The Geneva Risk and Insurance Review (2019) 44:137–175
https://doi.org/10.1057/s10713-019-00042-y

EGRIE KEYNOTE ADDRESS

The behavioral welfare economics of insurance

Glenn W. Harrison[1,2]

Received: 5 September 2019 / Accepted: 14 September 2019 / Published online: 17 October 2019

Abstract
Behavioral economics poses a challenge for the welfare evaluation of insurance products and policy. It demands that we recognize that the descriptive account of behavior toward insurance depends on risk and time preferences that might not be the ones we were all taught, and still teach, and that subjective beliefs might not accord with actuarial assessments of loss probabilities. Challenging as that can be, things become even harder when we jettison naive notions of revealed preferences as the basis for evaluating the individual welfare of insurance decisions. These challenges demand theory, datasets that allow us to identify structural models, datasets that allow us to observe those that do *not* purchase insurance, appropriate econometric methods, and particularly pay close attention to the methodological nonsense that is often used to justify policy interventions.

Keywords Behavioral welfare economics · Insurance · Methodology · Preferences · Subjective beliefs

1 Introduction

Decisions to purchase insurance should be a perfect place to see economic theory at work in general, and behavioral economics at work in particular. We have well-developed descriptive theories of the demand for insurance products. These theories

---

This article was the 30th Geneva Risk Economics Lecture of the European Group of Risk & Insurance Economists.

A debt is owed to long conversations with Don Ross and Elisabet Rutström, and comments from Michael Hoy and participants at the 45th Seminar of the *European Group of Risk & Insurance Economists* in Nuremberg, September 2018.

✉ Glenn W. Harrison
gharrison@gsu.edu

1 Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, USA

2 School of Economics, University of Cape Town, Cape Town, South Africa

extend relatively easily to the insights of behavioral economics.[1] When we turn to normative issues, however, things are not nearly so settled. The informational requirements needed to undertake welfare evaluations are severe, and are well recognized. And when we turn to *behavioral* welfare economics, we face numerous subtleties that place an even higher premium on theoretical rigor. This lecture reviews these normative issues, the existing literature, and the challenges, with emphasis on behavior toward insurance. Despite the emphasis on the need for theory, and the striking lack of theory in much of what passes for policy evaluation these days, the implications for policy are serious, and not just challenging theoretical puzzles.

From a theoretical perspective, one can quickly identify several "behavioral moving parts" in canonical insurance contracts. The first is obviously risk aversion, which can derive from various psychological pathways. The second is, also obviously, subjective beliefs about loss probabilities, as well as nonperformance risk and other basis risk when applicable. The third concerns time preferences, thinking of insurance as an explicitly time-dated contract: in general, I give you a known premium *now* in the expectation that if something happens to me *over the coming year* you will honor that contract and help me mitigate the loss. The fourth involves the interaction of risk and time preferences, in the form of intertemporal risk aversion; as explained below, this is not the same as atemporal risk aversion.

Sections 2 and 3 provide "helicopter tours" of key issues in theory and experiments that are central to evaluating the literature. Relying on a detailed survey in Harrison and Ng (2019), Sect. 4 provides another helicopter tour of *descriptive* behavioral evaluations of insurance. Section 5 begins the real business of closely reviewing normative behavioral evaluations of insurance. Section 6 points to the deeper methodological issues in behavioral welfare economics that need to be given more attention.

The main theme is the tension between "old time welfare economics" and "new behavioral economics." As our theoretical and empirical understanding of behavior has become richer and more nuanced, the normative applications of those insights have, in general, become "dumb and dumber." For example, the popularity of unconditional nudges presumes a striking hypocrisy: I am paternalistically confident that I know what choice is best for you, but I protect myself by claiming that I am a "paternalistic liberal" since you have the freedom to do what you want even when I know that statistically you will do what I nudge you to do.

A secondary theme is that the truly impressive technical skills on display in recent structural models of the welfare effects of insurance policy might easily lead a casual reader to think that they are taking the "new behavioral economics" seriously. A close review makes it clear that they do not, as valuable as they are as intellectual base camps for the mountain of scholarship that remains.

A related theme is that the laboratory experiment provides the perfect place to test out the way in which we might identify and measure the many behavioral moving parts of any rigorous welfare evaluation of insurance. As easy as it is to sleight

[1] Behavioral economics is not just the study of anomalies or irrationality.

such experiments in this age of field experimentation, the numbing absence of deep insights into welfare from these field experiments[2] should be a clear enough warning call. Nonetheless, even in the lab, with the "cleanest breakers" we can imagine, relatively little has yet been done.

2 Theory

Insurance is a staple of any classroom discussion of risk attitudes and risk management. Indeed, it is often used to immediately explain why we should be interested in knowing the risk attitudes of an agent. The very definition of a risk premium, as the amount of money one is willing to leave on the table, in expectation, in order to remove risk, defines willingness to pay for a full indemnity insurance contract with no deductible.

The good news is that the notion of a risk premium is one of the core concepts that different theories of risk preferences actually agree on. Expected Utility Theory (EUT) posits a psychological pathway in which aversion to variability drives a risk preference, where variability can be much more than just variance. Rank-Dependent Utility (RDU) posits an additional psychological pathway in which probability optimism or pessimism can augment, positively or negatively, any risk premium due to aversion to variability. And Cumulative Prospect Theory (CPT) posits yet another psychological pathway on top of these, where sign dependence relative to some reference point affects risk preferences. All agree on the same risk premium, and simply decompose it differently.

Important extensions to these basic insights include considerations of downside risk aversion that differs from the loss aversion of CPT, and is related to literature on "higher order risk preferences;" considerations of "regret" or "disappointment" that can arise from insurance decisions and outcomes; and allowance for multiattribute risk aversion, across insurance product lines or between foreground and background risks.

Theories of time preference range from Exponential discounting to Hyperbolic and Quasi-Hyperbolic models. The differences can best be understood by thinking of the lender of money as having some cost to not having her money for a time period. Exponential discounting assumes a constant variable cost with respect to time and no fixed cost; Hyperbolic discounting assumes a declining variable cost with respect to time and no fixed cost; and Quasi-Hyperbolic discounting assumes a fixed cost and a constant variable cost.[3] An alternative approach from psychology is to view the perception of time horizon as subjective: if the agent perceives

[2] Support for this sweeping claim can be found in Harrison (2011a, b, 2013, 2014a, b) and Harrison and Ng (2019).

[3] The Quasi-Hyperbolic model assumes a rather strange fixed cost, a constant *percentage* of the principal. One can write down models that assume that the fixed cost is a scalar amount of money, or a scalar level of utility.

time units contracting as the horizon gets longer, declining discount rates will arise. Andersen et al. (2014b) review these models, and evidence for them.

The bad news is that virtually all theories of time preference assume an *additive* intertemporal utility function, in which utility over time is a discount factor weighted sum of utility for each distinct period. In this respect, the alternative theories behind the discount factor tend to agree, and also use an additive intertemporal utility function. This seemingly technical assumption, however, has dramatic implications for behavior: it implies that agents are neutral toward risk *over time*, even if they are averse to risk *at a point in time*. In words, agents might be atemporally risk averse to risk resolved at a point in time, but must then be intertemporally risk neutral to risk resolved over time. A nasty corollary is that atemporal risk preferences and time preferences are formally "tied at the hip," in the sense that the intertemporal elasticity of substitution *must* be equal to the inverse of relative risk aversion. This corollary sits uncomfortably with intuition and the stylized data one encounters in aggregate data, forcing problematic calibrations in macroeconomic models. A simple resolution of this impasse is to allow nonadditive intertemporal utility functions, such that interactions between atemporal risk aversion between time periods matter to the agent: see Andersen et al. (2018b) for a review of the theory.

The static theory of subjective beliefs is dominated by Subjective Expected Utility (SEU), which assumes that agents behave as if satisfying the Reduction of Compound Lotteries (ROCL). The effect is that nondegenerate subjective belief distributions can be replaced by the weighted average belief, and then EUT applied as usual. It is noteworthy that SEU does not assume that the subjective belief distributions that agents hold satisfy Bayes Rule when updated over time, despite Savage being a staunch advocate for each. Bayes Rule is a separate model of (dynamic) risk perception, which may or may not apply with SEU. Relaxations of ROCL that still assume that the agent has a well-defined subjective belief distribution characterize uncertainty, and models of decision-making that do not assume a well-defined subjective belief distribution characterize ambiguity: see Harrison (2011b) for an exposition.

3 Experiments

There are various methods for eliciting and estimating risk preferences, reviewed in detail by Harrison and Rutström (2008). Unfortunately some of the methods in use have well-known weaknesses and biases. One of the most flexible is to ask the agent to make a series of Unordered Binary Choices over risky lotteries, where each lottery typically has between 1 and 4 outcomes. This method provides enough flexibility to estimate risk preferences at the level of the individual, as illustrated in the case of insurance experiments by Harrison and Ng (2016, 2018). For normative analysis, recognizing the heterogeneity of risk preferences across individuals is critical. Moreover, heterogeneity here means much more than the risk premium: it also refers to the *type* of risk preferences. It makes a significant difference for the normative evaluation of insurance products if the agent is an EUT or RDU decision-maker. In general these models will imply different utility functions, and it is the utility function that is used to calculate the certainty equivalent (CE) of insurance products.[4]

There is considerable evidence that laboratory and field samples, at least in developed countries, consist of roughly 50% best characterized by EUT and 50% best characterized by RDU. This classification refers to estimated models at the level of the individual: comparable classifications arise if one uses mixture models over data that is pooled over individuals, as proposed by Harrison and Rutström (2009). There is "never" any evidence for Dual Theory, which proposes the special case of RDU in which utility functions are linear, and the entire risk premium derives from probability weighting.

There is actually very little evidence for CPT in controlled, incentivized experiments. This may come as a shock to some. Harrison and Swarthout (2016) provide an extensive literature review, which finds that most reported evidence for "loss aversion" is actually evidence for probability weighting. They also report evidence of (at least local) asset integration in the laboratory, which is fatal for the empirical adequacy of CPT. Harrison and Ross (2017) review further evidence, and consider the implications for welfare assessment of the conjecture that the many reported "horse race" victories of CPT over EUT were really wins for RDU in disguise, where the successes of CPT stemmed from its allowance for probability weighting rather than "utility" loss aversion relative to an idiosyncratic reference point.

Another critique of EUT that has arisen in experimental settings is the so-called calibration critique popularized by Rabin (2000). This is the concern that "small stakes risk aversion," supposedly common in lab experiments, implies implausibly large "high stakes risk aversion" under EUT. This concern has also arisen in the behavioral evaluation of insurance deductibles. The point was originally made by Hansson (1988), and has been viewed as an indirect rationale for wanting to consider (utility) loss aversion from CPT as playing an important role in decision-making over low stakes. However, the general experimental literature on risk aversion does not support the theoretical premise of the calibration critique: that premise needs to have the *same person* face small stakes lottery choices over a range of wealth levels. An elegant design to implement this test has been proposed by Cox and Sadiraj (2008, p. 33), and builds on the ability to vary "lab wealth" for a given subject. Evidence from university undergraduates in the U.S. shows that the premise is simply false (Harrison et al. 2017a), although evidence from representatives of the adult Danish population shows that the premise is valid for the range of lab wealth considered (Andersen et al. 2018a). In the latter case, there are alternative assumptions about the degree of asset integration between field wealth and lottery prizes that allow one to reconcile small stakes risk aversion with plausible high stakes risk aversion (Cox and Sadiraj 2006), and these assumptions appear to apply to the Danish population.

There is much less evidence for "hyperbolicky" discounting than conventionally assumed. Prior to Coller and Williams (1999), there were very few experiments that provided designs that allowed one to infer monetary discount rates rigorously. This

4 These claims are illustrated in Sect. 5.3.

might seem like a simple point, but prior literature typically generated annualized discount rates in the hundreds or thousands of percent (and chose not to report them as such, for obvious reasons). Another important insight, often neglected completely, has been to correct for the effect of diminishing marginal utility on inferences drawn from "smaller, sooner" amounts of money and "larger, later" amounts of money about utility discount rates. Modest levels of diminishing marginal utility generate first-order changes in inferred discount rates (Andersen et al. 2008). Variations in designs allows one to test Exponential discounting against all major alternatives, and Exponential discounting clearly characterizes the data best in such settings (e.g., Andersen et al. 2014b). Nor is there any evidence for the alleged "magnitude effect," whereby elicited discount rates appeared to be lower for higher stakes (Andersen et al. 2013).

The significance of the interaction of time preferences and risk preferences has become a key issue recently. Casaburi and Willis (2018) provide striking evidence that the temporal nature of (index) insurance contracts may be a factor in low take-up. They consider a field experiment in which premium payments were deferred to the time of harvest, rather than months prior, and find significant increases in take-up. Using a model that assumes intertemporal risk neutrality, by assuming additive intertemporal utility, they suggest that liquidity constraints play the most likely role in explaining the change in behavior. Allowing for intertemporal risk aversion, however, provides a simple, conventional explanation for these findings: see Andersen et al. (2018b) for evidence from the Danish population. Explicit recognition of the temporal nature of the insurance contract is a major insight, with many potential behavioral implications.

