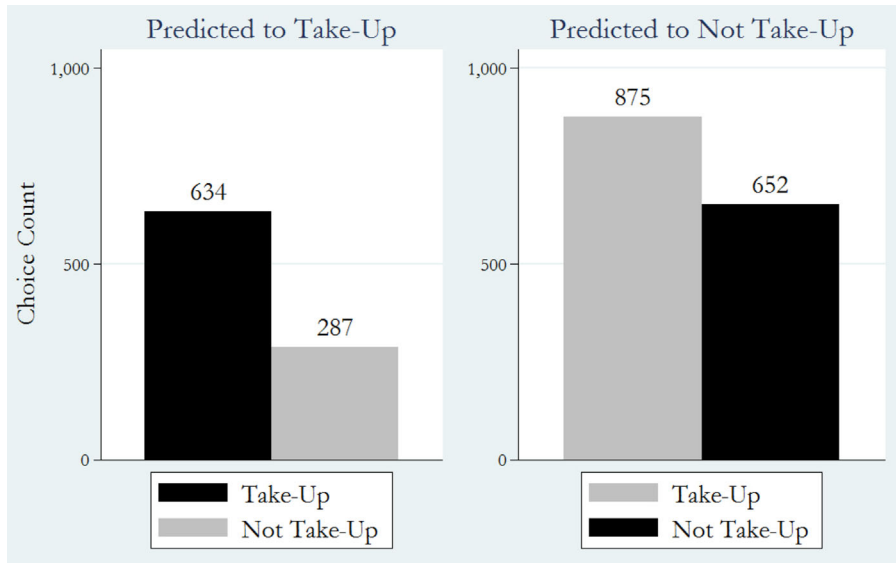
1,509 decisions to purchase insurance were estimated to result in foregone expected welfare gain.

Assuming a multivariate normal distribution on the risk parameters, 500 draws on the risk parameters for each individual were used to calculate the expected CS for each decision. Each decision was tested to determine if they were statistically significantly "incorrect." In other words, for decisions where insurance was actually purchased, was the expected CS significantly negative? And if insurance was not actually purchased, was the expected CS significantly positive?

Even after allowing for bootstrapping of the calculated welfare, the conclusion remains the same. Given the best-fitting decision-making model and risk preferences, a significant proportion of decisions made result in negative expected welfare gain.

IMPLICATIONS, EXTENSIONS, AND LIMITATIONS

Implications

Even though coming from deliberately stylized laboratory tasks, where one can methodologically examine behavior under controlled conditions, these results have significant implications for the evaluation of insurance products.

Many evaluations of insurance products use the metric of take-up, as we noted earlier. We have shown that this simply generates the wrong answer: many people take up a product when they should not, and fail to take it up when they should. Of even greater

significance, take-up is silent on the size of the welfare cost of suboptimal decisions. Even if it managed to "sign" the correct and incorrect decisions, we have no way of determining if a large fraction of incorrect take-up decisions are *de minimis* in terms of consumer welfare.

It follows that anyone that attempted to "nudge" behavior toward more take-up could very easily be nudging people in the wrong direction in consumer welfare terms. It is presumptuous to assume that the subjective guesses of actuaries can be used to substitute for the subjective evaluations of individual agents, but that presumption is implied by evaluations that solely determine success by increased take-up and the presumption that every decision maker is a risk-averse EUT type. This general methodological problem with nudges is well known, and acknowledged by the original proponents of the approach; for instance, see Thaler and Sunstein (2008, ch. 17).

From a normative perspective, a great deal of attention has been devoted to design better insurance *products*. We argue that comparable attention should be devoted to designing better insurance *decisions*.[15] Of course, what many behavioral economists call better products, worthy of a regulatory nudge here or there, are really better decision scaffolds to facilitate better decisions. We see no tension here, just the need to have a clear and structured ability to say something about the welfare effect of product innovations *per se* and the decision process surrounding the product.

Understanding Behavior Toward Other Insurance Products

This approach of measuring expected welfare gain can easily be extended beyond the current basic indemnity product. Other features of insurance, such as deductibles and coinsurance, can be modeled in to see how they affect the expected welfare gain and if take-up accurately reflects these preferences per standard predictions. Alternative insurance products can also include index insurance to investigate how well this model works for basis risk, insurance that covers catastrophic risk, or insurance that covers relatively high probability, low loss risks.

One strength of our approach is that one can rigorously identify which axioms of a normative model of risk preferences fail when one observes expected welfare losses. For instance, are the subjects that suffer losses when faced with an index insurance product those for whom the Reduction of Compound Lotteries axiom fails behaviorally? Precise characterizations of such failures can be identified in experiments (e.g., Harrison, Martínez-Correa, and Swarthout, 2015), just as the lottery battery employed here allows us to identify behavioral failures of the Compound Independence axiom. Thus our approach provides a way to *structurally* identify the behavioral basis of suboptimal insurance decisions.

---

[15]Grüne-Yanoff and Hertwig (2015) distinguish boosts from nudges along the same lines. We do not see these as mutually exclusive, but the distinction is a useful one if only to "nudge" behavioral economists away from only thinking about policy "tricks" to exploit presumed behavioral biases.

Source Independence

One of the maintained assumptions of our approach is the assumption that we can independently estimate risk preferences from one task and use those risk preferences to infer the welfare gain or loss from insurance choices. We discussed earlier why some such assumption is needed from a conceptual, logical, methodological level if one is to undertake normative evaluations. To operationalize this assumption in a simple manner, we used 80 binary lottery choices to measure risk preferences and applied them to 24 binary insurance choices.

One way to relax this assumption slightly is to estimate risk preferences with the 80 choices over lotteries and 23 of the 24 insurance choices, and use those risk preferences to infer the welfare gain or loss from the 1 "hold-out" insurance choice. Then repeat this exercise for each of the other 23 insurance choices, in each case just leaving out of the estimation sample the single insurance choice being evaluated. This jacknife-inspired approach does not remove the assumption of having separate measures of risk preferences in order to assess welfare gain or loss, but it does significantly weaken it.

More generally, for normative evaluations we do not just seek the "best descriptive model" of risk preferences. Obviously, we would like the normative metric to be a good descriptive model, but it must also be possible to use it to make welfare evaluations with some known normative basis.

Skewness Risk Preferences

One important feature of many insurance contracts of interest is that the distribution of payouts is highly skewed, with protection generally only provided for extreme tail events. In this case, it is particularly important to pay attention to models of risk preferences that allow flexible evaluation of those rare events. At the risk of some simplification, and only under EUT, we can associate the aversion to skewness of outcomes with the third derivative of the utility function (prudence) and aversion to kurtosis of outcomes with the fourth derivative (temperance).[16] If it is skewness we care about, then simple utility functions such as CRRA will do a poor job of characterizing risk preferences.

This issue is of paramount importance in field evaluations of many proposed index insurance products for rainfall. Giné, Townsend, and Vickrey (2007) evaluate the distribution of expected payouts for a product that has been extensively studied in the literature, and that is in many ways typical. They pose the issue well:

---

[16]The association is not precise, even if helpful to nudge research away from a dogmatic focus on *variance* in outcomes and toward *variability* of outcomes. Recent interest originated in attempts to formalize the concept of "downside risk aversion" by Menezes, Geiss, and Tressler (1980). Brocket and Kahane (1992), Chiu (2005, 2010), and Eeckhoudt and Schlesinger (2006) provide careful statements of the connection between concepts of risk aversion and preferences over moments. The title of §4 of Brocket and Kahane (1992) is as blunt as one can be in terms of general statements: "$U'' < 0$ and $U''' > 0$ Are Not Related to Variance Avoidance or Skewness Preference."

> Does the insurance contract pay off regularly, providing income during periods of moderately deficient rainfall? Or does it operate more like disaster insurance, infrequently paying an indemnity, but providing a very high payout during the most extreme rainfall events? Our evidence suggests the truth is closer to the second case (p. 1248).

They later calculate the distribution of payout amount and the rank of the payout in terms of size. They find,

> The payout is zero up to the 89th percentile, indicating that an indemnity is paid in only 11% of [rainfall season] phases. The 95th percentile of payouts is around Rs 200, double the average premium. In a small fraction of cases (around 1%) the insurance pays the maximum indemnity of Rs 1,000, yielding an average return on the premium paid of 900%. [This finding] suggests that the ... policies we study primarily insure farmers against extreme tail events of the rainfall distribution.