There have been important advances in the manner in which subjective beliefs can be elicited. One strand of literature concerns the estimation of subjective probabilities over binary events, using incentivized scoring rules and corrections for the effect of risk aversion on reports (Andersen et al. 2014a). Many losses in insurance are well characterized as binary events. Another strand tackles the more challenging problem of inferring whole subjective belief distributions for continuous or nonbinary events (Harrison et al. 2017b). In the latter case, one can directly make statements about the level of "confidence" that individuals have in their beliefs. Many loss distributions in insurance are real-valued or take on more than two discrete outcomes. The application of these methods has not been widespread in behavioral insurance as yet. One implication is that many studies are forced to assume that agents have subjective probabilities that magically match actuarial claim rates, which is palpably tenuous. We return to this identification gap later.[5]

[5] This is also one reason to be wary of claims to measure risk preferences by responses to surveys about participation in "risky activities" such as smoking, drinking, use of seat belts, and having higher job risk. The link between these activities and risk is clear enough, but identification of risk preferences from these summary measures of "risk tolerance" is complicated by numerous confounds (e.g., knowledge of the ease of quitting when starting to smoke, compensating wage differentials for job risk). Cutler et al. (2008) correlate such survey responses with insurance purchases, and *ex post* insurance utilization, and draw strong conclusions about different alleged patterns of adverse selection across different insurance markets. Surely we can do better than these correlations.

Finally, the issue of hypothetical bias. There is ample evidence, across a wide range of choice tasks, that hypothetical bias exists and affects inferences: see Harrison (2005, 2014c) and Harrison and Rutström (2006) for general reviews, and Laury et al. (2009) and Jaspersen (2016) for reviews specific to insurance applications.[6] This is a debate for another time and place, if at all.

4 Descriptive behavioral models

Underlying all normative behavioral models is some descriptive behavioral model. To finally set the stage, we need to quickly review some of the main themes of recent literature.

4.1 Inferences from observational data

Cohen and Einav (2007) examine a rich dataset of choices over menus of deductibles and premium payments for auto insurance that varied across individuals. They know the premium offered, but do not know the subjective perception of the risk of a claim, or the risk that the claim will be paid in full. To proxy the latter, they assume that individuals have accurate point estimates of the true distribution, a tenuous assumption even for experienced drivers. Moreover, they must assume EUT, since they have no way to identify non-EUT models of risk preferences, and hence the calibration implications of such preferences. Certain non-EUT models of risk preferences, such as RDU, have been shown to dramatically affect the valuation of insurance when calibrated to estimates from real choices in the field (Hansen et al. (2016). This identifying assumption, that individuals know the actuarial loss rates and claim values, turns out to play a critical role in most of the observational literature.[7]

The same confounding issue arises in the evaluation by Sydnor (2010) of choices over deductibles on home insurance. By choosing lower deductibles the individual is paying a lower, certain premium, in return for a risky return given by the claim rate, and the *subjective* perception of how often the individual expects to make a claim in the next year. Since these are lower deductibles, there is no risk attached to the *amount* that is saved by the lower deductible per se, but lower deductibles mean that the individual accepts more risk, so risk preferences must still play a role in this decision even under EUT. Moreover, it is easy to imagine an RDU agent viewing the

[6] Having surveyed this topic for decades now, I have reached the point of commenting in seminars that someone that claims that there is no evidence of hypothetical bias must be ignorant of the literature or knowingly dissembling. Recent claims to have hypothetical survey instruments that are "calibrated" to provide reliable measures of incentive-compatible risk preferences are questionable, even if the concept of calibration for hypothetical bias is a sensible one in general [e.g., Blackburn et al. (1994)] and in need of a formal Bayesian makeover.

[7] Indeed, in a survey, Ericson and Sydnor (2017, p. 54) correctly note that "When economists analyze health insurance markets, they typically assume that people are aware of the distribution of their possible medical bills for the year and choose their health plan with that information in mind."

actual claims rate "optimistically" enough to justify these deductibles.[8] Again, nothing in these data allow one to identify the parameters of the simplest RDU model.

Barseghyan et al.'s (2013) review is an important advance in the analysis of insurance deductible choice. They exploit the fact that the decision-makers in their sample had a choice from multiple deductibles, and recognize that this allows them to identify the role of diminishing marginal utility and probability weighting, since these two channels for a risk premium have different implications at different deductible levels. They also explicitly acknowledge that what they call probability weighting might also be simply subjective risk perceptions that differ from the true claims rate, noting that their analysis "does not enable us to say whether households are engaging in probability weighting per se or whether their subjective beliefs about risk simply do not correspond to the objective probabilities." (p. 2527).

Their striking result is that probability overweighting with respect to claims is, along with diminishing marginal utility, a central determinant of the risk preferences of these deductible choices. They use semi-parametric methods to infer the probability weighting function. Although such methods have some obvious attractions, they can lead to a priori implausible results, such as the massive jump discontinuity from the infamous probability weighting function sketch of Kahneman and Tversky (1979, Fig. 4, p. 283): claims rates of zero imply weighted claims rates of 6.5%, with 95% confidence intervals spanning 6% and 10% (Fig. 1). They also estimate CRRA coefficients of 0.37 or 0.21 over different specifications (p. 2524).

4.2 Inferences from experimental data

There is a long literature in experimental economics examining behavior toward insurance products, reviewed in detail by Harrison and Ng (2019). Here we focus on applications that have clear implications for the normative behavioral evaluation of insurance, revisited in Sect. 5.3.

Harrison and Ng (2016) conducted lab experiments with full indemnity contracts defined over losses from an endowment, with known loss probabilities and no deductibles. Using a battery of binary choices, they estimate risk preferences for each subject, and classify subjects as EUT or RDU.[9] They also estimate parameters

[8] For example, a typical choice from the sample was to pay $100 to get a $500 reduction in the deductible (p. 182). The actual claims rate was 0.042 in this case, at least for the claims that resulted in a payout. An RDU decision-maker with a power probability weighting function $\pi(p) = p^\gamma$ would only need $\gamma = 0.5$ to have a weighted probability and decision weight of 0.205, exceeding the 0.2 needed to justify the purchase. And it is reasonable to expect that some households might perceive the true probability as higher than 0.042, requiring even less optimism (at least with respect to the return on the insurance contract, as distinct from the underlying loss) to justify the purchase. The estimated probability weighting function of Barseghyan et al. (2013; Figs. 2 or 4), for comparable choices by samples from comparable populations, implies a weighted probability of roughly 0.11 if one uses the actual claims rate of 0.042. Of course, this is still a violation of EUT, which is the general point being made by Sydnor (2010).

[9] Since EUT is nested in RDU, if one was to use log-likelihood levels alone for the classification every subject would be RDU. Instead, they take the view that EUT is the natural null hypothesis, and someone is classified as RDU only if there is statistically significant evidence (at the 5% level) of probability weighting.

for structural models of risk preferences, which play a key role in their normative analysis, described later. After the risk aversion choices, subjects make a series of binary choices to purchase insurance or not. Since insurance contracts vary with respect to premia, loss probabilities and loss amounts, they could directly evaluate the extent to which the demand for insurance varies with these "actuarial" characteristics of the contract.

The same design is extended by Harrison and Ng (2018) to consider nonperformance risk. This risk is modeled, theoretically and in the experiments, as an extra probability that a loss will actually be approved by the implied insurance company. Nonperformance risk was modeled as one probability, reflecting both solvency probability and repayment percentage if insolvent. In this manner, one can look at the realistic cases of complete insolvency or "pennies on the dollar" insolvency. The key conceptual issue raised by nonperformance risk is the extent to which individuals process compound risks the same way that they process simple risks. Hence their risk battery included lotteries from Harrison et al. (2015) to identify consistency with the ROCL axiom. Those tests of ROCL are simple to implement, and yield a data-based measure of ROCL consistency.[10] Focusing solely on their descriptive finding, they show that the usual actuarial characteristics, particularly premium levels and loss probabilities, play an important role determining *take-up*. They also find that the fraction of repayment had the expected effect on *take-up*, although the fact of solvency did not. The count of ROCL violations, which varied from subject to subject, had no significant effect on *take-up*. Of course, take-up is not welfare: we return to those normative inferences below.

5 Normative behavioral models

5.1 Early insights

Feldstein (1973) proposed that, on average, U.S. households carried too much health insurance. Armed with estimates of a representative measure of risk aversion, a price elasticity of demand for health care, the (gross) price change induced by lower insurance coverage, and the decrease in health care quality induced by lower insurance coverage, he estimated that the CE of the EU loss from reduced insurance coverage would be more than offset by the gain from reduced purchases of lower-priced health care.[11]

[10] Subjects are given a series of binary choices between some simple lottery and a compound lottery, and then later or earlier given choices between the same simple lottery and the actuarially equivalent simple lottery to the original compound lottery. One then just counts the number of choices in these pairs that are the same.

[11] Feldman and Dowd (1991) updated these calculations with later, improved estimates of the moving behavioral parts. They also corrected the estimates of deadweight loss from Manning et al. (1987), which did not include the CE of the EU loss from reduced insurance. Friedman (1974) estimated risk preferences from health plan choices of Federal employees in the U.S., simulating claim probabilities and amounts and assuming that individuals responded to those parameters under EUT.

Townsend (1994) initiated a major stream of research by examining the response of household consumption to income shocks. Examining data from villages in rural India, he found that "household consumptions are not much influenced by contemporaneous own income, sickness, unemployment, or other idiosyncratic shocks, controlling for village consumption (i.e. for village level risk)" (p. 539).[12] Under certain, strong assumptions, evidence that consumption remains "stable" over time in relation to relative volatility of income indicates that there is likely to be small welfare gains, if any, from "social insurance" schemes.[13] Indeed, Chetty and Looney (2006, p. 2352) note, with citations, that "the presumption that consumption fluctuations give a measure of the welfare costs of risks, and therefore the value of additional insurance, remains prevalent." The subsequent literature on full or partial insurance, inferred from such correlations, continues. For example, Blundell et al. (2008) conclude from the U.S. data that there is "some partial insurance of permanent shocks, especially for the college educated and those nearing retirement [and that there is] full insurance of transitory shocks except among poor households." (p. 1887).

However, Baily (1978) had much earlier identified an important tradeoff between the factors causing benefits from consumption smoothing (higher risk aversion) and the factors causing costs of smoothing consumption in the design of optimal unemployment insurance. Focus on the latter: in a world of complete and perfect markets, these costs are low. Absent these imaginary markets, it is often presumed that private or informal insurance mechanisms at the individual, household, village, state or national level somehow act as if providing "full insurance" against consumption variability. Or in the debate over the roles of social *versus* private insurance, that private insurance serves to do what social insurance proposes doing. However, the logic proposed by Baily (1978) implied that evidence of consumption smoothing in Townsend (1994) might be evidence of *extremely* high risk aversion and *inefficient* risk management options: as long as the demand for risk reduction is high enough, even wasteful risk management schemes will be tolerated. A review of the vignettes from the *Portfolios of the Poor* financial diaries, by Collins et al. (2009), tells of the myriad, costly risk management schemes needed to understand "how the world's poor lives on $2 a day." Chetty (2006) and Chetty and Looney (2006) show precisely

[12] This conclusion was qualified in some villages for those that did not own land.

[13] Because of the influence of this approach, it is worth noting the explicit methodological position that motivated it. Townsend (1994) was well aware of the long list of mechanisms and institutions that might provide informal insurance, noting family transfers among villages, informal credit markets, plot and crop diversification, and animal sales. And rigorously documenting this type of long list has occupied him in later work in rural Thailand: see Samphantharak and Townsend (2010). However, Townsend (1994, p. 540) argues that "in studying one market or institution only, the researcher may miss smoothing possibilities provided by another. For example, transfers may be small or missing, but this may not leave the family vulnerable if credit markets function well. [Hence this study] presents a general equilibrium framework which overcomes the problem of looking at risk-sharing markets or institutions one at a time. Specifically, the general equilibrium model inevitably leads the researcher to focus on outcomes, namely, consumption and labor supply, so that all actual institutions of any kind are jointly evaluated." One concern with this position is that a general equilibrium structure is used to generate "reduced form" results which are then empirically evaluated, without being able to go back and verify the structure.

how this logic applies to understand the identification issues that plague the conclusions from Townsend (1994).

5.2 Inferences from observational studies, natural experiments, and surveys

Einav et al. (2010b) develop a structural empirical model of the demand for annuities in the United Kingdom between 1988 and 1994 for which the annuitant was still alive at the start of 1998. Data on gender, age at annuitization, and age at death if prior to 2006, is observed, as well as the level of annuitization and the choice of a 0, 5 or 10-year guarantee. Annuitization itself is compulsory for most of the accumulated balances from tax-preferred, defined-contribution pension payments. Annuity payment rates decline with longer guarantee periods (Table II, p. 1039) and this pattern was held constant over the period of annuitizations.

A key issue for the effects of adverse selection on welfare evaluation is whether there is any private idiosyncratic information that individuals have when they decide on the length of guarantee. Subjective beliefs about longevity, conditional on reaching the age at which this decision is made, are what is relevant for *ex ante* welfare evaluation. However, *ex post* mortality rates can provide some partial indicator of the potential extent of the problem. Over all 9364 annuitants, 10%, 87%, and 3% chose the 0-year, 5-year, and 10-year guarantee, respectively (Table I, p. 1037). Conditional on choosing the 0, 5, or 10 year guarantee, mortality rates were 16%, 21%, or 19%, respectively. Across the three contracts to choose from, the mortality rates were 20%, so 1-in-5 received the *ex post* benefit of the guarantee. Of course, this can only be one piece of the puzzle: subjective beliefs about these mortality rates, even if they are assumed to match the realized rates, do not tell us subjective beliefs about longevity *beyond* the guarantee period.