Of course, such events are what we have many insurance products for, but they have sharp implications for the need to use flexible models of risk preferences to undertake reliable welfare evaluations in the domain of practical interest in the field.

Other Models of Risk Preferences or Uncertainty Aversion

Our approach to evaluating expected welfare gain of insurance from reported beliefs can be readily extended beyond EUT or RDU to any model of risk preferences that allows a welfare evaluation. The one constraint, and it is an important one, is to determine the parameters of the appropriate models for an individual independently of the elicitation of insurance preferences.

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

It would not be appropriate, however, to consider any descriptive model of behavior towards risk that did not allow a welfare evaluation. Even if such models were descriptively more accurate than models that allow a welfare evaluation, if they do not imply some measure of consumer surplus then they are of no value for normative evaluation of insurance. As it happens, we do not believe that these heuristic-based models are at all impressive descriptively, but that is another debate for another day.[17]

---

[17]For instance, the popular "priority heuristic" of risky choices proposed by Brandstätter, Gigerenzer, and Hertwig (2006) has an appalling predictive power, as shown by Andersen et al. (2010b, §7).

Field Experiments

The experiment conducted was deliberately limited to the laboratory, where the probability of loss was given. If the experiment was run in the field, additional tasks would be required to estimate the subjective loss probabilities to the individual.

An additional feature of the field, in developed countries and particularly in developing countries, is the added risk of nonperformance. The chance that insurance companies will not deliver on their contracts as promised is not a casual one, as anyone that has spent any time in developing countries can attest. A practical example of these problems is explained by Banerjee, Duflo, and Hornbeck (2014, p. 297):

> And as it turns out, SKS clients were correct *ex post* in not wanting to purchase this particular health insurance policy. Implementation of the insurance was mismanaged by the partnership of SKS and ICICI-Lombard. In our sample of clients, few claims were submitted and very few clients received any reimbursement. By the endline survey, and in our regular monitoring data, very few people report using insurance, largely because clients were never given documentation to be able to use the insurance or clients did not know how to use the insurance. There is no particular reason to think that this was expected by SKS clients *ex ante*, at least beyond the normal pessimism in developing countries about the prospects of formal health insurance. By the time the product was voluntary, however, these failures were probably quite obvious and could explain why only 29 people [out of roughly 5,000] purchased insurance voluntarily. The fact that client pessimism was well-grounded suggests that offering products that do work, and letting people experience them, should come before trying to solve issues like adverse selection that can only arise once insurance actually delivers a valuable service.

We mention these problems because we have a relatively rich theory of the implications of default risk for the demand for insurance (Doherty and Schlesinger, 1990), as well as evidence that "compound risks" provide a particularly acute source of behavioral problems for individuals (Harrison, Martínez-Correa, and Swarthout 2015). These problems can be studied in the laboratory, in stylized fashion, before one dives into the complications of the field.

What Should the Normative Metric Be?

Our statement of efficiency losses takes as given the type of risk preferences each individual employs, and uses that as the basis for evaluating welfare effects of insurance decisions: *periculum habitus non est disputandum*. One could go further and question if the RDU models themselves embody an efficiency loss for those subjects we classify as RDU. Many would argue that RDU violates some normatively attractive axioms, such as the independence axiom. Forget whether that axiom is descriptively accurate or not. If RDU is not normatively attractive then we should do a calculation of CS in which we *only* assume EUT parameters for subjects: we could

estimate the EUT model and get the corresponding CRRA coefficient estimate (we would not just use the CRRA coefficient estimate from the RDU specification). Then we repeat the calculations. For subjects best modeled as EUT there is no change in the inferred CS, of course.

This issue raises many deeper issues with the way in which one should undertake behavioral welfare economics. For now, we take the agnostic view that the risk preferences we have modeled as best characterizing the individual are those that should be used, in the spirit of the "welfarism" axiom of welfare economics. Even though the alternatives to EUT were originally developed to relax one of the axioms of EUT that some consider attractive normatively, it does not follow that one is unable to write down axioms that make those alternatives attractive normatively.

We view this methodological issue as urgent, open, and important. There is a large, general literature on behavioral welfare economics, including Bernheim (2009), Bernheim and Rangel (2009), Manzini and Mariotti (2012, 2014), Rubinstein and Salant (2012), Salant and Rubinstein (2008), and Sugden (2004). Our general concern with this literature is that although it identifies the methodological problem well, none provide "clear guidance" so far to practical, rigorous welfare evaluation with respect to risk preferences as far as we can determine.