Heroic assumptions are needed to generate welfare estimates of alternative policies for individuals. This is not a criticism, just a recognition that if one is to go beyond the qualitative identification of the existence of adverse selection or moral hazard in insurance purchases and quantify their importance, one must let theory, parametric structure and assumptions play a central role. In this case, the empirical model assumed EUT decision-makers, Exponential discount rates, and additive intertemporal utility functions (p. 1041): these assumptions rule out the behavioral alternatives reviewed earlier in Sect. 2, which could be important for some individuals. In addition, and a common assumption when working with observational data, individuals are assumed to know the relevant risks that are being insured, in this case their own longevity risk (p. 1042). Furthermore, the same CRRA utility function over consumption applies to all individuals, and with one caveat the same CRRA utility function used for consumption applies to all individuals with respect to the utility of bequests at death (p. 1043).[14] Values for RRA and the discount rate are assumed, not estimated. In fact, RRA is set to 3, and the discount rate is set to 4.3% p.a. on the

[14] The caveat is that individuals have a multiplicative weight that they put on the latter argument of utility, interpreted as “the relative weight that individual *i* puts on wealth when dead relative to consumption when alive” (p. 1043).

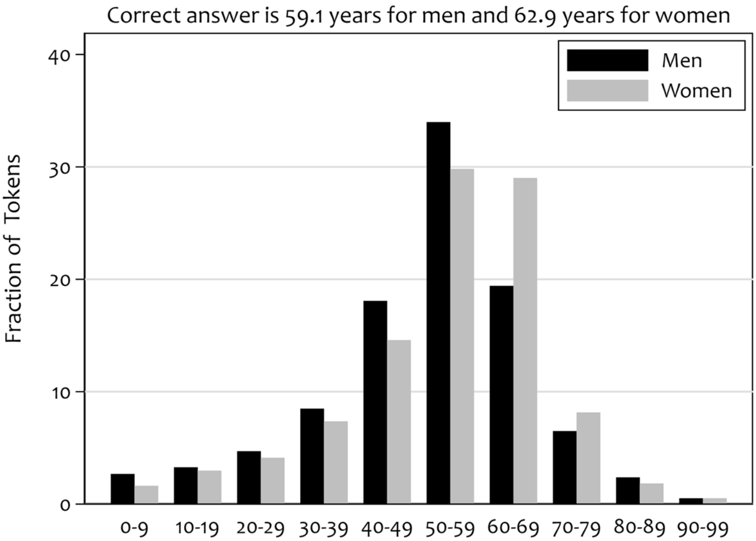

Fig. 1 Elicited beliefs for remaining lifetime. Correct answer is 59.1 years for men and 62.9 years for women

basis of a real interest rate at the beginning of 1992. In addition, since annuitization rates are in nominal currency units, they have to assume an expected annual inflation rate of 5% to infer the real annuity payout stream that individuals are choosing over.[15] Although there are several references to "estimates of the joint distribution of risk and preferences" (e.g., p. 1082), there is no sense at all in which risk preferences toward consumption variability or bequest risk are estimated, let alone time preferences, let alone any interaction between risk and time preferences.

Some of these heroic assumptions are treated as being relatively unimportant. Perhaps the most important for this insurance contract is longevity risk, as noted earlier. Einav et al. (2010b, p. 1079) claim that, "Throughout we made a strong assumption that individuals have perfect information about their actual mortality rate […]. This is consistent with empirical evidence that individuals' perceptions about their mortality probabilities covary in sensible ways with known risk factors, such as age, gender, smoking, and health status […] Of course, such work does not preclude the possibility that individuals also make some form of error in forecasting their mortality." In fact, there is evidence that individuals, as well as epidemiologists and actuaries, struggle with forecasts of longevity risk (Elder 2013). Figure 1 displays results of an incentivized subjective belief elicitation of longevity risk from 120 graduate students in England, reported in detail in Di Girolamo et al. (2015). The question asked was, "Based on 2010 National Statistics, if a man lived to be 20 in the United

[15] Hansen et al. (2016) illustrate how one can combine estimates of key behavioral parameters such as these, from field experiments with representative populations, and insurance data from the same population.

Kingdom, how many more years would he expect to live? Note that this is not the age he would die at, but how many more years he would expect to live"; the obvious variant for longevity risk of women was also asked. Although respondents correctly perceived the gender differential in favor of women, there is considerable imprecision in these pooled beliefs, and significant heterogeneity in beliefs at the individual level.

The grand utility function over consumption flows while alive, and bequest motives when dead, is also assumed to be additive in these two components, quite apart from the additivity of the former over remaining time periods of life. Inflation risk is also assumed away (footnote 11, p. 1050). The point of mentioning all of these factors is that tradeoffs between these many risks are ruled out by additive structures. The implications for behavior toward retirement planning, and related life-cycle decisions, of allowing tradeoffs between risk preferences and longevity risk in intertemporal settings has been extensively explored by Bommier (2006, 2010, 2013) and Bommier and Rochet (2006).

Welfare results are provided to show the effect of the existing annuity scheme in comparison to a scheme in which annuitization rates could be adjusted to the mortality risks of annuitants so as to offset the immediate effects of adverse selection on those risks, and in comparison to mandating 0, 5, or 10 year guarantees. One important methodological innovation is to consider welfare gains in relation to the "maximum money at stake" in their decisions, defined as the smallest amount of money needed to compensate the annuitant for receiving their least-preferred contract.[16] They find that the welfare gain is greatest for the risk-adjusted contracts that offset adverse selection and for a mandated 10-year guarantee. However, they stress the strong assumptions needed to draw these conclusions, and emphasize that their "results highlight the practical difficulties involved in trying to design mandates to achieve social welfare gains." (p. 1069). Even with calibrated risk preferences and time preferences, rather than estimates, their results do show that "the welfare-maximizing choice for a mandated contract would not be apparent to the government without knowledge of the joint distribution of risk and preferences." (p. 1082). Since there is no heterogeneity allowed in terms of risk preferences, time preferences, or subjective perception of longevity risk, the informational challenge to the design of welfare-improving policies is even steeper.[17]

Einav et al. (2010a) apply a "sufficient statistics" approach to measure changes in social welfare from insurance.[18] The approach rests on assuming that (naive) revealed preference applies (p. 879), so that the demand curves for insurance products are "sufficient statistics" for willingness to pay for the product. It is also solely directed at social welfare, with no basis for making any inferences about the

[16] In this setting, "money" is correctly defined in terms of a Wealth Equivalent (p. 1060), which is the appropriate analogue of the CE of a standard lottery.

[17] In addition, annuity markets are one of the few insurance markets where moral hazard issues can be assumed to be less important. Hence, the methods employed here would need to attend to those issues in general.

[18] Chetty (2009) provides an excellent exposition of this general approach.

distribution of welfare gains (let alone gains and losses) among the population. Finally, it is limited to evaluating the welfare effects of changes in the *pricing* of *existing* contracts (p. 878).

Bundorf et al. (2012) undertake welfare calculations of the effects of self-selection of health insurance plans by employees, particularly when uniform contribution policies enforce the same price per plan for employees within a firm. Their policy interest is the use of "risk scoring" algorithms to allow for risk-adjusted contributions in a counterfactual in which individuals are risk-classified. These are important questions to ask, and to try to answer. A severe limitation of the analysis, however, is that risk preferences do not play any important role. Indeed, the expression "risk aversion" does not get a mention until §V, in an afterthought, and there one finds a single, *assumed* CARA value for all individuals. To be sure, this afterthought calculation is intended to show that making some allowance for risk aversion with respect to "reclassification risk" does not offset the efficiency gain of allowing reclassification. However, reclassification risk, in the sense used here and even assuming EUT, is not the core risk preference that needs to be accounted for with respect to the individual welfare gain from insurance.

That "core" risk preference comes in the utility functional used to evaluate the product by consumers. In the formal specification, Bundorf et al. (2012, p. 3225) adopt a money-metric utility function.[19] Household utility depends additively on plan characteristics, plan contributions, demographics, household health risk, and "an idiosyncratic preference" term. The household health risk comes from a commercial simulation model and algorithm that uses rich data for the individual, akin to how an actuary might calculate the risks. Whatever their actuarial merits for pricing and reserving calculations, these are not, of course, the subjective health risk perceptions of households. The "idiosyncratic preference" term, where one might hope to have heterogenous risk preferences considered, is a nuisance parameter, literally, whose potential endogeneity problems are handled by instruments (p. 3229).

The deeper problems with this approach for welfare analysis stem from the use of money-metric utility, building on the general approach of Small and Rosen (1981). In effect, one is approximating the area to the left of compensated demand curves, with approximation risks that are now well known.[20] Hence, from a behavioral welfare perspective, one is assuming naive revealed preference to generate the estimates

[19] This defines the minimum income needed at fixed, reference plan attributes and individual characteristics to pay for an insurance bundle that is at least as good as the one that is consumed. For those reference attributes and characteristics, this is a particular normalization of the household's ordinal utility function.

[20] Important results on the implications of unobserved heterogeneity of demand on inferences about *average* welfare effects of price changes are provided by Hausman and Newey (2016). They stress the difficulty of precise identification, and develop a bounds approach. And this is still just for inferences about the average welfare effect, not the full distribution. A deeper issue is the normative status of the demand curve for an insurance product. This issue was at the heart of the debate between Pauly (1968) and Arrow (1968), with them agreeing that interactions between the insurance contract and the demand for the asset or services being insured invalidates direct inferences about welfare. Newhouse (2015) provides a cautionary reminder of the complications that this insight has for normative inferences from data drawn from social experiments on health insurance (let alone nonexperimental, observational data).

of CS; this is fine, as long as we know what is happening beneath the details. To summarize a long and important technical literature, money-metric utility functions are wonderful devices to descriptively model demand curves, but severely problematic when it comes to doing welfare analysis. One immediate point is that all evaluations of welfare must be undertaken for very small perturbations in plan attributes and household characteristics. Referring to traditional applications with a reference *price* vector, rather than a vector of attributes and characteristics, Blackorby and Donaldson (1988, p. 128) warn that

> small changes in the reference price vector may yield large changes in optimal solutions. Since reference price vectors are typically picked by the analyst in an ethically arbitary and mechanical way, ethically acceptable social judgments are not guaranteed.

More generally, after a long exegesis of the manner in which money-metric utility helps free *descriptive* demand analysis from all cardinality of utility, Samuelson (1974, p. 1266) sharply concludes:

> Whatever the merits of the money-metric utility concept developed here, a warning must be given against its misuse. Since money can be added across people, those obsessed by Pareto optimality in welfare economics as against interpersonal equity may feel tempted to add money-metric utilities across people and think that there is ethical warrant for maximizing the resulting sum. That would be an illogical perversion, and any such temptation should be resisted. [footnotes omitted]

Similar warnings appear in Slesnick (1998, p. 2141) and Blundell et al. (1994).

Handel and Kolstad (2015) seek to tell a story about the role played by "risk preferences" and the role played by "information frictions" in determining the demand for health insurance products. They also seek to tell a story about the welfare implications of the inclusion of "information frictions." I use the expression "seek to tell a story" to be clear that this is academic rhetoric, for the purpose of shifting discussion away from just assuming that "risk preferences" alone explain insurance behavior.[21] Others might not see this type of rhetoric as the right way to model behavior, but that position neglects any appreciation of the paucity of data with which to draw inferences in the field.

Handel and Kolstad (2015) start with a rich administrative dataset in which individuals with certain demographic characteristics had to choose between two health insurance plans. One plan, the Preferred Provider Option (PPO), provides "comprehensive risk protection" (p. 2451); the other plan, a High Deductible Health Plan (HDHP), provided access to "the same medical providers and treatments as the

[21] They reference (p. 2450) Cohen and Einav (2007) and Bundorf et al. (2012) as conducting welfare analysis of health insurance plans in which they use "observed choices to identify risk preferences." In fact, risk preferences are not identified by Bundorf et al. (2012), as explained earlier. And Cohen and Einav (2007) undertake no welfare analysis. Similarly, Einav et al. (2010a, p. 878) claim that Einav et al. (2010b) and Bundorf et al. (2012) "recover the underlying (privately known) information about risk and preferences." Neither of these claims are true.

PPO, lower relative upfront premiums, and larger relative risk exposure." (p. 2451). In addition to the administrative data, for a significant subsample of the population, they also had a linked survey of beliefs about these plans. The intuition of their results can be seen by one example (p. 2451): if individuals incorrectly believed that the PPO provided greater medical access to providers and treatments (20% of the sample), or were not sure about that (30%), they were more likely to choose the PPO than individuals that knew that the plans provided the same access. Call these subjective beliefs about some core attributes of the products. Given these subjective beliefs, apply SEU to these choices, and what we see is just a better apple or a less risky apple being selected over a poor apple. The first 20% subjectively perceive a more useful product, and the second 30% subjectively perceive a less risky product.