CONCLUSIONS

The laboratory is the correct place to start the welfare evaluation of insurance products, since it provides the best chance of controlling the environment and becoming an actual behavior. Every issue that one might have with the manner in which theory is implemented in the laboratory has the same force in the field, but is usually buried by the added weight of confounds that cannot be controlled easily in the field. We have already noted several issues that would arise as we extend our analysis, in terms of complications to the tasks from more interesting insurance products, and the modeling of risk preferences and subjective beliefs. As a way of illustrating how one might operationalize an answer to this question, then, the laboratory is the appropriate place to begin. The only sense in which the laboratory is limited is the nature of the population from which we sample, and there are obvious solutions to those concerns that are well developed in the literature on "artefactual field experiments" (e.g., Andersen et al., 2010a).

We also view the laboratory as the appropriate place to "wind tunnel" the normative welfare evaluation of new products or decision scaffolds. Figures 10 and 11 stand as explicit, rigorous "target practice" for anyone proposing nudges or clubs to improve welfare from insurance decisions.

Our approach has one methodological perspective that might seem alien to many of the currently popular applied evaluations in economics: the use of theory. We reject as incomplete the claim that one can study welfare effects without theory, or with minimal use of theory (Harrison, 2011a). Nor can one study *any* claims of causality that involve latent constructs, such as welfare, without the explicit or implicit use of theory (Harrison, 2011b, 2014). We show, in fact, that conclusions about causality or welfare that try to avoid using latent constructs from theory, such as those focusing on

insurance take-up as the metric of evaluation, can generate the wrong conclusions. The methodological reason for this conclusion is that one needs behavioral latent theoretical constraints at the beginning and end of the causal chain being evaluated: risk preferences and expected welfare.

Substantively, we provide a clear decomposition of the efficiency losses that come from people making poor decisions about insurance products. We find significant welfare losses from the insurance decisions individuals make. Now begins the exciting task of scientifically considering ways to reduce those behavioral welfare losses.

References

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström, 2008, Eliciting Risk and Time Preferences, *Econometrica*, 76(3): 583-618.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström, 2010a, Preference Heterogeneity in Experiments: Comparing the Lab and Field, *Journal of Economic Behavior and Organization*, 74: 209-224.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström, 2010b, Behavioral Econometrics for Psychologists, *Journal of Economic Psychology*, 31: 553-576.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström, 2014, Discounting Behavior: A Reconsideration, *European Economic Review*, 71: 15-33.

Banerjee, A., E. Duflo, and R. Hornbeck, 2014, Bundling Health Insurance and Microfinance in India: There Cannot Be Adverse Selection if There Is No Demand, *American Economic Review (Papers and Proceedings)*, 104(5): 291-297.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum, 2013, The Nature of Risk Preferences: Evidence From Insurance Choices, *American Economic Review*, 103(6): 2499-2529.

Bernheim, B. D., 2009, Behavioral Welfare Economics, *Journal of the European Economic Association*, 7: 267-319.

Bernheim, B. D., and A. Rangel, 2009, Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics, *Quarterly Journal of Economics*, 124: 51-104.

Brandstätter, E., G. Gigerenzer, and R. Hertwig, 2006, The Priority Heuristic: Making Choices Without Trade-Offs, *Psychological Review*, 113(2): 409-432.

Briys, E., and H. Loubergé, 1985, On the Theory of Rational Insurance Purchasing: A Note, *Journal of Finance*, 40(2): 577-581.

Brocket, P. L., and Y. Kahane, 1992, Risk, Return, Skewness and Preference, *Management Science*, 38(6): 851-866.

Cai, H., Y. Chen, H. Fang, and L.-A. Zhou, 2015, The Effect of Microinsurance on Economic Activities: Evidence From a Randomized Field Experiment, *Review of Economics and Statistics*, 97(2): 287-300.

Camerer, C. F., 1989, An Experimental Test of Several Generalized Utility Theories, *Journal of Risk and Uncertainty*, 2: 61-104.

Camerer, C. F., 1992, Recent Tests of Generalizations of Expected Utility Theory, in: W. Edwards, ed., *Utility Theories: Measurements and Applications* (Boston: Kluwer).

Camerer, C. F., and T.-H. Ho, 1994, Violations of the Betweenness Axiom and Nonlinearity in Probability, *Journal of Risk and Uncertainty*, 8: 167-196.

Chiu, W. H., 2005, Skewness Preference, Risk Aversion, and the Precedence Relations on Stochastic Changes, *Management Science*, 51(12): 1816-1828.

Chiu, W. H., 2010, Skewness Preference, Risk Taking and Expected Utility Maximisation, *Geneva Risk and Insurance Review*, 35: 108-129.

Cohen, A., and L. Einav, 2007, Estimating Risk Preferences From Deductible Choice, *American Economic Review*, 97(3): 745-788.

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickrey, 2013, Barriers to Household Risk Management: Evidence From India, *American Economic Journal Applied Economics*, 5(1): 104-135.

Cole, S., D. Stein, and J. Tobacman, 2014, Dynamics of Demand for Index Insurance: Evidence From a Long-Run Field Experiment, *American Economic Review (Papers and Proceedings)*, 104(5): 284-290.

Dercon, S., R. V. Hill, D. Clarke, I. Outes-Leon, and A. S. Taffesse, 2014, Offering Rainfall Insurance to Informal Insurance Groups: Evidence From a Field Experiment in Ethiopia, *Journal of Development Economics*, 106: 132-143.

Doherty, N. A., and H. Schlesinger, 1990, Rational Insurance Purchasing: Consideration of Contract Nonperformance, *Quarterly Journal of Economics*, 105(1): 243-253.

Eeckhoudt, L., and H. Schlesinger, 2006, Putting Risk in Its Proper Place, *American Economic Review*, 96(1): 280-289.

Fischbacher, U., 2007, Z-Tree: Zurich Tool Box for Ready-Made Economic Experiments, *Experimental Economics*, 10(2): 171-178.

Ghiradoto, P., F. Maccheroni, and M. Marinacci, 2004, Differentiating Ambiguity and Ambiguity Attitude, *Journal of Economic Theory*, 118: 133-173.

Gilboa, I., and D. Schmeidler, 1989, Maxmin Expected Utility With Non-Unique Prior, *Journal of Mathematical Economics*, 18(2): 141-153.

Giné, X., R. Townsend, and J. Vickrey, 2007, Statistical Analysis of Rainfall Insurance Payouts in Southern India, *American Journal of Agricultural Economics*, 89(5): 1248-1254.

Giné, X., R. Townsend, and J. Vickrey, 2008, Patterns of Rainfall Insurance Participation in Rural India, *World Bank Economic Review*, 22(3): 539-566.

Grüne-Yanoff, T., and R. Hertwig, 2015, Nudge Versus Boost: How Coherent Are Policy and Theory? *Minds and Machines*. Advance online publication. doi:10.1007/s110123-015-9367-9

Gul, F., 1991, A Theory of Disappointment Aversion, *Econometrica*, 59(3): 667-686.

Hansen, J. V., R. H. Jacobsen, and M. I. Lau, 2013, Willingness to Pay for Insurance in Denmark, *Journal of Risk and Insurance*. Advance online publication. doi:10.1111/j.1539-6975.2013.12011.x

Harless, D. W., 1992, Predictions About Indifference Curves Inside the Unit Triangle: A Test of Variants of Expected Utility, *Journal of Economic Behavior and Organization*, 18: 391-414.

Harrison, G. W., 2011a, Experimental Methods and the Welfare Evaluation of Policy Lotteries, *European Review of Agricultural Economics*, 38(3): 335-360.

Harrison, G. W., 2011b, Randomisation and Its Discontents, *Journal of African Economies*, 20(4): 626-652.

Harrison, G. W., 2014, Impact Evaluation and Welfare Evaluation, *European Journal of Development Research*, 26(1): 39-45.

Harrison, G. W., and J. A. List, 2004, Field Experiments, *Journal of Economic Literature*, 42: 1009-1055.

Harrison, G. W., J Martínez-Correa, and J. T. Swarthout, 2015, Reduction of Compound Lotteries With Objective Probabilities: Theory and Evidence, *Journal of Economic Behavior and Organization*, 119: 32-55.

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

Harrison, G. W., and E. E. Rutström, 2009, Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral, *Experimental Economics*, 12: 133-158.