The first formal step in the analysis is just to recover risk preferences from observed choices between the PPO and HDHP. In this case, the model assumes EUT, and critically *assumes that individuals know the actuarial probabilities* of receiving benefits from each insurance plan. Intuitively, think of the PPO as the safe lottery and the HDHP as the risky lottery.[22] To borrow an expression, the resulting estimates of risk aversion are "just wild and crazy guys," to be laughed at because they are so high (p. 2452). Of course, we know from RDU models of risk preferences that this *might* actually be a combination of (very) pessimistic beliefs about receiving the benefits of the HDHP and a (modestly) concave utility function.[23] And we know that 50% of the subjects confessed, albeit in a hypothetical survey, to entertaining just such pessimistic beliefs. This is also just another story. The point is that the available data is unable to differentiate them, hence we cannot claim to have identified risk preferences without accepting the maintained assumption of EUT for all individuals, and where EUT assumes prescient knowledge of the actuarial risks of what are clearly compound subjective lotteries.

The second formal step in the analysis is to correctly recognize (p. 2455ff.) that modern health insurance plans have many attributes that differentiate them. We are not in a world, at least for these product lines, of just trading off lower deductibles for higher premia. In the absence of these "nonfinancial attributes," the utility function has, as an argument, $W_k - P_{kj} - s_i$ where W_k is wealth for household k, P_{kj} is the premium that household k faces for insurance plan j, and s_i is the out-of-pocket payments for some sad event i. Then there is some actuarial probability mass function, let us assume, defined over the s_i, and that depends on the household k and plan j in question. Now consider the effect of "nonfinancial attributes," such as "the network of physicians and hospitals available, the time and hassle costs associated with

[22] The effort to construct these actuarial probabilities (p. 2480) is impressive. It uses *ex post* information to predict the utilization of four types of health expenditure in the coming year, and then *ex post* data on the costs of each of these expenditure types to predict spending distributions. One could use these objective calculations as the basis for eliciting subjective probability distributions with incentive-compatible experiments, which is what we need to estimate an SEU model of insurance choice.

[23] This reference to RDU is to make the simple point about alternatives to EUT, and not to say that RDU decision weights are subjective probabilities, although some do adopt that interpretation. It is possible to apply RDU to subjective probabilities, in appropriately identified settings: see Andersen et al. (2014a). Hence one could actually have pessimistic subjective probabilities as well as some probability weighting.

dealing with claims, and the tax benefits of linked financial accounts." (p. 2455). For short, call this BLOB_j for plan *j*, recognizing that BLOB has potentially many arguments reflecting a vector of perceived attributes.[24] The argument of the utility function then becomes $W_k - P_{kj} - s_i + \mathrm{BLOB}_j$. This specification is at the heart of the analysis.

There are two problems with this way of handling "nonattribute frictions." The first problem was alluded to earlier: many of these attributes might just be naturally subsumed into the *subjective* probability mass function, rather than being added as arguments to the utility function. Indeed, one might plausibly make a case that they could be added to the *objective* probability mass function reflecting compound risks, and in turn we would plausibly allow different households to hold different subjective perceptions of these risks. In any event, it is immediately apparent that this alternative would lead to very different estimates of "risk preferences." The second problem is that they are included in an additive manner. This implies that they are known quantities if one knows the household *k* and plan *j*, so they are not themselves risky.[25] This also implies that even if they were assumed to be risky, they *cannot* tradeoff with other "financial risks." The literature on multiattribute risk aversion shows that *additive* utility functions defined over risky attributes exhibits multiattribute risk neutrality: see Andersen et al. (2018b) and Gangadharan et al. (2019) for discussion and applications. The general point is that we are talking about "risk preferences" here, albeit in the form of an exciting cocktail of multiattribute risk preferences, but just risk preferences nonetheless.[26]

The modeling upshot is that I am suggesting a different "story" here, and there is no possible way for these data, as rich as they are in comparison to most observational datasets, to tell them apart. However, this story has very different implications for how one does welfare evaluations. The exercise undertaken by Handel and Kolstad (2015) is to assume their structural model is valid and to counterfactually eliminate the PPO plan, the "safe" lottery choice here.

Handel (2013) exploits a natural experiment where a large firm changed health insurance options from an active choice mode to a passive mode in which the previously selected choice was the default choice in later years unless action was taken. This change allowed inferences about the role of "inertia" in insurance plan choice. The behavior of new employees, who needed to make an active choice when previous employees were faced with passive choices, provides intuition for the significance of inertia, assuming comparability of other characteristics between the two employee groups. In addition, some passive employees faced dominated choices over time as

[24] Indeed, BLOB could be viewed as a nested utility function defined over these attributes, as proposed in footnote 12 (p. 2456) and in the empirical model. In the empirical model (p. 2475), these attributes are all treated as binary, and included additively.

[25] The only stochastic aspects of these attributes (p. 2456) is that they are *observed* with error by the researcher, reflecting unobserved but deterministic heterogeneity.

[26] Handel and Kolstad (2015, p. 2452) include "inertia" in their structural model, and comment that "incorporating inertia into the model matters a lot for risk preference estimates." They refer here to *atemporal* risk preferences. The deeper implications for risk preferences, having to do with *intertemporal* risk preferences, is discussed below with reference to Handel (2013), where "inertia" is the main story.

insurance parameters changed, and their sluggishness in the face of these incentives provides indicators of inertia. Risk preferences are assumed to be distributed randomly over the population sampled, and be consistent with EUT. Individuals know their own risk preferences, but this is unobserved by the analyst. This might cause identification problems if the "nonfinancial attributes," to use the expression of Handel and Kolstad (2015), also varied across all plan choices, but three PPO plans had no differences in these attributes: hence their variations in "financial attributes," such as deductible, coinsurance, and out-of-pocket maxima, could be used to identify risk preferences.[27] In keeping with other observational studies, the distribution of claims was simulated using sophisticated models akin to how an actuary would undertake the task, and individuals were assumed to know the risks they faced exactly.

Since the focus is on "inertia" over time, an important behavioral omission is the implicit assuming that individuals are *intertemporally risk neutral*. Hence, whatever the implied atemporal risk aversion from the random coefficient estimation, individuals are unable to exhibit inertia in choices due to intertemporal risk aversion. This is quite separate from the assumption that "consumers are myopic and do not make dynamic decisions whereby current choices would take into account inertia in future periods" (p. 2662). That assumption has to do with sophistication with respect to the effect of current consumption on future consumption, akin to "rational addiction" models. Intertemporal risk aversion is just a taste for not having variability in claims risks over time, and that is met simply by choosing the same plan year over year. Just as one is willing to pay a risk premium in terms of expected value to reduce atemporal risk aversion, the willingness to put up with lower expected value plans can be seen as a risk premium to reduce intertemporal risk aversion. This has fundamental implications for the resulting welfare analysis (pp. 2669–2679). The story here is that "consumers enroll in suboptimal health plans over time, from their perspective, because of inertia. After initially making informed decisions, consumers don't perfectly adjust their choices over time in response to changes to the market environment (e.g., prices) and their own health statuses" (p. 2669). Another story, equally consistent with the observed choices and EUT, is that consumers have a preference for avoiding intertemporal risk in the health plan lotteries they choose.

Loewenstein et al. (2013) report the results of hypothetical surveys to evaluate if individuals understand the health insurance products they are being asked to purchase.[28] One survey asked about some basic insurance concepts (deductible, copay, coinsurance, and out-of-pocket maximum), and then presented a standard, commercial health insurance contract with all of these concepts in play and asked some questions about what the contract entailed. Accepting the methods to measure insurance literacy for the moment,[29] the conclusion is that there is "strong evidence that

[27] Again, the presumption is that individuals do not subjectively believe that these attributes differ across these PPO plans.

[28] Ericson and Sydnor (2017, p. 58ff.) review the broader literature on "confusion" in health insurance choice.

[29] Some of the questions are not ideal measures of literacy in this domain, reflecting poor survey design for the inferences intended. For example, the questions about the concept of deductibles has multiple-choice answers (p. 853) to the question "Which of the following best describes a Deductible?", and responses are coded as "true or false." Two of the responses are clearly false, one is "I'm not sure," and

consumers do not understand traditional plans" (p. 850). So how do we evaluate this? We are told that "limited understanding is likely to lead to suboptimal decisions," (p. 852), but how do we know? If someone responds to a survey question "I'm not sure," that is a *plausible* signal for someone that is likely to seek a cognitive scaffold prior to making an actual decision (e.g., check the internet, check with an expert, or just check with a friend).[30] Access to a scaffold does not ensure an optimal decision, but the response is at least flagging *some* lack of confidence in the answer, and that surely has some implications for behavior beyond just assuming a priori that someone will pick at random or in systematic error.

Loewenstein et al. (2013) do flag two further ways in which understanding, or literacy as it should be termed, might affect individual welfare. One is whether individuals choose health insurance policies that minimize their expected costs. This is a problematic metric, hinted at with the comment that "while cost minimization is not necessarily equivalent to utility maximization, it is a useful benchmark." (p. 852). A more accurate statement would be that "cost minimization is not equivalent to expected utility maximization, or even maximization of some other interesting utility function, and is not a useful benchmark." We simply have to minimally attend to risk preferences, time preferences, and subjective beliefs before we start making claims about individual welfare. The second way in which literacy failings might impact individual welfare is to see if a "lack of understanding was correlated with their insurance choices," as in Handel and Kolstad (2015). In the absence of more nuanced evaluations of these choices, in terms of preferences and beliefs, such correlations mean little.

Bhargava et al. (2017) study a remarkable dataset from a company that offered employees a menu of 48 health insurance plans that differed solely in terms of "financial attributes." In particular, there are blocks of four plans that literally differed solely in terms of the deductible and the premium. In one case, Plan A (p. 1329), an increase of $\$1204=\$2134-\$930$ in the premium was accompanied by a reduction of $\$650=\$1000-\$350$ in the deductible, and this difference was representative across other plans. Roughly 55% of employees selected a dominated plan, after allowance for after-tax adjustments. Average medical expenditures were $3567 (p. 1336) and those that chose dominated plans "could have saved an average of $352 with little risk of losing money" (p. 1339). In nominal cost terms, this is just under a 10% savings compared to expenditures.

Footnote 29 (continued)

the two others are "The amount you pay before your insurance company pays benefits" and "The amount you pay before your health expenses are covered in full." The last one is false in the sense that it ignores possible co-pays that might apply over the period covered, and ignores possible coinsurance payments. However, it certainly covers the essential idea reasonably well. Another question poses a specific scenario about the commercial plan, Plan T on page 861, for which the correct answer requires the arithmetic evaluation of $\$1500+0.8\times(\$100{,}000-\$1500)=\$1500+0.8\times\$98{,}500$. The respondent is asked for an open-ended response: do *you* know the exact answer without calculating it?

[30] Clark (1998) refers to these external elements to one's physical brain as "cognitive scaffolding." Ross (2005, 2014) develops the role of scaffolding in specifying and identifying utility functions using sophisticated revealed preference theory.

Of course, expected savings is not the same as risk-adjusted savings. While it is true that "no beliefs about health care needs or standard preferences for avoiding risk would rationalize the choice of the low-deductible plan" (p. 1321), various assumptions could make these welfare losses *de minimis*. An EUT calculation, assuming that individuals again use actual distributions of medical expenditure as their subjective distribution of medical expenditure (p. 1342), leads to comparable estimates of the CE of the foregone savings. These CE range from \$372 down to \$167 (p. 1344) depending on the level of risk aversion assumed, as one might expect a priori. Of course, an EUT calculation does not take probability weighting into account, even if one continued to assume that subjective expenditure probabilities equaled historical probabilities, and this could have a first-order effect on the implied CE. And there is no accounting for aversion to variability of payments over time: a deductible of \$1000 over several years allows more room for variability of out-of-pocket expenditures than a \$350 deductible.

A potentially valuable complement to the evaluation of observational data was the use of experiments to evaluate alternative explanations in stylized, but "naturalistic" settings. Unfortunately, these were all hypothetical surveys conducted online. These can be useful to set up tests of hypotheses,[31] but suffer from the general problem of hypothetical bias referred to earlier.

5.3 Inferences from experimental studies

Harrison and Ng (2016) provided an explicit welfare analysis of the simplest full indemnity insurance contract in controlled laboratory experiments. They used the risk preferences for each individual, estimated from a risk aversion task, to infer if the individual was an EUT or RDU decision-maker, and to provide parameter estimates for their specific risk preferences. Armed with estimates of the utility function of each subject, they were able to directly calculate the expected CS of purchasing insurance or not purchasing insurance, in each case using the CE difference between the two actions.

Table 1 illustrates their calculations for a single subject. The first column shows the decision number, ordered here by premium for simplicity. The second column shows the premium on offer. The third column reports the decision of this subject, to purchase the contract or not. If this subject is classified as an EUT decision-maker, we know her risk preference parameters from the prior risk aversion task and can use her estimated utility function to calculate the CE of purchasing the insurance contract and the CE of not purchasing the insurance contract.[32] The difference between

[31] In particular, one intriguing hypothesis (p. 1353) posits that agents might "value costs associated with plan premiums differently than those paid (perhaps unexpectedly) out-of-pocket." In effect, this relaxes the perfect asset integration assumption that some associate with EUT. Cox and Sadiraj (2006) and Andersen et al. (2018a) show how to evaluate partial asset integration specifications using (incentivized) experiments.

[32] For immediate pedagogic purposes, the point estimates of risk preferences are used. The full analysis accounted for covariances in estimates using bootstrapping.

Table 1 *Ex ante* consumer surplus for one subject

Choice	Premium	Choice	Consumer surplus if EUT	Consumer surplus if RDU
1	\$0.20	Buy	\$1.57	\$2.12
2	\$0.40	Buy	\$1.37	\$1.93
3	\$0.60	Buy	\$1.17	\$1.73
4	\$0.80	Not Buy	\$0.97	\$1.53
5	\$1.00	Buy	\$0.77	\$1.33
6	\$1.20	Buy	\$0.57	\$1.13
7	\$1.40	Buy	\$0.38	\$0.94
8	\$1.60	Buy	\$0.17	\$0.73
9	\$1.80	Buy	– \$0.02	\$0.54
10	\$2.00	Buy	– \$0.23	\$0.33
11	\$2.20	Buy	– \$0.43	\$0.13
12	\$2.40	Buy	– \$0.63	– \$0.07
13	\$2.60	Not buy	– \$0.82	– \$0.26
14	\$2.80	Buy	– \$1.02	– \$0.46
15	\$3.00	Not buy	– \$1.22	– \$0.66
16	\$3.20	Not buy	– \$1.43	– \$0.87
17	\$3.40	Not buy	– \$1.63	– \$1.07
18	\$3.60	Buy	– \$1.82	– \$1.26
19	\$3.80	Not buy	– \$2.02	– \$1.46
20	\$4.00	Buy	– \$2.22	– \$1.66
21	\$4.20	Not buy	– \$2.42	– \$1.86
22	\$4.40	Not buy	– \$2.63	– \$2.07
23	\$4.60	Not buy	– \$2.82	– \$2.26
24	\$4.80	Buy	– \$3.03	– \$2.47

these CE is her *ex ante* consumer surplus from purchasing. Hence we observe that if this subject was an EUT decision-maker she should have purchased up to and including the premium of \$1.60, and then not purchased for higher premia. If this subject is classified as an RDU decision-maker, using the specification employed here (a Prelec probability weighting function), we infer different CE for each possible choice, and different CS from purchasing. If she was an RDU decision-maker she should have purchased up to and including the premium of \$2.20.

This simple table of elementary calculations has important implications. First, even if we assume an individual is an EUT decision-maker, we need to know *how risk averse* she is to say if her decision to "take-up" the product is the right one or not. The same point applies generally to the case in which she is an RDU decision-maker. A pox on unconditional nudges to take up the insurance product that ignore risk preferences! Second, we need to know which *type* of risk preferences best characterizes her. Consider decisions 9, 10 and 11: in this case we get the *sign* of the welfare effect wrong unless we know the type of decision-maker. Third, we see CS numbers in dollars, reflecting the equivalent variation in income. We can distinguish what are "small" welfare effects from what are "large" welfare effects. Finally, we see the CS from purchasing or not from purchasing. This might seem trivial, until one realizes that many observational datasets, not all, suffer from the selection bias of only seeing those that purchased insurance.

Figure 2 displays the data from Table 1 in a way that allows one to see the manner in which estimated risk preferences and observed insurance choices translate into realized welfare gains and welfare losses for this subject. The subject in Table 1 was actually best characterized as an EUT decision-maker. The horizontal axis of Fig. 2 shows each premium, arrayed horizontally to match the vertical array in Table 1. The vertical axis shows the *realized* CS, defined as the CS gained by making the right decision or the CS foregone by making the wrong decision: in Fig. 2 we only use the "CS if EUT" column from Table 1. The large "V" spanning Fig. 2 shows the correct take-up decision for each premium: to the left of the V the correct decision is to purchase, and to the right of the V the correct decision is not to purchase. We see immediately how this approach allows us to see welfare gains and losses by the same subject. We also see that some of these losses arise from taking up the insurance contract, to warn us from automatically associating take-up with a welfare gain (i.e., to warn us from a naive application of revealed preference). Finally, we observe that welfare gains and welfare losses can arise from *not purchasing* the product, reminding us of the dangers of only studying data from those that purchase the product.

Undertaking similar calculations for each subject and choice, respecting their individual risk preferences, one can see the correct policy targets of our welfare evaluation in Fig. 3. The left panel displays all *realized* CS values, and the right panel displays a *realized* Efficiency measure for each subject. Efficiency here is defined in the standard manner from experimental economics, as the percentage of realized CS by an individual over all of her decisions compared to the maximum potential CS she could have realized. In this instance, the maximum potential CS for the subject in Fig. 3 is just the absolute value of the CS values in the "CS if EUT" column of Table 1, since it assumes that she would have made all of the right decisions. Figure 2 shows the right behavioral targets for interventions designed to improve welfare. Armed with either of these targets,[33] one can fire up the theoretical engines, or even the nudge engines if one must and has no theory as a guide, to see if some intervention can shift this distribution to the right. In fact, one should do more than just look at the distribution, let alone just the average of the distribution, since that could mask gross winners and gross losers even if there is a shift to the right. A full analysis of the heterogeneity of welfare impacts must dig beneath the summary distributions in Fig. 3, but still using the individual welfare evaluations used to construct Fig. 3. To the extent that the CS distribution in the left panel of Fig. 3 has any negative values, and the Efficiency distribution in the right panel is not all piled up at 100%, there are welfare gains to be had from appropriate interventions.

[33] There are circumstances, discussed by Harrison and Ng (2018), such as variations in nonperformance risk across choices, where Efficiency is a superior measure of welfare than CS.

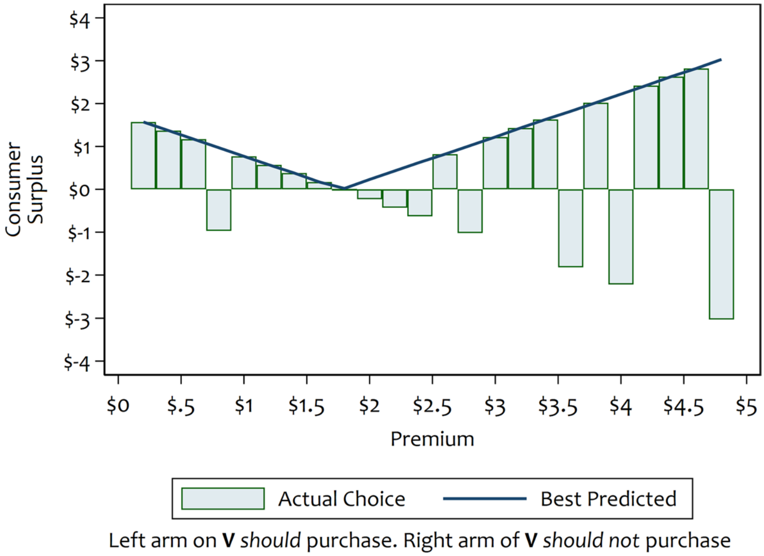

Fig. 2 Realized consumer surplus for one subject

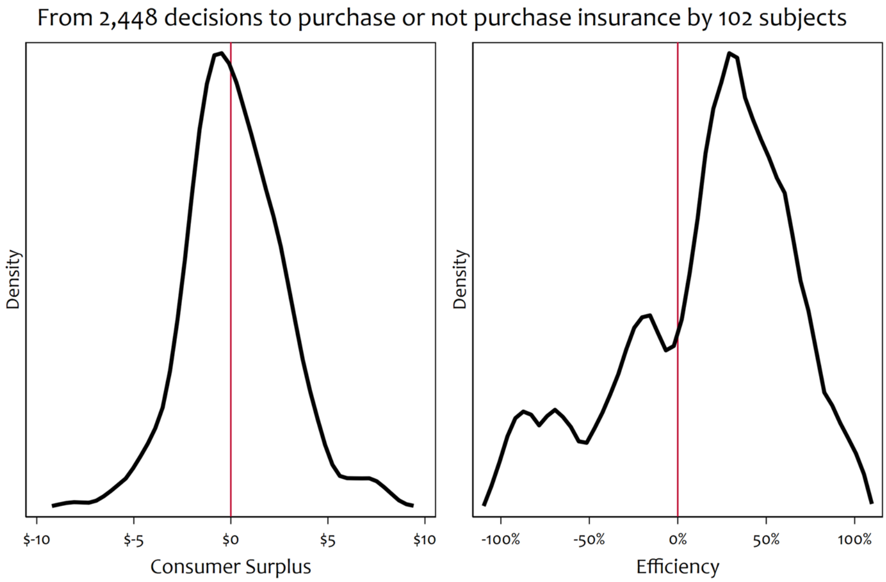

Fig. 3 Realized welfare measures

In summary, Harrison and Ng (2016) find significant evidence of welfare loss, deriving from individuals that should have purchased but did not, as well as subjects that did purchase but should not have. Their central message is that take-up is not a reliable indicator of welfare, consistent with the rhetorical behavioral concerns with take-up and revealed preferences noted earlier.

Harrison and Ng (2016) make the simplest possible assumption to undertake behavioral welfare analysis in the absence of assuming naïve revealed preference: that the risk task identifies the risk preferences for the individual, and that one can then 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 the risk task and the insurance task, for whatever "framing" reason one might think of. This assumption might be descriptively correct, and indeed would be implied conceptually if one found, as was the case, that risk preferences in the risk task do not explain every insurance choice. However, note how their assumption, or something equivalent to it, 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.

An extensive discussion of the methodological implications of this approach is provided by Harrison and Ng (2016, pp. 111–116) in the specific context of insurance. Harrison and Ross (2018) provide a similar discussion in the context of portfolio choice, and also offer a general philosophical exposition of what they characterize as the "quantitative intentional stance" toward behavioral welfare economics. We return to these issues in §5 below.

Harrison and Ng (2018) extend the methodology to consider nonperformance risk with full indemnity contracts with no deductibles. For normative purposes, their key insight is that the driving factor behind *welfare* effects of nonperformance is the extent to which the individual processes compound risks using ROCL. This structural insight is only possible if we know something more about risk preferences than in Harrison and Ng (2016): in this instance, whether they are or are not consistent with ROCL. They also find that the standard "actuarial parameters," that did have a significant effect on take-up, do not have a major effect on welfare. Again, take-up is an unreliable proxy for welfare.[34]

6 The methodologies of behavioral welfare economics

The literature just reviewed illustrates many different ways of generating welfare estimates for decisions over insurance contracts. Surprisingly, perhaps, I find the structural models of observational data for field contracts to be perfect complements to the artefactual, stylized experiments in lab settings. The former are brave attempts to directly answer important policy questions in the field, and without those full structural attempts one cannot see what remains to be done from a behavioral

[34] Harrison and Ng (2019) elaborate on how the existing literature has defined welfare gains from insurance and how it has been measured. They broadly categorize these "second-best" methods of calculating welfare gain from insurance into four groups: take-up of insurance, willingness to pay for insurance, risk reduction proxies, and "some other metric."

perspective. Generally, we only see EUT models of risk preferences rather than RDU models; we only see Exponential discounting models of time preference rather than "hyperbolicky" alternatives (if empirically relevant); we only see additive intertemporal utility functions instead of specifications that allow for intertemporal risk aversion; we only see actuarial loss probabilities assumed for individuals rather than subjective probabilities or subjective belief distributions; and we rarely see any (EUT) risk preferences or time preferences *estimated* at all. The reason for these behavioral restrictions are clear: it is hard enough to marshal the data and structural model assumptions with the data available, and identification of these behavioral preferences and beliefs is well beyond what one could reasonably expect in the first analyses. The place to demonstrate the behavioral need for augmenting these observational and administrative datasets with field experiments is in the laboratory.

The laboratory experiments show the fundamental need for knowing atemporal risk preferences, and for knowing what type of risk preferences (EUT and RDU) best characterize behavior for a population.[35] This is decidedly not just to have some measure of risk preferences to "toss on the right hand side of some regression," as one sees too often: it is to evaluate the CE needed to compute the welfare-equivalent of the insurance contract lottery under consideration. Furthermore, the laboratory evidence has shown the value of structural knowledge of deviations from EUT, by showing the significance of knowing if subjects violate ROCL when evaluating the welfare effects of nonperformance risk for insurance contracts. All of these insights leave aside roles for time preferences, intertemporal risk preferences, and subjective beliefs about loss probabilities. Much remains to be done, even in the lab.

In contrast to the studies reviewed in Sect. 5, however, most policy evaluations of insurance products use the metric of take-up. It has been shown in the laboratory experiments 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 is *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 for actuarially fair or subsidized products *and* the qualitative presumption that every decision-maker is a risk averse EUT type. This general methodological problem with nudges is well known, and even acknowledged by the original proponents of the approach: for instance, see Thaler and Sunstein (2008, ch. 17).

[35] It is EUT *and* RDU deliberately: even if individuals can be "typed" as one or the other with reasonable confidence, if we are considering pooled data then we must recognize that both are typically present by using a mixture specification (Harrison and Rutström 2009).

In this section, we step back from the specifics of behavioral evaluation of insurance contracts to make sure that the major alternative approaches to behavioral welfare evaluation are considered critically. The insurance literature already contains casual, glancing references[36] to these deeper issues, and needs to integrate them more fundamentally. There is a large, general literature on behavioral welfare economics, including Bernheim (2009, 2016), Bernheim and Rangel (2008, 2009), Manzini and Mariotti (2012, 2014), Rubinstein and Salant (2012), Salant and Rubinstein (2008) and Sugden (2004, 2009). A 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. That is what the approach advocated by Harrison and Ng (2016, 2018), Harrison et al. (2016) and Harrison and Ross (2018) seeks to do.

6.1 Nudges and boosts

A principal source of interest in behavioral economics has been its advertised contributions to policies aimed at "nudging" people away from allegedly natural but self-defeating behavior toward patterns of response thought more likely to improve their welfare. Leading early promotions of this kind of application of behavioral studies are Camerer et al. (2003) and Sunstein and Thaler (2003a, b). Grüne-Yanoff and Hertwig (2016) have distinguished nudging, which is based on the heuristics-and-biases branch of behavioral economics research associated with Kahneman et al. (1982), from policies aimed at "boosting," which apply the simple heuristics research program of Gigerenzer et al. (1999) and Hertwig et al. (2013). Nudging and boosting are contrasted as follows. Nudges aim to change a decision-maker's ecological context and external cognitive affordances in such a way that the decision-maker will be more likely to choose a welfare-improving option without having to think any differently than before. Boosts aim to supplement cognitive processes with heuristics that are viewed as reliable guides, despite glossing some information and avoiding computationally intensive algoirthms, produce good inferences, choices, and decisions when applied in the appropriate ecological contexts.[37]

[36] For example, Handel and Kolstad (2015, pp. 2451, 2456, 2490) or Ericson and Sydnor (2017, p. 70).

[37] An alternative way to define boost builds on the role that "scaffolds" play in aiding cognition (Clark 1998). Access to the internet or experts, for example, might be expected a priori to make individuals more literate on facts that affect their cognition, as better inputs to their "cognitive production function." Hence boosts need not rely on the use of heuristics. The key distinction between an algorithm and a heuristic has to do with the knowledge claim that they each allow one to make. If an algorithm has been applied correctly, then the result will be a solution that we know something about. For example, we may know that it is a local optimum, even if we do not know that it is a global optimum. Heuristics are lesser epistemological beasts: the solution provided by a heuristic has no claim to be a valid solution in the sense of meeting some criteria. In the computational literature, if not some parts of the psychological literature, heuristics are akin to "rules of thumb" that simply have good or bad track records for certain classes of problems. The track record may be defined in terms of the speed of arriving at a candidate solution, or the ease of application. Harrison (2008, §4.2) provides more discussion of the role of heuristics in decision-making, particularly their crucial role in "behaviorally plausible" homotopy, or path-following, algorithms.

An additional contrast between nudges and boosts is that a nudge would normally be expected to have effects only on the specific behavior to which it is applied, and only in the setting that the nudge adjusts. A boost, on the other hand, to the extent that it alters standing cognitive capacities and associated behavioral propensities across ranges of structurally similar choice problems, might be hoped to generate "rationality spillovers" discussed by Cherry et al. (2003). Furthermore, boosting might plausibly capacitate people with defenses against nonbenevolent nudging by narrowly self-interested parties such as marketers and demagogues.

Nudging is open to the charge that it is manipulative. Its defenders point out that if people are naturally prone to systematic error, then any scaffolding built by any institution unavoidably involves manipulation, so the manipulation in question might as well be benevolent. Of course, as stressed above, what is actually "benevolent" is typically conditional on some unobserved preference or belief ascribed to the decision-maker. How this ascription occurs is the deeper question, addressed in Sect. 6.5.

Boosting, by contrast, involves endowing decision-makers with enhanced cognitive capacities by teaching them more effective decision principles, which they can choose to apply or not once they have been enlightened. Thus, boosting avoids manipulating the agents to whom the policies in question are applied, and is to that extent less paternalistic. Boosting also begs the question of what are more "effective" decision principles. 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 many consider to be normatively attractive, it does not follow that one is unable to write down axioms that make those alternatives attractive normatively.[38] How one ascribes more effective decision principles is also considered in Sect. 6.5.

6.2 Randomized evaluations in search of "what works"

One of the slogans that burdens behavioral economic policy, and the focus on randomized evaluations, is the claim that they are only interested in "what works." It is hard to imagine a less informative, or more dangerous, slogan. The core problem is that it characterizes approaches that only look at observables.

The problem with just looking at observables is that they tell us nothing about the virtual variables that are of interest in welfare evaluation. For that we need to make inferences about CS, and for that we need to know a lot more about the preferences that people bring to their choices, such as risk preferences and time preferences.

[38] For instance, consider inverse-S probability weighting in an RDU setting, which leads the decision-maker to place greater weight on the probabilities associated with the best and worst outcomes. This might be a reasoned heuristic for recognizing that "tail probabilities" are known to be inferred less reliably, and are more reliant on parametric forms for probability distributions being correct. In fact, it characterizes one approach to "actuarial prudence" in the calculus of risk management. In terms of decision theory, it may be viewed as one way to extend the reasoning from the "small worlds" of Savage to his "large worlds" (Binmore 1999).

We also need to know a lot more about the subjective beliefs that people bring to their choices. The reason that there is this dogmatic focus on observables is easy to discern and openly discussed: a desire to avoid having to take a stand on theoretical constructs as maintained assumptions, since maintained assumptions might be wrong. The same methodological precept guides the choice of statistical methods, but that is another story about modeling costs and benefits. One can fill in these blanks in our knowledge about virtual preferences and beliefs with theories and guessed-at numbers, or with theories and estimated numbers. However, one has to use theory to make conceptually coherent statements about preferences and beliefs, and then undertake welfare evaluations. That is the rub: an agnosticism toward theory.

Advocates of randomized evaluations portray the tradeoff here in overly dramatic fashion. Either one uses the methods that avoid these theoretical constructs, or one dives head first into the shoals of full structural modeling of behavior. This is a false dichotomy, raised as a cheap rhetorical device to still debate over the role of theory. The missing middle ground becomes apparent when empirical puzzles emerge, leading to casual theorizing and even more casual behaviorism, documented in Harrison (2011a).

In any event, randomized evaluations can be wonderful tools for gathering information about the cost-effectiveness of alternative policies toward some given goal, but are silent on the real question of the net welfare of those policies.

6.3 Happiness

No thanks. The same applies to surveys of "life satisfaction," "well-being," or even "peace of mind."

6.4 Behavioral welfare analysis using "frames"

Bernheim and Rangel (2008, 2009) and Bernheim (2009, 2016) present an approach to behavioral welfare economics that recognizes the methodological challenge of evaluating welfare when one does not accept that one can rely on (naive) revealed preference. They propose that one develops two frames with which to ask a question bearing on financial choices, where two conditions are met, and are couched here in terms of a financial literacy application:

1. Each frame is a priori presumed to generate actions that have the same welfare consequences for the individuals.
2. However, where one frame is simple and transparent to understand, so a priori does not require any significant degree of literacy to comprehend, the other frame requires some degree of financial literacy to comprehend.

Note that both conditions rely on a priori judgments. There is nothing wrong with this, but of course the "proof is in the pudding" when one gets to specific applications, and different readers might have different priors on the validity of these two conditions.[39] The application of these ideas in Ambuehl et al. (2014) and Ambuehl et al. (2018), reviewed in Bernheim (2016, p. 51ff.), provide just such an instance, focused squarely on financial literacy.[40]

The application in each case is the same, and tests comprehension of the concept of compound interest as it affects intertemporal choices between a smaller, sooner (SS) amount of money and a larger, later (LL) amount of money. This is a canonical task for the elicitation of time preferences: see Coller and Williams (1999) for an extensive review of the older literature and clean experimental implementation of this task. To illustrate, consider these two statements, which very slightly paraphrase those actually used:

A. You will receive \$88 in 72 days.
B. We will invest \$22 at 3% interest, compounded daily, for 72 days.

Subjects are then asked, in response to one of these statements, to say "what is the present amount that is equivalent?" Responses are elicited using an Iterative Multiple Price List (iMPL) procedure developed by Andersen et al. (2006), and can be assumed for present purposes to lead subjects to reveal their true answer in an incentive-compatible manner.

If subjects exhibit financial literacy they "should" give the same answers in response to statements A and B, since we observers know that the amount of money in B will end up being \$88 in 72 days. If the answers to A and B differ, then we have identified a financial literacy gap, and it is asserted that we can take the absolute value of the difference in valuations as a measure of the welfare loss from that gap. Since the present value amounts are stated in deterministic form, this welfare loss is in the form of a certainty equivalent. In effect, here, the observed choice is a willingness to exchange the LL amount mentioned or implied by statement A or B for the SS amount stated in the response elicited by the iMPL procedure.

[39] This is the approach adopted in Ambuehl et al. (2014), to view one of the frames as revealing true, virtual valuations. In Ambuehl et al. (2017), this position was qualified, allowing that there might be some normative metric that does not lead one to accept that either frame represents the true, virtual valuation. The example provided is when subjects exhibit Quasi-Hyperbolic discounting in response to both questions, with Exponential discounting a priori deemed to be normatively attractive and Quasi-Hyperbolic discounting deemed a priori to be normatively unattractive. In this case, they claim, both responses might be "contaminated" by the "passion for the present" one expects from Quasi-Hyperbolic responses. They then present a formal mathematical result that essentially says that if the responses to statements A and B are equally contaminated, then as one takes the *limit of the difference between the responses as that difference goes to zero*, then a first-order approximation to a valid welfare measure can be obtained. However, that says nothing about whether the difference between the responses that are nonzero, or not close to zero, have any valid interpretation, unless one wants to invoke stringent path-independence assumptions from welfare economics (see Boadway and Bruce (1984, p. 199) or Harrison et al. (1993). The bulk of responses of interest are decidedly nonzero, and not close to zero, as illustrated in Ambuehl et al. (2018, p. 16, Fig. 1).

[40] An application of the same methodology to retirement savings plans is provided by Bernheim et al. (2015).

Now consider if statements A and B meet the conditions required for inferences about welfare loss due to financial illiteracy.

One immediate concern is that statement B might be interpreted, from a conversational perspective, as already providing the answer: surely it is $22. The interpretation is that you have been asked what amount of money today would generate the implied $88 in 72 days, and this must be a "trick question" because the statement already tells you that it was $22. Of course, we analysts are expecting subjects to tell us the present discounted amount that is equivalent to $88 in 72 days, where the discount rate need not be the same as the interest rate, but that is just one interpretation of the question. One might expect, if inspecting the raw data, to see many respondents simply say $22 in this instance.

Another, more subtle, interpretation issue concerns the information about a 3% interest rate. A subject might reasonably presume that this is taken to be the market (borrowing and lending) interest rate for this question. Then we know from the Fisher Separation Theorem that we cannot recover estimates of discount rates due to censoring: see Coller and Williams (1999). All that we would recover is their knowledge of the interest rate, which is again included in statement B, hence we would again expect a spike of responses at $22.

Extending this point, the mere mention of interest rates provides a scaffold that might affect responses differently for statement B compared to statement A. In effect, statement B offers a cognitive scaffold that could be expected to change the response compared to statement A, where there is no such explicit scaffold mentioned. Thus, what is claimed to be the welfare effect of literacy might just be the welfare effect of having access to a scaffold, and that is ambiguous as a theoretical matter.

Finally, any difference between responses to statements A and B might simply reflect an inability to apply the principle of compound interest in evaluating statement B, to arrive at the implied $88 correctly. A subject might understand what compound interest is, and just not be able to "do the math" on the spot, even with a calculator provided. The issue here is whether one labels any difference in present value responses a welfare-significant failure of literacy with respect to the concept of compound interest or a welfare-significant failure of the ability to *apply* the correct concept (recall the distinction between literacy and capability). And the focus throughout Ambuehl et al. (2014, 2017, 2018) is on the effect of an intervention to improve decision-making, whether or not it is literacy or capability that is driving the effect.

One overarching concern here is that to apply the method of Berneim and Rangel (2008, 2009), one must find frames that convince readers that they meet the two conditions noted earlier, and this is not likely to be an easy task across domains. Their method is not, in this sense, a general method.

6.5 Welfare analysis from the intentional stance

Harrison and Ng (2016) use the best descriptive model of risk preferences to make normative evaluations of the insurance product choices of their subjects.[41] By contrast, Bleichrodt et al. (2001) maintain that EUT is the appropriate normative model, and correctly note that if an individual is an RDU or CPT decision-maker, then recovering the utility function from observed lottery choices requires allowing for probability weighting and/or sign-dependence. They then implicitly propose using *that* utility function to infer the CE using EUT. These are radically different normative positions.

Some notation will help. Let RDU(x) denote the evaluation of an insurance policy x in Harrison and Ng (2016) using the RDU risk preferences of the individual, including the probability weighting function. They calculate the CE by solving $U^{\mathrm{RDU}}(\mathrm{CE}) = \mathrm{RDU}(x)$ for CE, where U^{RDU} is the utility function from the RDU model of risk preferences for that individual. However, Bleichrodt et al. (2001) evaluate the CE by solving $U^{\mathrm{RDU}}(\mathrm{CE}) = \mathrm{EUT}(x)$ where EUT(x) uses that utility function in an EUT manner, assuming no probability weighting. This seems normatively illogical. The logical approach here would be to estimate the "best fitting EUT risk preferences" for the individual from their observed lottery choices, and then use the resulting utility function U^{EUT} as the basis for evaluating the CE using $U^{\mathrm{EUT}}(\mathrm{CE}) = \mathrm{EUT}(x)$.

The choice of this approach by Harrison and Ng (2016) and Harrison and Ross (2018) is evidently of direct relevance with respect to the extent of paternalism involved in normative assessment. It can be justified on deeper philosophical grounds.

Harrison and Ross (2018) included a case study from a consulting project undertaken for a client that is an investment bank. They recommended additional cognitive preparation for RDU choosers before they selected investment products, but did not recommend trying to teach them the concept of probability weighting so they could then apply this characterization to themselves. This is only partly motivated by the questionable practicality of the pedagogical task that would be required. It also reflects wariness about telling subjects a story about themselves they would surely interpret as telling them that they possess a kind of internal psychological "defect" when such a story would outrun the available data and is, in any case, doubtful according to sophisticated philosophy of mind.

It is unlikely that most people choosing insurance contracts or investment funds attempt to compute internally represented optima, either from EUT or RDU bases, and then make computational errors that could be pointed out to them. This echoes a point made by Infante et al. (2016) when they complain that behavioral welfare economists typically follow Hausman (2011) in "purifying" empirically observed

[41] Harrison and Ross (2018) employ the same methodological approach in the evaluation of choices over alternative investment portfolios. The exposition in this subsection is adapted from Harrison and Ross (2018, §5).

preferences. Infante et al. (2016) argue that purification reflects an implicit philosophy according to which an inner Savage-rational agent is trapped within a psychological, irrational shell from which best policy should try to rescue her. They provide no general philosophical framework within which they motivate their skepticism about "inner rational agents." However, such a framework is available.

Dennett (1987) provides a rich account of the relationships between beliefs, preferences and "propositional attitudes" that provides a rigorous philosophical foundation for behavioral welfare economics. He argues that the attribution of preferences and beliefs involves taking an intentional stance toward understanding the behavior of an agent. This stance consists in assuming that the agent's behavior is guided by goals and is sensitive to information about means to the goals, and about the relative probabilities of achieving the goals given available means. Goals, like preferences and beliefs, are *not internal states of agents*, but are rather relationships between agents, environments, and those of us that are attributing these relationships in order to do behavioral welfare economics. Hence there is a crisp rejection at the outset of the realist conception of economic agents presented in naïve behavioral rhetoric, such as the "humans" and "econs" of Thaler and Sunstein (2008, Part I).

Hence the behavioral welfare economist, by this account, has to try to interpret and predict the agent's actions by means of controlled speculation about that agent's context and information-processing capacities. Agents themselves are trained, during socialization while growing up, to adopt the intentional stance toward themselves. For the sake of coordination in action and communication, agents' self-ascriptions are made so as to at least approximate alignment with the ascriptions of others. These ascriptions and self-ascriptions are not guesses about "true" beliefs and preferences hidden from direct view in people's heads. Rather, beliefs and preferences are constructed rationalizations of agents' behavioral and cognitive ecologies.[42]

Beliefs and preferences are virtual states[43] of whole intentional systems rather than particular physical states of brains; but being virtual is a way of being real, not a way of being fictitious. If a claim about intentional states is the sort of claim that can have a truth value, then it had better be possible to specify possible evidence that would undermine it. The holistic nature of intentional stance description of agent behavior allows for error, but also complicates it: as stressed by Hey (2005), the "behavioral error" stories that we append to our structural models are part of the economics.[44]

[42] Critics have sometimes misinterpreted this view as instrumentalism, a doctrine according to which beliefs and preferences are mere useful fictions, unconstrained from the "facts of the matter." Dennett (1987) has consistently maintained, however, that there are facts about agents' goals and access to information, and hence also facts about their propositional attitudes, that should constrain these rationalizations.

[43] Economists often use the expression "latent states" to mean the same thing. Unfortunately, there are significant complications with the use of the term "latent" when one interacts with philosophers and psychologists, and behavioral welfare economics must interact with them. Harrison and Ross (2018, fn.17, p. 65) and Ross (2014, §4.2) discusses the complications and why they matter.

[44] To add complication, they interact directly with the stochastic specifications that attend to sampling errors in the econometrics, and hence inferences about preferences: see Wilcox (2008, 2011) for a masterful review in the case of risk preferences.

Ross (2014) argues that this marks a main basis for the distinction between economics and psychology. Psychologists are professionally interested directly in how individuals process information, including information that influences decisions. Economists, by contrast, are concerned with this only derivatively. If a system of incentives will lead various people, through a heterogeneous set of psychological processes, to all make the same choice then the people form, at least for an analysis restricted to that choice, an equivalence class of economic agents. However, it is a strictly empirical matter when this psychological heterogeneity will and won't matter economically. Economists, like all scientists, seek generalizations that support out-of-sample predictions. Different data-generating processes tend to produce, sooner or later, different data, including different economic data. Economics is thus crucially informed by psychology in general, while not collapsing into the psychology of valuation as some behavioral economists have urged [e.g., Camerer et al. (2005)].

Applying this philosophy of mind and agency to the applications to insurance in Harrison and Ng (2016), we assume the intentional stance to make sense of our experimental subjects' overall behavioral patterns, and use the lottery choice experiment as a relatively direct source of constraint on the virtual preference structures we assign when we perform welfare assessment of their insurance contract choices. The more precisely we specify the contents of propositional attitudes, especially in quantitative terms, the less weight in identification will rest on "inboard" elements of data-generating processes relative to external aspects of the agents' overall behavioral ecologies (i.e., cognitive scaffolds). Our technical tools allow us to identify virtual intentions that most subjects are not able to identify when they take the intentional stance to themselves, and that they could not deliberately use to evaluate their own decisions.[45] On the other hand, certain experimental treatments[46] might provide evidence that attention to certain informational patterns induces a significant number of subjects to act as if they were stochastically closer to expected utility optimizers. These patterns therefore enter into a fully informed analyst's specification of the subjects' beliefs and preferences.[47]

6.6 The opportunity criterion

Sugden (2004, 2009, 2018) develops an important framework for normatively evaluating agents' outcomes under alternative institutional arrangements in a way that privileges their autonomy as choosers (i.e., their consumer sovereignty) without depending on their specific preference orderings, and thus without requiring their

[45] Hence, again, the irrelevance of the derisive comments of some behavioral economists toward their straw man account of the agent being modeled, on the grounds that nobody actually makes decisions the way our intentional stance posits.

[46] For example, the informational treatment of Harrison and Ross (2018) with respect to investment decisions, or the "actuarially-equivalent" insurance contract treatment of Harrison et al. (2016) with respect to index insurance decisions.

[47] In this philosophical framework, it makes sense to say that we *boost* the subjects' informational access in a way that *nudges* their (subdeliberative) cognition.

preferences to even be consistently ordered, let alone fully EUT-compliant. According to Sugden (2004, 2009), agents are made better off to the extent that their *opportunity sets* are expanded, and worse off to the extent that their *opportunity sets* are contracted. Against this standard, "pure" boosts will typically make agents better off and "pure" nudges will typically make them worse off.

This idea is indeed attractive as a way of addressing normative questions in circumstances where welfare analysis in the technical sense is not possible due to preference reversals. Thus, for example, this approach can generate recommendations in cases where the method of Bernheim (2009, 2016) and Bernheim and Rangel (2008, 2009) would find Pareto indifference and therefore yield no guidance. However, we should not abjure ever doing standard welfare analysis merely because it can't be undertaken in every context. In both the Harrison and Ng's (2016) case and in the situation presented to Harrison and Ross (2018) by their consulting client, the complications arise from the existence of preferences that violate EUT but are nevertheless well ordered. Arguably, this is the standard situation where relevant utilities are defined over expected monetary values that are risky.

7 Conclusion

From a normative perspective, a great deal of attention has been devoted to design better insurance *products*. It is apparent from the existing evidence that comparable attention should be devoted to designing better insurance *decisions*. 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 real tension here, just the need to have a clear, structured ability to say something about the welfare effect of product innovations *and* the decision process surrounding the product.

These requirements generate a derived demand for thinking carefully about the methodologies of behavioral welfare economics, and that requires that insurance economists, and economists in general, think more deeply about the methodology and philosophy of their subject. In this respect, thought experiments and laboratory experiments stand as ideal places to begin this long, slippery path. One major risk we face, and that is tragically illustrated by what currently passes for behavioral economic policy, is that the new behavioral economics causes us to forget the old welfare economics.[48] In the words of Homer Simpson (season 5, episode 22), "every time I learn something new, it pushes some old stuff out of my brain."

[48] More generally, Atkinson (2001, 2009, 2011) eloquently considers "the strange disappearance of welfare economics."

References

Ambuehl, Sandro, B. Douglas Bernheim, Fulya Ersoy, and Donna Harris. 2018. Peer Advice on Financial Decisions: A Case of the Blind Leading the Blind? *Working Paper 25034*, National Bureau of Economic Research.

Ambuehl, Sandro, B. Douglas Bernheim, and Annamaria Lusardi. 2014. The Effect of Financial Education on the Quality of Decision Making. *Working Paper 20618*, National Bureau of Economic Research.

Ambuehl, Sandro, B. Douglas Bernheim, and Annamaria Lusardi. 2017. A Method for Evaluating the Quality of Financial Decision Making, with an Application to Financial Education. *Working Paper 20618*, National Bureau of Economic Research.

Andersen, Steffen, James C. Cox, Glenn W. Harrison, Morten I. Lau, E.Elisabet Rutström, and Vjollca Sadiraj. 2018a. Asset Integration and Attitudes to Risk: Theory and Evidence. *Review of Economics & Statistics* 100 (5): 816–830.

Andersen, Steffen, John Fountain, Glenn W. Harrison, and Elisabet Rutström, 2014a. Estimating Subjective Probabilities. *Journal of Risk & Uncertainty* 48 (3): 207–229.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2006. Elicitation Using Multiple Price List Formats. *Experimental Economics* 9 (4): 383–405.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. Eliciting Risk and Time Preferences. *Econometrica* 76 (3): 583–618.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2013. Discounting Behavior and the Magnitude Effect: Evidence from a Field Experiment in Denmark. *Econometrica* 80: 670–697.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2014b. Discounting Behavior: A Reconsideration. *European Economic Review* 71: 15–33.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E.Elisabet Rutström. 2018b. Multiattribute Utility Theory, Intertemporal Utility, and Correlation Aversion. *International Economic Review* 59 (2): 537–555.

Arrow, Kenneth J. 1968. The Economics of Moral Hazard: Further Comment. *American Economic Review* 58 (3): 537–539.

Atkinson, Anthony B. 2001. The Strange Disappearance of Welfare Economics. *Kyklos* 54 (2/3): 193–206.

Atkinson, Anthony B. 2009. Economics as a Moral Science. *Economica* 76: 791–804.

Atkinson, Anthony B. 2011. The Restoration of Welfare Economics. *American Economic Review (Papers & Proceedings)* 101 (3): 157–161.

Baily, Martin Neil. 1978. Some Aspects of Optimal Unemployment Insurance. *Journal of Public Economics* 10 (3): 379–402.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. The Nature of Risk Preferences: Evidence from Insurance Choices. *American Economic Review* 103 (6): 2499–2529.

Bernheim, B.Douglas. 2009. Behavioral Welfare Economics. *Journal of the European Economic Association* 7 (2–3): 267–319.

Bernheim, B. Douglas. 2016. The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics. *Journal of Benefit-Cost Analysis* 7 (1): 12–68.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. The Welfare Economics of Default Options in 401(k) Plans. *American Economic Review* 105 (9): 2798–2837.

Bernheim, B. Douglas, and Antonio Rangel. 2008. Choice-Theoretic Foundations for Behavioral Welfare Economics. In *The Foundations of Positive and Normative Economics: A Handbook*, ed. A. Caplin and A. Schotter. Oxford: Oxford University Press.

Bernheim, B. Douglas, and Antonio Rangel. 2009. Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics* 124 (1): 51–104.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. Choose to Lose: Health Plan Choices from a Menu with Dominated Options. *Quarterly Journal of Economics* 132 (3): 1319–1372.

Binmore, Ken. 1999. *Rational Decisions*. Princeton: Princeton University Press.

Blackburn, McKinley, Glenn W. Harrison, and E.Elisabet Rutstrom. 1994. Statistical Bias Functions and Informative Hypothetical Surveys. *American Journal of Agricultural Economics* 76 (5): 1084–1088.

Blackorby, Charles, and David Donaldson. 1988. Money Metric Utility: A Harmless Normalization? *Journal of Economic Theory* 46 (1): 120–129.

Bleichrodt, Han, J.L. Pinto, and Peter P. Wakker. 2001. Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. *Management Science* 47: 1498–1514.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. Consumption Inequality and Partial Insurance. *American Economic Review* 98 (5): 1887–1921.

Blundell, Richard, Ian Preston, and Ian Walker. 1994. Introduction to Applied Welfare Economics. In *The Measurement of Household Welfare*, ed. R. Blundell, I. Preston, and I. Walker. New York: Cambridge University Press.

Bommier, Antoine. 2006. Uncertain Lifetime and Intertemporal Choice: Risk Aversion as a Rationale for Time Discounting. *International Economic Review* 47: 1223–12246.

Bommier, Antoine. 2010. Portfolio Choice Under Uncertain Lifetime. *Journal of Public Economic Theory* 12: 57–73.

Bommier, Antoine. 2013. Life-Cycle Preferences Revisited. *Journal of the European Economic Association* 11: 1290–1319.

Bommier, Antoine, and Jean-Charles Rochet. 2006. Risk Aversion and Planning Horizons. *Journal of the European Economic Association* 4: 708–734.

Boadway, Robin, and Neil Bruce. 1984. *Welfare Economics*. Oxford: Blackwell.

Bundorf, M.Kate, Jonathan Levin, and Neale Mahoney. 2012. Pricing and Welfare in Health Plan Choice. *American Economic Review* 102 (7): 3214–3248.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism. *University of Pennsylvania Law Review* 151: 1211–1254.

Camerer, Colin, George Loewenstein, Ted O'Donoghue, and Drazen Prelec. 2005. Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature* 43: 9–64.

Casaburi, Lorenzo, and Jack Willis. 2018. Time vs. State in Insurance: Experimental Evidence from Contract Farming in Kenya. *American Economic Review* 108 (12): 3778–3813.

Cherry, Todd, Thomas Crocker, and Jason Shogren. 2003. Rationality Spillovers. *Journal of Environmental Economics and Management* 45: 63–84.

Chetty, Raj. 2006. A General Formula for the Optimal Level of Social Insurance. *Journal of Public Economics* 90 (10–11): 1879–1901.

Chetty, Raj. 2009. Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced Form Estimates. *Annual Review of Economics* 1: 451–488.

Chetty, Raj, and Adam Looney. 2006. Consumption Smoothing and the Welfare Consequences of Social Insurance in Developing Countries. *Journal of Public Economics* 90: 2351–2356.

Clark, Andy. 1998. *Being There*. Cambridge: MIT Press.

Cohen, Alma, and Liran Einav. 2007. Estimating Risk Preferences from Deductible Choices. *American Economic Review* 97 (3): 745–788.

Coller, Maribeth, and Melonie B. Williams. 1999. Eliciting Individual Discount Rates. *Experimental Economics* 2 (2): 107–127.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford, and Orlanda Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton: Princeton University Press.

Cox, James C., and Vjollca Sadiraj. 2006. Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory. *Games and Economic Behavior* 56: 45–60.

Cox, James C., and Vjollca Sadiraj. 2008. Risky Decisions in the Large and in the Small: Theory and Experiment. In *Risk Aversion in Experiments*, ed. J. Cox and G.W. Harrison. Bingley: Emerald, Research in Experimental Economics.

Cutler, David M., Amy Finkelstein, and Kathleen McGarry. 2008. Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. *American Economic Review (Papers & Proceedings)* 98 (2): 157–162.

Dennett, Daniel. 1987. *The Intentional Stance*. Cambridge: MIT Press.

Di Girolamo, Amalia, Glenn W. Harrison, Morten I. Lau, and J.Todd Swarthout. 2015. Subjective Belief Distributions and the Characterization of Economic Literacy. *Journal of Behavioral and Experimental Economics* 59: 1–12.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010a. Estimating Welfare in Insurance Markets Using Variation in Prices. *Quarterly Journal of Economics* 125 (3): 877–921.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2010b. Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market. *Econometrica* 78 (3): 1031–1092.

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

Ericson, Keith Marzilli, and Justin Sydnor. 2017. The Questionable Value of Having a Choice of Levels of Health Insurance Coverage. *Journal of Economic Perspectives* 31 (4): 51–72.

Feldman, Roger, and Bryan Dowd. 1991. A New Estimate of the Welfare Loss of Excess Health Insurance. *American Economic Review* 81 (1): 297–301.

Feldstein, Martin S. 1973. The Welfare Loss of Excess Health Insurance. *Journal of Political Economy* 81 (2): 251–280.

Friedman, Bernard. 1974. Risk Aversion and the Consumer Choice of Health Insurance Option. *Review of Economics and Statistics* 56 (2): 209–214.

Gangadharan, Lata, Glenn W. Harrison, and Anke Leroux. 2019. Are Risks Over Multiple Attributes Traded Off? A Case Study of Aid. *Journal of Economic Behavior & Organization* 164: 166–198.

Gigerenzer, Gerd, Peter Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.

Grüne-Yanoff, Till, and Ralph Hertwig. 2016. Nudge vs. Boost: How Coherent Are Policy and Theory? *Minds and Machines* 26: 149–183.

Handel, Bengamin R. 2013. Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review* 103 (7): 2643–2682.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review* 105 (8): 2449–2500.

Hansen, Jan V., Rasmus H. Jacobsen, and Morten I. Lau. 2016. Willingness to Pay for Insurance in Denmark. *Journal of Risk and Insurance* 83 (1): 49–76.

Hansson, Bengt. 1988. Risk Aversion as a Problem of Conjoint Measurement. In *Decisions, Probability, and Utility*, ed. P. Gardenfors and N.-E. Sahlin. New York: Cambridge University Press.

Harrison, Glenn W. 2005. Hypothetical Bias Over Ucertain Outcomes. In *Using Experimental Methods in Environmental and Resource Economics*, ed. J.A. List. Northampton: Edward Elgar.

Harrison, Glenn W. 2008. Neuroeconomics: A Critical Reconsideration. *Economics & Philosophy* 24 (3): 303–344.

Harrison, Glenn W. 2011a. Randomisation and Its Discontents. *Journal of African Economies* 20: 626–652.

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

Harrison, Glenn W. 2013. Field Experiments and Methodological Intolerance. *Journal of Economic Methodology* 20 (2): 103–117.

Harrison, Glenn W. 2014a. Impact Evaluation and Welfare Evaluation. *European Journal of Development Research* 26: 39–45.

Harrison, Glenn W. 2014b. Cautionary Notes on the Use of Field Experiments to Address Policy Issues. *Oxford Review of Economic Policy* 30 (4): 753–763.

Harrison, Glenn W. 2014c. Real Choices and Hypothetical Choices. In *Handbook of Choice Modeling*, ed. S. Hess and A. Daly. Northampton: Edward Elgar.

Harrison, Glenn W., Morten I. Lau, Don Ross, and J.Todd Swarthout. 2017a. Small Stakes Risk Aversion in Experiments: A Reconsideration. *Oxford Review of Economic Policy* 160: 24–28.

Harrison, Glenn W., Jimmy Martínez-Correa, Jia Min Ng, and J. Todd Swarthout. 2016. Evaluating the Welfare of Index Insurance. *CEAR Working Paper 2016-07*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.

Harrison, Glenn W., Jimmy Martínez-Correa, and J.Todd Swarthout. 2015. Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence. *Journal of Economic Behavior & Organization* 119: 32–55.

Harrison, Glenn W., Jimmy Martínez-Correa, J.Todd Swarthout, and Eric Ulm. 2017b. Scoring Rules for Subjective Probability Distributions. *Journal of Economic Behavior & Organization* 134: 430–448.

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

Harrison, Glenn W., and Jia Min Ng. 2018. Welfare Effects of Insurance Contract Non-performance. *Geneva Risk and Insurance Review* 43 (1): 39–76.

Harrison, Glenn W., and Jia Min Ng. 2019. Behavioral Insurance and Economic Theory: A Literature Review. *Risk Management & Insurance Review* 22: 133–182.

Harrison, Glenn W., and Don A. Ross. 2017. The Empirical Adequacy of Cumulative Prospect Theory and Its Implications for Normative Assessment. *Journal of Economic Methodology* 24 (2): 150–165.

Harrison, Glenn W., and Don A. Ross. 2018. Varieties of Paternalism and the Heterogeneity of Utility Structures. *Journal of Economic Methodology* 25 (1): 42–67.

Harrison, Glenn W., Thomas F. Rutherford, and Ian Wooton. 1993. An Alternative Welfare Decomposition for Customs Unions. *Canadian Journal of Economics* 26 (4): 961–968.

Harrison, Glenn W., and E.Elisabet Rutström. 2006. Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In *Handbook of Experimental Economics Results*, ed. C.R. Plott and V.L. Smith. Amsterdam: North-Holland.

Harrison, Glenn W., and E.Elisabet Rutström. 2008. Risk Aversion in the Laboratory. In *Risk Aversion in Experiments*, ed. J.C. Cox and G.W. Harrison. Bingley: Emerald, Research in Experimental Economics.

Harrison, Glenn W., and E.Elisabet Rutström. 2009. Expected Utility *And* Prospect Theory: One Wedding and A Decent Funeral. *Experimental Economics* 12 (2): 133–158.

Harrison, Glenn W., and J. Todd Swarthout. 2016. Cumulative Prospect Theory in the Laboratory: A Reconsideration. *CEAR Working Paper 2016-05*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University. In *Prospect Theory as a Model of Risky Choice: Descriptive and Normative Assessments*, ed. G.W. Harrison, and D. Ross. Bingley: Emerald, Research in Experimental Economics, forthcoming.

Hausman, Daniel. 2011. *Preference, Value, Choice and Welfare*. Cambridge: Cambridge University Press.

Hausman, Jerry A., and Whitney K. Newey. 2016. Individual Heterogeneity and Average Welfare. *Econometrica* 84 (3): 1225–1248.

Hertwig, Ralph, Ulrich Hoffrage, and the ABC Research Group. 2013. *Simple Heuristics in a Social World*. Oxford: Oxford University Press.

Hey, John D. 2005. Why We Should Not Be Silent About Noise. *Experimental Economics* 8 (4): 325–345.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016. Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioral Welfare Economics. *Journal of Economic Methodology* 23: 1–25.

Jaspersen, Johannes G. 2016. Hypothetical Surveys and Experimental Studies of Insurance Demand: A Review. *Journal of Risk and Insurance* 83 (1): 255–271.

Kahneman, Daniel, Paul Slovic, and Amos Tversky (eds.). 1982. *Judgment Under Uncertainty: Heuritics and Biases*. Cambridge: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47: 263–291.

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

Loewenstein, George, Joelle Y. Friedman, Barbara McGill, Sarah Ahmad, Suzanne Linck, Stacey Sinkula, John Beshears, James J. Choi, Jonathan Kolstad, David Laibson, Brigitte C. Madrian, John A. List, and Kevin G. Volpp. 2013. Consumers' Misunderstanding of Health Insurance. *Journal of Health Economics* 32 (5): 850–862.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz, and Susan M. Marquis. 1987. Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *American Economic Review* 77: 251–277.

Manzini, Paola, and Marco Mariotti. 2012. Categorize then Choose: Boundedly Rational Choice and Welfare. *Journal of the European Economic Association* 10 (5): 1141–1165.

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

Newhouse, Joseph P. 2015. "Introduction", in Finkelstein, Amy, *Moral Hazard in Health Insurance*. New York: Columbia University Press.

Pauly, Mark V. 1968. The Economics of Moral Hazard: Comment. *American Economic Review* 58 (3): 531–537.

Rabin, Matthew. 2000. Risk Aversion and Expected Utility Theory: A Calibration Theorem. *Econometrica* 68: 1281–1292.

Ross, Don. 2005. *Economic Theory, and Cognitive Science: Microexplanation*. Cambridge, MA: MIT Press.
Ross, Don. 2014. *Philosophy of Economics*. London: Palgrave Macmillan.
Rubinstein, Ariel, and Yuval Salant. 2012. Eliciting Welfare Preferences from Behavioral Datasets. *Review of Economic Studies* 79 (1): 375–387.
Salant, Yuval, and Ariel Rubinstein. 2008. (A,f): Choice with Frames. *Review of Economic Studies* 75 (4): 1287–1296.
Samphantharak, Krislert, and Robert M. Townsend. 2010. *Households as Corporate Firms: An Analysis of Household Finance Using Integrated Household Surveys and Corporate Financial Accounting*. New York: Cambridge University Press.
Samuelson, Paul A. 1974. Complementarity: An Essay on the 40th Anniversary of the Hicks–Allen Revolution in Demand Theory. *Journal of Economic Literature* 12 (4): 1255–1289.
Slesnick, Daniel T. 1998. Empirical Approaches to the Measurement of Welfare. *Journal of Economic Literature* 36 (4): 2108–2165.
Small, Kenneth A., and Harvey S. Rosen. 1981. Applied Welfare Economics with Discrete Choice Models. *Econometrica* 49 (1): 105–130.
Sugden, Robert. 2004. The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *American Economic Review* 94 (4): 1014–1033.
Sugden, Robert. 2009. Market Simulation and the Provision of Public Goods: A Non-Paternalistic Response to Anomalies in Environmental Evaluation. *Journal of Environmental Economics and Management* 57: 87–103.
Sugden, Robert. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. New York: Oxford University Press.
Sunstein, Cass, and Richard H. Thaler. 2003a. Libertarian Paternalism. *American Economic Review (Papers & Proceedings)* 93: 175–179.
Sunstein, Cass, and Richard H. Thaler. 2003b. Libertarian Paternalism is Not an Oxymoron. *University of Chicago Law Review* 70: 1159–1202.
Sydnor, Justin. 2010. (Over)insuring Modest Risk. *American Economic Journal: Applied Economics* 2 (4): 177–199.
Thaler, Richard H., and Cass Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth and Happiness*. New Haven: Yale University Press.
Townsend, Robert M. 1994. Risk and Insurance in Village India. *Econometrica* 62 (3): 539–591.
Wilcox, Nathaniel T. 2008. Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison. In *Risk Aversion in Experiments*, ed. J.C. Cox and G.W. Harrison. Bingley: Emerald, Research in Experimental Economics.
Wilcox, Nathaniel T. 2011. 'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk. *Journal of Econometrics* 162 (1): 89–104.