Harrison, G. W., and J. T. Swarthout, 2014, Experimental Payment Protocols and the Bipolar Behaviorist, *Theory and Decision*, 77(3): 423-438.

Hey, J. D., and C. Orme, 1994, Investigating Generalizations of Expected Utility Theory Using Experimental Data, *Econometrica*, 62(6): 1291-1326.

Holt, C. A., and S. K. Laury, 2002, Risk Aversion and Incentive Effects, *American Economic Review*, 92(5): 1644-1655.

Humphrey, S. J., 1995, Regret-Aversion or Event-Splitting Effects? More Evidence Under Risk and Uncertainty, *Journal of Risk and Uncertainty*, 11: 263-274.

Hurwicz, L., 1951, Optimality Criteria for Decision Making Under Ignorance, Cowles Commission Paper, 370.

Klibanoff, P., M. Marinacci, and S. Mukerji, 2005, A Smooth Model of Decision Making Under Ambiguity, *Econometrica*, 73(6): 1849-1892.

Laury, S. K., M. M. McInnes, and J. T. Swarthout, 2009, Insurance Decisions for Low-Probability Losses, *Journal of Risk and Uncertainty*, 39: 17-44.

Lee, H. J., and G. E. Pinches, 1988, On Optimal Insurance Purchasing, *Journal of Risk and Insurance*, 55(1): 145-149.

Loomes, G., and R. Sugden, 1998, Testing Different Stochastic Specifications of Risky Choice, *Economica*, 65: 581-598.

Machina, M. J., 1995, Non-Expected Utility and the Robustness of the Classical Insurance Paradigm, *Geneva Papers on Risk and Insurance Theory*, 20: 9-50.

Machina, M., 2000, The Robustness of the Classical Insurance Paradigm, in: G. Dionne, ed., *Handbook of Insurance* (Boston: Kluwer Academic Publishers).

Manzini, P., and M. Mariotti, 2012, Categorize Then Choose: Boundedly Rational Choice and Welfare, *Journal of the European Economic Association*, 10: 1141-1165.

Manzini, P., and M. Mariotti, 2014, Welfare Economics and Bounded Rationality: The Case for Model-Based Approaches, *Journal of Economic Methodology*, 21(4): 343-360.

Menezes, C., C. Geiss, and J. Tresseler, 1980, Increasing Downside Risk, *American Economic Review*, 70(5): 921-932.

Mossin, J., 1968, Aspects of Rational Insurance Purchasing, *Journal of Political Economy*, 79: 553-556.

Plott, C. R., and V. L. Smith, 1978, An Experimental Examination of Two Exchange Institutions, *Review of Economic Studies*, 45(1): 133-153.

Prelec, D., 1998, The Probability Weighting Function, *Econometrica*, 66: 95-113.

Quiggin, J., 1982, A Theory of Anticipated Utility, *Journal of Economic Behavior and Organization*, 3(4): 323-343.

Razin, A., 1976, Rational Insurance Purchasing, *Journal of Finance*, 31(1): 133-137.

Rubinstein, A., and Y. Salant, 2012, Eliciting Welfare Preferences From Behavioral Datasets, *Review of Economic Studies*, 79: 375-387.

Salant, Y., and A. Rubinstein, 2008, (A, f): Choice With Frames, *Review of Economic Studies*, 75: 1287-1296.

Savage, L. J., 1951, The Theory of Statistical Decision, *Journal of the American Statistical Association*, 46(253): 55-67.

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

Smith, V. L., 1968, Optimal Insurance Coverage, *Journal of Political Economy*, 76(1): 68-77.

Starmer, C., and R. Sugden, 1993, Testing for Juxtaposition and Event-Splitting Effects, *Journal of Risk and Uncertainty*, 6: 235-254.

Sugden, R., 2004, The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences, *American Economic Review*, 94: 1014-1033.

Sydnor, J., 2010, (Over) Insuring Modest Risks, *American Economic Journal: Applied Economics*, 2(4): 177-199.

Thaler, R., and C. Sunstein, 2008, *Nudge: Improving Decisions About Health, Wealth and Happiness* (New Haven, CT: Yale University Press).

Tversky, A., and D. Kahneman, 1992, Advances in Prospect Theory: Cumulative Representations of Uncertainty, *Journal of Risk and Uncertainty*, 5: 297-323.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendices