Received: 1 May 2019 | Accepted: 1 May 2019

DOI: 10.1111/rmir.12119

INVITED ARTICLE

WILEY

Behavioral insurance and economic theory: A literature review

Glenn W. Harrison[1,3] | **Jia Min Ng**[2]

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

[2]Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, Georgia

[3]School of Economics, University of Cape Town, Cape Town, South Africa

Correspondence
Glenn W. Harrison, Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, Georgia.
Email: gharrison@gsu.edu

Jia Min Ng, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, Georgia.
Email: jiamin.ng@outlook.com

Abstract

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 theories of the demand for, and welfare evaluation of, insurance products. These theories extend relatively easily to the insights of behavioral economics. Unfortunately, the empirical literature has not maintained this tight connection. In fact, much of the empirical literature illustrates the dangers of the modern passion with agnostic economics: avoiding theory at all costs to focus on "what works." We identify these dangers and the implications in the literature.

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 theories of the demand for, and welfare evaluation of, insurance products. These theories extend relatively easily to the insights of behavioral economics.[1] Unfortunately, the empirical literature has not maintained this tight connection with theory. In fact, much of the empirical literature illustrates the dangers of the modern passion with agnostic

[1]Behavioral economics is not defined here, nor should it be, as just the study of anomalies or irrationality.

economics: avoiding theory at all costs to focus on "what works." We identify these dangers and the implications for the behavioral insurance literature.[2]

To keep the focus tight we limit discussions to empirical studies that use the methods of experimental economics, whether it be in the laboratory or the field. This includes studies that exploit naturally occurring data that offer many of the controls of experiments.[3] We generally avoid any mention of studies using hypothetical surveys, because of the overwhelming evidence of hypothetical bias across most behavioral domains.[4]

From a theoretical perspective, we 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 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 then 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 1 and 2 provide "helicopter tours" of key issues in theory and experiments that we view as central to evaluating the literature. Section 3 reviews descriptive behavioral experiments on insurance, and Section 4 reviews normative behavioral experiments on insurance.

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 (WTP) for a full indemnity insurance contract with no deductible.

And 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

[2]General discussions of the methodological implication of a theoretical behavioral research are provided by Harrison (2013, 2014a, 2014b) and Spiegler (2019).

[3]Harrison and List (2004) carefully review the taxonomy of different types of experiments. The dictionary always provides a useful check on semantic confusions. Consider the *Oxford English Dictionary* (*Second Edition*), and the definitions of the noun "experiment" in science: "An action or operation undertaken in order to discover something unknown, to test a hypothesis, or establish or illustrate some known truth." There is no direct mention of randomization or causality. The verb "control" is defined in the following manner: "To exercise restraint or direction upon the free action of; to hold sway over, exercise power or authority over; to dominate, command." So the word means something more active and interventionist than is suggested by its colloquial clinical usage. Control can include such mundane things as ensuring sterile equipment in a chemistry lab, to restrain the free flow of germs and unwanted particles that might contaminate some test. When controls are applied to human behavior, we are reminded by this definition that someone's behavior is being restrained to be something other than it would otherwise be if the person were free to act. Thus, we are immediately on alert to be sensitive, when studying responses from a controlled experiment, to the possibility that behavior is unusual in some respect. The reason is that the very control that defines the experiment may be putting the subject on an artificial margin. Even if behavior on that margin is not different than it would otherwise be without the control, there is the possibility that constraints on one margin may induce effects on behavior on unconstrained margins. This point is exactly the same as the one made in the "theory of the second best" in public policy. If there is some immutable constraint on one of the margins defining an optimum, it does not automatically follow that removing a constraint on another margin will move the system closer to the optimum.

[4]See Cummings, Harrison, and Rutström (1995); Cummings, Ellot, Harrison, and Murphy (1997); and Harrison (2006a, 2014c) for evidence on hypothetical bias across a range of domains and elicitation methods. There remains a limited role for hypothetical surveys, to explore tentative hypotheses quickly and cheaply, and as statistical complements to incentivized responses that can be used to calibrate hypothetical responses and correct for biases. See Blackburn, Harrison, and Rutström (1994) and Harrison (2006b).

which probability optimism or pessimism can augment, positively or negatively, any risk premium due to an 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 risk.

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.[5] An alternative approach from psychology is to view the perception of time horizon as subjective: if the agent perceives time units contracting as the horizon gets longer, declining discount rates will arise.

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 temporally risk neutral to risk resolved at a point in time but *must* be intertemporally risk averse to risk resolved over time.[6] 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, Harrison, Lau, and Rutström (2018) 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 (2011) for an exposition.

[5]So stated, the Quasi-Hyperbolic assumption is that these fixed costs are some constant fraction of the loan principal, which has the implication that it stays important for all magnitudes of the loan. An alternative is to assume some fixed cost in terms of money, which of course declines in importance as loan magnitudes increase.

[6]This is a different matter than the agent having preferences over *when* risk is resolved.

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.[7]

There is considerable evidence that laboratory and field samples, at least in developed countries, are well characterized as consisting of roughly 50% of subjects 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.[8] 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 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. As we will see, this concern has 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

[7]A persistent "cottage industry" in experimental economics pursues the issue of stability of risk preferences across elicitation methods. Unfortunately, this literature rarely includes the Unordered Binary Choice method. More seriously, it invariably evaluates results by looking for empirical correlations between observable choices. There is no coherent statistical theory that predicts that linear correlations of observed choices reflect similar latent risk preferences, which are invariably nonlinear functions of those observed choices (the nonlinearity deriving from $U'' \neq 0$ and/or probability weighting). The correct way to compare elicitation methods, as explained by Harrison and Rutström (2008, §2.5), is to determine if the implied, latent risk preferences differ.

[8]In other words, if a subject is given a \$100 house endowment and a 50:50 chance of losing \$10 from it or gaining \$11, local asset integration occurs when the subject views this as a 50:50 chance of gaining \$90 or \$111. Using \$100 as the putative reference point for exposition, CPT requires that the argument of the utility function be −\$10 and +\$11, not +\$90 and +\$111. In this sense, CPT does not nest EUT or RDU. If the subject does not perceive different signs on outcomes, CPT is conceptually "dead on arrival" before one starts the estimation engines.

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.[9] Evidence from university undergraduates in the United States shows that the premise is simply false (Harrison, Lau, Ross, & Swarthout, 2017), although evidence from representatives of the adult Danish population shows that the premise is valid for the range of lab wealth considered (Andersen, Cox, Harrison, Lau, Rutström, & Sadiraj, 2018). 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 & Sadiraj, 2006), and these assumptions appear to apply to the Danish population (Andersen, Cox et al., 2018).

There is far 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 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, Harrison, Lau, & Rutström, 2008). Variations in designs allow one to test Exponential discounting against all major alternatives, and Exponential discounting clearly characterizes the data best in such settings (e.g., Andersen, Harrison, Lau, &, Rutström, 2014).[10] Nor is there any evidence for the alleged "magnitude effect," whereby elicited discount rates appeared to be lower for higher stakes (Andersen, Harrison, Lau, & Rutström, 2013).[11]

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, Harrison, et al. (2018) for evidence from the Danish

[9]Give each subject choices between a safe lottery of w for sure, and a risky lottery of a 50:50 chance of $w-x$ or $w+y$, where $w-x \geq 0$ and $y>x$. The key idea is to vary w in the lab, and ask each subject to make lottery choice decisions at different levels of w. Consider, for illustration, values of wealth w from the ordered set, $S=\{w_1, w_2, w_3, \ldots\}$, where $w_1<w_2<w_3$. These values of lab wealth may be plausibly much less than the W that the subject has in the field prior to the experiment. The experimenter does not need to know W for a given subject, but by varying "lab wealth" from S for that subject the experimenter has considered small-stakes lottery choices over 50–50 probabilities of a low prize of $w-x$ and a high prize of $w+y$ against "lab wealth" w for sure, or "field + lab" wealth levels $W+w$, with w from S, for that subject. This step of the design presumes that we vary lab wealth for a given subject since then we can plausibly presume that field wealth W is constant for that subject during the experimental session. If we observe the agent choosing the safe lottery for small levels of lab wealth but the risky lottery for larger levels of lab wealth, then the empirical premise of the calibration critique is rejected for that agent.

[10]One controversy surrounds the use of monetary prizes or "consumption flows" as outcomes. This controversy arises from the dogma that the only true argument of a utility function is consumption, which in turn derives from nothing more than this being what someone was taught in a class on the principles of economics. Augenblick, Niederle, & Sprenger, 2015) argue that using "real effort" as a proxy for consumption flows generates hyperbolicky discounting. There are some difficulties with their design, which relies on the design of Andreoni and Sprenger (2012a). Fundamental conceptual and econometric issues arise with that design when (many) subjects choose corner solutions. Moreover, evidence that utility functions are proximately linear over the stakes considered flies in the face of decades of research on risk aversion (reviewed in Harrison & Rutström 2008, §3.9).

[11]Indeed, this is a prime example of how sloppy the behavioral literature is in referencing "stylized facts." There is literally no study in the entire literature that used incentives and obviously-confound-free designs that finds a magnitude effect, and yet it is commonly asserted as a robust finding.

population.[12] Explicit recognition of the temporal nature of the insurance contract is a major insight, with many potential behavioral implications.

Finally, 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, Fountain, Harrison, & Rutström, 2014). Most losses in insurance are, of course, well characterized as binary events. Another strand tackles the more challenging problem of inferring whole subjective belief *distributions* for continuous or nonbinary events (Harrison, Martínez-Correa, Swarthout, & Ulm, 2017). 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, of course, 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 match actuarial claim rates, which is obviously tenuous. We return to this identification gap later in the literature review.

4 | DESCRIPTIVE APPLICATIONS

4.1 | Laboratory experiments with indemnity insurance

McClelland, Schulze, and Coursey (1993) conducted laboratory experiments with real payments to see if insurance behavior is fundamentally different for low-probability events than for high-probability events. Their first study involved manipulating the probability of loss from very high (0.9) to very low (0.01), while keeping the size of the monetary loss fixed at $4. They used Vickrey auctions, where eight subjects at a time bid for insurance against the loss scenarios, and the top four bidders receive the insurance at the cost of the fifth highest bid. The loss was determined by drawing a chip from a bag, and the result of that event applied to all subjects. Average bids for insurance converged at the expected value of the insurance for most probabilities of loss. However, the results from the decisions with very low probabilities show bimodal behavior from the subjects: they either buy zero insurance or they bid much higher than the expected value. Their second study shows that this finding of bimodal demand continues to hold for a larger loss amount, or even as subjects gain experience.

Risk preferences were not taken into account at all, and McClelland et al. (1993, p. 110) note that risk preferences could possibly help explain their results:

[12]The implication for the claim by Andreoni and Sprenger (2012b) that "risk preferences are not time preferences" is immediate. If the intertemporal utility function that subjects use is actually nonadditive, then risk preferences over time streams of money need to be sharply distinguished from risk preferences over a-temporal payoffs. In effect, there are two possible types of risk aversion when one considers risky choices over time, not one. If one gives subjects choices over differently time-dated payoffs, which is what Andreoni and Sprenger (2012b) did, one sets up exactly the thought experiment that defines intertemporal risk aversion. They compare behavior when subjects make choices over time-dated payoffs that are not stochastic with choices over time-dated payoffs that are stochastic, and observe different behavior. In the former case virtually all choices in their portfolios were at extreme allocations, either all payoffs sooner or all payoffs later; in the latter case they observed more choices in which subjects picked an interior mix of sooner and later payoffs, diversifying intertemporally. Evidence that subjects behave differently, when there is an opportunity for intertemporal risk aversion to affect their choices compared to a setting in which it has no role, is evidence of intertemporal risk aversion. It is not necessarily evidence for the claim that there is a "different utility function" at work when considering stochastic and nonstochastic choices. We do not rule the latter hypothesis out, but there is a simpler explanation well within received theory. Intertemporal risk aversion provides an immediate explanation for the observed behavior in Andreoni and Sprenger (2012b). Just as a-temporal risk aversion encourages mean-preserving reductions in the variability of a-temporal payoffs, intertemporal risk aversion encourages mean-preserving reductions in the variability of the time stream of payoffs. Hence, when Andreoni and Sprenger (2012b) claim that "risk preferences are not time preferences," one can restate this correctly as "a-temporal risk aversion is not the same as intertemporal risk aversion," and of course that is true whenever there is a nonadditive intertemporal utility function.

> *Thus, at least for low probabilities, another theory such as [EUT] or [Prospect Theory] must be employed to explain the apparent oversensitivity to small probabilities observed in our experiments.*

Since risk preferences were not considered, even for the purpose of calculating "expected sensitivity" to parameters, their conclusion applies for all observed decisions and not just those with low probabilities

Irwin, McClelland, and Schulze (1992) explore the effects of hypothetical *versus* real money and experience on insurance purchasing behavior. They make use of the same Vickrey auction as in McClellend et al. (1993), but set the number of draws at 50 or 150 for each subject with a fixed loss probability of 0.01 for all draws. Following McClelland et al. (1993), they propose the expected value of the lottery as the optimal cost of insurance and do not take into account the risk preference of the individual. Their results show that the bimodal result from McClelland et al. (1993) was less pronounced if hypothetical rewards were used, as there was an increased number of very low and very high bids, and that there is some effect of having more than one round in the experiment.

Ganderton, Brookshire, McKee, Stewart, and Thurston (2000) disagree with the empirical findings of McClelland et al. (1993), and do not observe the bimodal distribution of bids for very low probability losses. They attribute the difference in results to differences in their experimental set-up. They employ a more complex decision setting to reflect naturally occurring disasters, and extract insurance choices from subjects in an extensive form game. Subjects face compound lotteries: each subject is first exposed to three possible outcomes (no event, a low probability event, and a very low probability event), then if a loss event has occurred each subject could experience either a small loss or a large loss. A subject could randomly face any treatment from 18 parameter combinations across five insurance cost levels for a random number of rounds and periods.

Ganderton et al. (2000) examined how insurance purchasing behavior would vary for varying insurance costs. They used subject's choices from choices over lotteries with constant mean payoffs but increasing variance to infer risk preferences, and used the method of Cameron (1988) to predict WTP from regression, rather than the implied CE.

As predicted by EUT, the results in their econometric models show that insurance purchase will be less likely when the cost of insurance is high, when the expected loss is low, and when the individual's wealth increases. But their results also show that repeated exposure to loss events results in a negative effect on insurance demand. They also show that subjects are relatively more sensitive to the low probability of a loss, rather than to the size of the potential loss. These results cannot be explained by EUT.

Laury, McInnes, and Swarthout (2009) tested the belief from Kunreuther, Novemsky, and Kahneman (2001) that individuals tend to underinsure against catastrophic events with a low probability and high loss, relative to higher-probability, low-loss events. Their objective was to undertake a systematic study of the effect of the probability of a loss on insurance purchase decisions. Their design focused on whether subjects were more or less likely to purchase insurance as the probability of loss increased, while holding constant the expected value of the loss and the insurance load.

In the first part of their study they replicated the results from Slovic, Fischhoff, Lichtenstein, Corrigan, and Combs (1977), a widely cited laboratory study of insurance purchasing decisions even though all tasks were hypothetical. As much as possible, they replicated the survey that elicited willingness to purchase actuarially fair insurance for up to eight different situations. The probability of loss was presented in terms of draws of orange and white balls from an urn, ranging from 0.0001 to 0.5. The loss amount and insurance price were expressed in points, and payments were hypothetical, even though the subjects were asked to treat the gambles as actual gambles. The loss amount was varied so that the expected loss and insurance price was kept at one level across all decisions. Subjects in their replication were more inclined to purchase insurance than those in the original study for probabilities between 0.0001 and 0.05. The main result from Slovic et al. (1977), however, was still replicated: the percentage of subjects purchasing insurance increases as the probability of a loss increases.

Laury et al. (2009) then conducted a new experiment to test if those results would hold if real money *and* incentive-compatible procedures were used to incentivize the subjects, as is standard in experimental economics. They varied the choices the subjects would make by loss probability (0.01, 0.10), premium load (0.8, 1.0, 4.0), and expected value of loss ($0.15, $0.30, $0.60). The loss probabilities of 1% and 10% were chosen because they could be implemented in a laboratory setting, and so that there was a substantial expected change in proportion of subjects purchasing insurance between the two probabilities (based on results from the previous experiment). Varying the load on the actuarially fair premiums allowed testing of the robustness of results against the premium size.

Taking into account the within-subjects, full-factorial design of the three varying factors, each subject was asked to make a choice for each of the 18 insurance decisions, with an initial endowment of $60 for each decision. The experiment had 40 subjects receiving an actual payment, while 37 subjects did the experiment receiving a hypothetical payment.

Employing the exact conditional McNemar test, a nonparametric procedure, they find that the results in this experiment significantly conflict with Slovic et al. (1977): the earlier finding that more insurance is purchased as the probability of loss increases is *not* observed when real rewards are used. In fact, the results of Laury et al. (2009) show that significantly *less* insurance is purchased as the loss probability increases. They also showed that less insurance was purchased when the payments were hypothetical, but that the same pattern still holds. Premium loading was also found to decrease purchase rates at the 5% significance level.

Laury et al. (2009) have shown that incentives matter for correctly inferring behavior in insurance experiments.[13] When real high-consequence losses were implemented, there was no evidence of underinsurance of low probability losses. This experiment shows that subjects overestimating low-probabilities is not the reason why individuals tend to under-insure against low-probability high-loss events, relative to high-probability loss-loss events, if indeed they do.

Laury and McInnes (2003) considered insurance purchases in which subjects actually received real rewards, but they did not elicit risk attitudes. They comment (p. 228) that the fact that a majority of subjects decided to purchase the actuarially fair insurance is consistent with them being risk averse, and that some levels of risk aversion are in turn consistent with the evidence from virtually every comparable experiment.[14]

Schade, Kunreuther, and Koellinger (2012) consider the purchase of insurance against the loss of a valuable object. They were motivated by deviations from EV in elicited WTP for insurance products noted by previous studies. Their review (p. 534) correctly notes that high levels of risk aversion can explain these extreme choices:

> *There is empirical evidence that many individuals exhibit behavior that implies that they are either unconcerned or extremely risk averse when deciding whether to*

[13]Jaspersen (2016) provides an expansive review of the place of hypothetical surveys and "survey experiments" in behavioral insurance.

[14]In their experiments 74% of choices were to purchase insurance. Three of 60 subjects never purchased insurance, and 17 of 60 always purchased insurance.

> *purchase insurance against events that have a small probability of occurring [...]. The unconcerned individuals are not willing to pay a penny even if premiums are subsidized, whereas those who appear to be highly risk averse opt for premiums that are more than 10 times the expected loss.*

Their experiments have some unfortunate procedural features. First, only 2 of 263 subjects were to be paid in a salient manner, and for the others the only motivation was a small fixed, nonsalient participation payment. Second, subjects were *not* told that probability at the outset (p. 535), and had no way of knowing how many subjects would be in the experiment. Third, they used the Becker-DeGroot-Marschak (BDM) procedure (Becker, DeGroot, & Marschak, 1964) to elicit WTP for insurance.[15] The BDM has been shown to have extremely poor behavioral properties when used to elicit precise CE.[16] They replicated prior findings in a qualitative sense, finding highly skewed distributions of WTP for insurance. They do not report if they observe the bimodality of WTP noted in prior research when one uses real rewards compared to hypothetical survey questions.

Di Mauro and Maffioletti (1996) consider, among other things, a "self-insurance" experiment which is, for our purposes, the same as an insurance purchase.[17] They frame it as self-insurance to contrast with self-protection experiments in which subjects could pay to have the probability of a loss reduced. In any event, each of 38 subjects has a stake of £10 and makes eight choices, four of which are over risky outcomes of interest here. The loss probabilities for the four risky choices are 3%, 20%, 50% and 80%. The subject reports a WTP in each case using a real-time English clock auction: as the price ticks along from £0 to £10, in increments that are not reported, the subject indicates when to "drop out" of the auction. There is considerable evidence from Rutström (1998) and Harstad (2000) that this English auction reliably elicits homegrown values from subjects, certainly by comparison with the theoretically isomorphic Vickrey sealed-bid auction or BDM procedure. Average and median bids are reported (p. 62), along with the standard deviation of bids. There is evidence of slight skewness in WTP, but not as severe as prior studies. There is no evidence presented in either direction about the existence of bimodality of WTP. Median WTP tracks EV closely for all but the highest loss probability, when it is 89% of EV. Average WTP exceeds EV for the 3% and 20% loss probabilities, and is 220% and 128% of EV, respectively; it is less than EV for the 50% and 80% probabilities, and is 87% and 81% of EV, respectively.

Harrison and Ng (2016) consider both descriptive and normative aspects of behavioral insurance. Their lab experiments involve 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.[18] They also estimate parameters 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

[15]The BDM version employed in this study used an attractive, credible method for generating the random purchase prices (p. 536), but is the same as the BDM applied for decades by experimental economists.

[16]Although formally incentive compatible, this elicitation method is widely avoided by experimental economists since subjects often fail to understand it without a great deal of hands-on training: see Plott and Zeiler (2005, p. 537). Moreover, even if subjects understand the incentives, the mechanism is known to generate *extremely* weak incentives for accurate reports: see Harrison (1992, 1994) and Rutström (1998).

[17]Di Mauro and Maffioletti (2001) report exactly the same experimental design and data.

[18]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.

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. In fact, they did not, and the results are consistent with those in Harrison and Ng (2018), discussed below, when repayment percentage there is set at 100%.

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 implicit insurance company. This 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, Martínez-Correa, and Swarthout (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.[19] Focusing solely on their descriptive finding for now, they show that the usual actuarial characteristics, particularly premium levels and loss probabilities, play an important role determining take-up when repayment percentage is less than 100%. 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.

4.2 | Probabilistic insurance

Kahneman and Tversky (1979, p. 269) introduced a concept of "probabilistic insurance," which incorporates the essential features of nonperformance risk when insurance claims are processed. Their hypothetical example also, however, allowed for the probabilistic reimbursement of the premium, which changes predictions from the pure nonperformance risk case.[20]

Herrero, Tomás and Villar (2006) is the first experimental study to examine probabilistic insurance using real rewards applied in an incentive-compatible manner. They examined the original version of probabilistic insurance from Kahneman and Tversky (1979), and compared demand for what they refer to as no insurance (NI), full insurance (FI), and probabilistic insurance (PI). They used an elegant design to first elicit the loss probability that made an individual subject indifferent between NI and PI for the same final outcomes, but without ever framing choices in terms of NI, FI, or PI to subjects. For instance, a subject might have a choice between \$0 with probability λ and x_2 with probability $(1-\lambda)$ or x_1 for sure. The x_1 for certain option is interpreted by the experimenters as FI, since it reflects the outcome of a full indemnity insurance contract with no deductible. Assume that the elicited probability for this subject and these outcomes is λ*. The subject was then given a series of binary choices between NI, FI, and PI, defined simply as three-prize lotteries:

[19]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.

[20]Wakker, Thaler, and Tversky (1997) considered the case of pure nonperformance risk, and showed that the same prediction under EUT applied: that a small risk of nonperformance should not lead to a large change in WTP for the product. Kahneman and Tversky (1979), Wakker et al. (1997), and Zimmer, Schade, and Gründl (2009), *inter alia*, report survey responses to suggest that subjects do not behave consistently with EUT, and appear to dislike probabilistic insurance in each of the forms proposed. Hypothetical survey responses are known to be generally unreliable, and the primary focus of these surveys was to question the empirical validity of EUT rather than evaluate the welfare effects of performance risk. Segal (1988) demonstrates that this "puzzle" from the perspective of EUT can be explained easily using RDU or recursive RDU, where the latter is a model of Segal (1990) that allow one to relax ROCL while assuming the CIA. Wakker et al. (1997) also show that the "puzzle" can be resolved by RDU, even when probabilistic insurance is presented in the pure form.

- NI is where 0 is received with probability $\lambda^*/2$, 0 is received with $\lambda^*/2$, and x_1 is received with probability $1-\lambda^*$.
- FI is where x_2 is received with probability $\lambda^*/2$, x_2 is received with probability $\lambda^*/2$, and x_2 is received with probability $1-\lambda^*$.
- PI is where 0 is received with probability $\lambda^*/2$, x_2 is received with probability $\lambda^*/2$, and $(x_1+x_2)/2$ is received with probability $1-\lambda^*$.

Subjects were incentivized with a subtle method that is indeed incentive-compatible, although that might not seem apparent.[21] They find that subjects tend to prefer FI to PI, where the loss probability is the one that this subject revealed to make her indifferent between FI and NI, and that the subjects tend to prefer PI to NI. The preference for FI over PI is inconsistent with EUT, and consistent with RDU; but the preference for PI over NI is consistent with EUT and inconsistent with RDU. Both patterns are consistent with the Regret Theory of Loomes and Sugden (1982, 1987).

Zimmer, Gründl, Schade, and Glenzer (2018) conduct the first *framed* lab experiment to examine probabilistic insurance using real rewards applied in an incentive-compatible manner. They framed the instructions in terms of insurance products, to make it easier for subjects to understand the task. As the exposition of the design of Herrero, Tomás, and Villar (2006) illustrates, unframed experiments can seem very different than insurance decisions, perhaps disconnecting behavioral responses from field counterparts. Their experiments used a design which gave subjects a 1-in-200 chance of being paid: while a probability of 0.005 might satisfy a theorist as constituting a strictly positive probability of reward, it is surely a concern that subjects might have viewed this as "effectively hypothetical" given the low chance of being rewarded. However, if rewarded, the stakes were high: up to €800. Subjects were told that there was a 5% loss probability, and that the loss would be complete, resulting in earnings of €0. Four full indemnity contracts were offered, with nonperformance risks of 0%, 1%, 2%, and 3%.[22]

Subjects were asked to state their maximum WTP for the insurance contract. The BDM elicitation method was used, with a well-known variation in which the random "buying price" is preselected and placed in an envelope. Despite these caveats, the evidence suggests a sharp reduction in the valuation of insurance for small increases in nonperformance risk, generally inconsistent with EUT.

Biener, Landmann, and Santana (2017) conducted the first *artefactual field* experiments on insurance with nonperformance risk. They presented subjects in the Philippines with insurance contracts that had a 10% default risk for an insurance product that had a loss probability of 30% and a premium of 50 Philippine Pesos, and asked if they wanted to purchase the product. Another product had a 0% default risk for the same loss probability and a premium of 60 Philippine Pesos. Their design does not allow a clean identification of the effect of default risk on take-up, since premium was varied as well, but the effect appears to be to *reduce* demand by at *least* 22.3%.[23] The effect of adding default risk is likely larger, since premiums were lowered as well, leading to an understatement of the pure effect of default risk. No evaluation of choices beyond take-up is provided.

[21]One choice was selected at random to be paid. If that choices was a direct binary choice of one lottery over another, the chosen lottery was paid out. If that choice was one in which the loss probability was elicited that made the subject indifferent, the subject would be paired with another subject. If the other subject reported a loss probability that was smaller (larger) than the subject being rewarded, the subject to be rewarded got to play the NI (FI) option. In effect, this is a variant on the BDM elicitation method, since the first subject does not know what probability the paired subject will state, and can only harm himself by stating a probability that is higher or lower than his true indifference probability. In practice, this logic is not obvious to subjects.

[22]The instructions did present the possibility of default with an unusually aggressive flourish, using the (translated) text: "Default risk. 3%, i.e., the insurer pays its valid claims in 97 out of 100 cases, and in the 3 out of 100 cases the insurer does not pay!" The use of exclamation marks is culturally specific, and Germans often use them for imperative sentences conveying simple advice. Given that the subjects were German, this exclamation mark should not be seen as biasing responses against the purchase of contracts with a default risk.

[23]This is the estimate from the correct probit specification for a binary dependent variable (Table C1, p. 50), with numerous controls.

The behavioral evaluation of nonperformance risk in Harrison and Ng (2018) goes beyond the EUT model of risk preferences used in the theory of Doherty and Schlesinger (1990), by allowing for individuals to be characterized by risk preferences that relax the compound independence axiom in the manner characterized by RDU. Indeed, this specification accounts for just under 45% of their sample. The core theorem of the probabilistic insurance thought experiment of Kahneman and Tversky (1979) does not survive generalization to relax the perfect asset integration assumption. That assumption is that agents treat wealth, income and loss amounts as perfect substitutes. In this case the intuition behind the claim that EUT agents always prefer probabilistic insurance stems from the fact that risk has a second-order effect under EUT and agents prefer a sure gain (from a cheaper policy) since it has a first-order effect: see the general case for the familiar optimality-of-deductibles theorem in Gollier and Schlesinger (1996). In the theoretical framework employed in Harrison and Ng (2018, §2), this assumption amounts to defining $U(A, \pi, L) = \mathrm{A} - \pi - L$ if insurance is purchased at premium π and loss L occurs out of endowment A. However, a generalization proposed by Cox and Sadiraj (2006) allows these arguments of the utility function to be less than perfect substitutes.[24] It is easy to show in this general case that the EUT agent does *not* always prefer probabilistic insurance to a traditional nonprobabilistic full indemnity contract.[25] Hence it would be valuable to consider the welfare evaluation of insurance contracts with nonperformance risk when these generalizations are allowed.

Clarke (2016) raises two empirical puzzles with regards to index insurance[26] demand. The first is that the demand for weather index insurance, which is expected to offer protection against extreme adverse weather events, is lower than expected. The second is that demand seems to be particularly low from the most risk averse, when they are the ones who should benefit most from insurance. He makes use of a rational demand model due to Doherty and Schlesinger (1990) on nonperformance risk to derive a theory to solve these puzzles, so he assumes the consumer is a price-taking, risk averse, expected utility maximizer.

The critical feature of this model with basis risk is the nature of the joint probability structure of the index insurance product and the consumer's loss. Since the payout from insurance is imperfectly correlated with the individual's loss, purchasing index insurance both *worsens* the *worse* possible outcome and *improves* the *best* possible outcome. Although purchasing more index insurance could reduce the loss exposure of the individual when the individual outcome matches the outcome of the index, it will also increase exposure to a worse possible outcome when the individual experiences a loss but the index does not. Depending on which factor has a stronger impact, it is no longer obvious what the optimal amount of insurance a risk inverse individual should purchase. It is no surprise that a sufficiently risk averse EUT agents might find it unattractive.

Solving for the optimal amount of coverage the individual should purchase to maximize expected utility, Clarke (2016) finds that for the classes of constant absolute and constant relative

[24]Some claim that EUT requires perfect asset integration, but this is not true. In contrast, whether or not EUT does require this assumption is irrelevant for present purposes.

[25]For simplicity we consider the original probabilistic insurance contract proposed by Kahneman and Tversky (1979), in which the premium is returned with some probability. Consider the functional form used by Andersen, Cox, et al. (2018), in which $v(A, \pi, L)$ is a constant elasticity of substitution function, and $U(v)$ is the usual CRRA function $U(v) = v^{(1-r)}/(1-r)$ over the composite good. This specification allows perfect asset integration, null asset integration, and partial asset integration as special cases. Andersen, Harrison, et al. (2018) show that the evidence for adult Danes supports the partial asset integration case. And the only case in which the probabilistic insurance contract dominates, with partial asset integration, is when A and L are perfect substitutes.

[26]Index insurance is a popular product in developing countries. Rather than evaluate claims at the level of the insured, claims are resolved by evaluation of some index common to many agents. An example might be a device measuring rainfall in a region. This generates a symmetric basis risk: on a "really good day" for the agent, the index shows a loss but the agent has not incurred a loss, or on a "really bad day" for the agent, the index shows no loss but the agent has incurred a loss. Nonperformance risk is a one-sided, downside basis risk.

risk aversion, demand for actuarially unfair indexed cover is hump-shaped in the degree of risk aversion. First it increases is risk aversion increases, then it decreases at higher levels of risk aversion. Demand for actuarially favorable indexed cover is either decreasing or decreasing-increasing-decreasing in risk aversion, and there is no monotonic relationship between demand and initial wealth, loss amount or premium loading.

Clarke and Kalani (2012) empirically test the results from Clarke (2016) by conducting a field experiment in villages in Ethiopia. They set up lottery choices in the gain frame which they call the benchmark, as well as insurance choices which they try to frame as losses, to test determinants for demand of index insurance, determinants of risk aversion, and effect of group insurance over individual insurance. Clarke and Kalani (2012) use the ordered lottery selection design of Binswanger (1980) to elicit risk preferences, and applied it in their benchmark treatment, as well as in four insurance treatments. Subjects were given 65 Birr, and were told they could lose up to 50 Birr, then they were asked how much insurance they would prefer to purchase. We describe the two of their insurance treatments.

The first treatment involves an individual indemnity contract. Subjects are shown that there are four tokens in a bag, three blue and one yellow. If a yellow token is drawn, subjects will lose 50 Birr. Subjects can choose to purchase between 0 and 5 units of indemnity insurance to reduce the loss amount. One unit of indemnity insurance costs a premium of 8 Birr and with each unit of insurance purchased the loss when a yellow token is drawn is reduced by 10 Birr.

The second treatment is the individual index treatment. This insurance decision is based on a two-stage probability structure. In the first stage, a fair wheel is spun to select between a blue bag, and a yellow bag. The blue bag contains three blue tokens and one yellow token, and a yellow bag contains one blue token and three yellow tokens. A token is drawn from the bag selected in the first stage, and if a yellow token is drawn, the subject will lose 50 Birr. Once again subjects can choose to purchase between 0 and 5 units of insurance, but for this treatment the insurance will only pay out if the yellow bag is selected in the first stage. One unit of index insurance cost a premium of 3 Birr and led to a claim payment of 5 Birr in the event of the yellow bag being selected, and zero otherwise. There is basis risk, hence there is a chance that a subject who purchased insurance might incur a loss but not receive a payout.

Clarke and Kalani (2012) use structural maximum likelihood to estimate risk preferences based on the choices made in the individual indemnity treatment. They assume a CRRA utility function, and that the population on average can have EUT or RDU risk preferences. They use the mean-variance (MV) utility decision theory developed by Giné, Townsend, and Vickrey (2008) to see how well risk choices fit that model. They also tested for how well the risk choices fit a mixture model between MV and RDU risk preferences. They find their data best fits the mixture model of MV and EUT.

Because Clarke and Kalani (2012) used the Binswanger (1980) risk elicitation task, each subject only makes one insurance choice per treatment, hence they are only able to elicit average risk preferences for the sample population, and are unable to elicit risk preferences on the individual level. They also notice framing effects in their study. Although the benchmark and individual indemnity treatment are made up of numerically identical choices, they do not produce numerically consistent choices.

They run an ordered probit model on the choices from the individual index treatment to determine how characteristics impact demand for index insurance. They find that subjects with intermediate levels of wealth have the highest take-up, with the poorest and richest subjects revealing a low demand for index insurance. This is consistent with the hump-shaped theoretical relationship between index insurance take-up and wealth derived by Clarke (2016) in an EUT framework. However, they do not use the risk preferences estimated from the benchmark or indemnity insurance to calculate WTP for insurance. In other words they do not compare how risk aversion should or could affect take-up. However, they do allude to this comparison in Clarke and Kalani (2012, p. 30):

> *This finding [of an "S-shaped" probability weighting function] is not surprising given the data; a large number of participants purchased more index insurance than is consistent with EUT or RDU with an inverse S-shape.*

4.3 | Naturally occurring data

Several studies of insurance data have attempted to estimate large-stakes risk aversion, and evaluate implications of the Hansson-Rabin calibration puzzle. The problem with naturally occurring data, of course, is identification. This is where the trade-off between controlled lab or field experiments and naturally occurring data is most clearly seen.

Cohen and Einav (2007) examine a rich data set 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.[27] 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 (Hansen, Jacobsen, & Lau, 2016).

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, so risk preferences do not play a role in this decision under EUT. But it is easy to imagine an RDU agent viewing the actual claims rate "optimistically" enough to justify these deductibles.[28] Again, nothing in these data allow one to identify the parameters of the simplest RDU model, hence identify the calibration implications for such a specification.

Barseghyan et al. (2013) 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 recognize that what they call probability weighting might also be simply subjective risk perceptions that differ from the true

[27]Cohen and Einav (2007) explicitly "take a neutral position" (p. 746) with respect to the calibration implications of their analysis, recognizing that "avoiding this debate is also a drawback" (p. 747) of their approach. Of course, their analysis was not intended to contribute to the debate over the calibration critique.

[28]For example, the modal choice from the sample was to pay $100 to get a $500 reduction in the deductible. The actual claims rate was 0.043 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 0.21, 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 0.043, requiring even less optimism to justify the purchase. The estimated probability weighting function of Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2013, figure 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 0.043. Of course, this is still a violation of EUT, which is the general point being made by Sydnor (2010).

claims rate, an important issue we return to later. 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 semiparametric 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, figure 4, p. 283): claims rates of zero imply weighted claims rates of 6.5%, with 95% confidence intervals spanning 6% and 10% (figure 1). They also estimate CRRA coefficients of 0.37 and 0.21 (p. 2524).

When it comes to implications for the calibration critique, Barseghyan et al. (2013, p. 2527) hedge, suggesting that their relatively low estimate of U'' "suggests that it may be possible" to explain low-stakes and high-stakes risk aversion while maintaining "standard risk aversion," by which they mean some degree of diminishing marginal utility. If one interprets their probability weighting in terms of an RDU model, they still require a deviation from EUT. In contrast, they openly acknowledge that their analysis "does not enable us to say whether households are engaging in probability weighting or whether their subjective beliefs about risk simply do not correspond to the objective probabilities." (p. 2527). The latter explanation, when it requires additivity, is just SEU, which does not require that subjective beliefs be correct or even updated according to Bayes Rule.[29] We return to the role of subjective beliefs below.

It is possible to write down non-EUT models that can explain small-stakes risk aversion as well as large-stakes risk aversion. For instance, Ang, Bekaert and Liu (2005), building on Epstein and Zin (1990), show that a recursive utility specification with a non-EU, first-order[30] risk averse CE, can account for both types of risk aversion. The theoretical approach in Andersen, Cox, et al. (2018) does not *require* than one adopt a non-EU specification, but does *allow* for that.

An important feature of Barberis et al. (2006) is the evaluation of small-stakes risks that are delayed, rather than resolved immediately. This requirement differentiates their specification from the model of Ang et al. (2005), who implicitly require these risks to be resolved immediately. Modeling risk over time raises many new issues, discussed by Andersen, Harrison, et al. (2018).

5 | NORMATIVE APPLICATIONS

Consider the humble question of the welfare valuation of some new insurance product, such as the "microinsurance" products being offered and promoted in developing countries. In general these policies currently are evaluated by the metric of product take-up. Although take-up is easy to measure, it does not automatically reflect the existence or size of the welfare gain of the insurance product to the insured. An insurance product usually involves the

[29]Some economists view Bayes Rule as a part of SEU, but they are distinct as noted earlier. The literature in behavioral finance is clear about these two being separate, even if it challenges the descriptive validity of both. Barberis and Thaler (2005, p. 1) open their survey by noting that "The traditional finance paradigm [...] seeks to understand financial markets using models in which agents are "rational." Rationality means two things. First, when they receive new information, agents update their beliefs correctly, in the manner described by Bayes's law. Second, given their beliefs, agents make choices that are normatively acceptable, in the sense that they are consistent with Savage's notion of Subjective Expected Utility (SEU)."

[30]First-order risk aversion refers to a utility functional that can exhibit risk aversion for small prizes. Under full asset integration, and assuming wealth is significant, a differentiable utility function does not exhibit first-order risk aversion, though it can at nondifferentiable points (Segal and Spivak, 1990). Under no asset integration it does. In context, the reference in the text is to a "disappointment aversion" model.

individual[31] giving up a certain amount of money ex ante some event in the expectation of being given some money in the future if something unfortunate occurs, as noted at the outset. Welfare evaluation therefore generally requires that one knows risk and time preferences of the individual, since the benefits of the product are risky, and in the future, while the costs are normally[32] certain and up front. We must also know the subjective beliefs that the individual used to evaluate possible losses.[33]

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

Harrison and Ng (2016) provided the first explicit welfare analysis of the demand for insurance. 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 consumer surplus of purchasing insurance or not purchasing insurance, in each case using the CE difference between the two actions. They 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. A key feature of their conceptual design, and the value of undertaking these choices in a laboratory environment, is the ability to observe those who did not purchase insurance as well as those that did. Often the empirical evaluation in the field is limited to those that did purchase insurance, which of course leaves out a significant potential source of welfare gains or losses from those that did not purchase insurance. Their central message is that take-up is not a reliable indicator of welfare, consistent with the rhetorical behavioral concerns with take-up and naïve revealed preferences noted above.

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 our 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. But 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.

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

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

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

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 (§5) a general philosophical exposition of what they characterize as the "quantitative intentional stance" toward behavioral welfare economics.

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.[34] Again, take-up is an unreliable proxy for welfare.

Here we elaborate on how the existing literature has defined welfare gains from insurance, by comparison, and how it has been measured. We broadly categorize the various methods of calculating welfare gain from insurance into four groups: take-up of insurance, WTP for insurance, risk reduction proxies, and "some other metric." Table 1 lists each study, and several salient characteristics of each. We only cover the more important studies from each group here in greater detail.

5.1 | Take-up of insurance

Hill and Robles (2011) developed a market for weather securities in southern Ethiopia to replace the more traditional index insurance contract. Their motivation is to develop a risk management product that better meets the heterogeneous needs of rainfall protection for farmers, which can be dependent on crop choice, land quality or production practices. These factors can vary even among farmers within close proximity of each other. This study has used take-up of weather securities as a proxy for protection from uncertain rainfall. They suggest that a high take-up rate of 20% reflects a welfare gain, but do not specify if a lower take-up reflects a smaller welfare gain or if it would reflect a negative welfare gain.

Hill and Robles (2011) conducted an experiment offering six different weather securities: one against severe drought and one against moderate drought, in each of the three main month of the rainy season. Farmers are given an endowment and can choose which securities they would like to purchase if at all. Securities were priced at expected value. Payouts were given in real time, depending on actual rainfall levels, to closer model the field. The same securities were subsequently offered in a pilot program a year later. Weather securities designed in this way can better meet the heterogeneous rainfall risks of the farmers, relative to a standard index insurance contract. The regression results from the experiment and the pilot program are similar. Farmers who grew barley were much more likely to purchase securities later in the season, when barley grows, and less at the beginning of the season. Use of fertilizer did not affect whether a farmer purchased securities, but did affect which securities he was likely to buy. Those who use soil conservation were more inclined to purchase securities at the beginning of the season. Welfare gain in this study was measured as take-up of the weather securities, and Hill and Robles (2011) were interested in the determinants of securities choices. They do however clarify that, though their results have some merit in understanding the

[34]They make the point that welfare is better measured in this setting by what experimental economists call efficiency: the fraction of consumer surplus that the subject *actually* extracts from the experiment divided by the maximum consumer surplus the subject *could have* extracted. This measure is conceptually better than expected consumer surplus in this context, because one is comparing behavior in the face of insurance contracts that have no nonperformance risk (apples) with insurance contracts that have varying levels of nonperformance risk (a variety of non-apple fruits). All subjects faced the same set of insurance choices, with the same nonperformance risk variations.

TABLE 1 Review of alternative welfare metrics for microinsurance products

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
A. Welfare measured by take-up					
Gumber (2001)	Take-up of health insurance and financial protection	Average	Household survey		Usage of private or public health facilities is price-sensitive.
Schneider and Diop (2001)	Take-up of health insurance	Average	Household survey		Low take-up, despite insurance improving financial access to care across all income levels. Social capital is an important determinant for participation.
Giesbert (2008)	Take-up of health insurance	Average	Actual insurance sold and survey		Understanding of concept of insurance beyond health insurance is mixed, though potential demand for insurance in survey area seems to be high.
Giné et al. (2008)	Take-up of rainfall index insurance	Average	Household survey		Lack of understanding, but also credit constraints, limited familiarity, and risk aversion discourage insurance purchase. Being previously insured, connected to village networks and self-indentifying as “progressive” encourage insurance purchase.
Ito and Kono (2010)	Take-up of health insurance	Average	Actual insurance sold and survey	MPL (real)	Finds weak empirical support for a risk-loving attitude toward losses which might explain the low take-up rates. Households with hyperbolic preference were more likely to purchase insurance,

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					consistent with their theoretical prediction of a demand for commitment. Also some evidence on the existence of adverse selection: households with a higher ratio of sick members were more likely to purchase insurance.
Thornton et al. (2010)	Take-up of social security health insurance	Average	Actual insurance sold and survey		Low take-up and retention rates for insurance. Health services utilization did not increase with insurance. Microfinance institutions were not a more effective delivery agent than the government.
Hill and Robles (2011)	Take-up of varying weather securities	Average	Field experiment, actual insurance sold and survey	Choices on components of weather securities package (real)	High take-up in average and variance experimental game and pilot as weather securities are easily understood and fit heterogeneous farmers' needs. Crop and production choices, and soil characteristics have some explanatory power for security choices.
Clarke and Kalani (2012)	Take-up of index insurance, reduction of risk aversion	Average, variance, and MEU	Field experiment	Binswanger (real)	Take-up is hump-shaped against wealth, where subjects with immediate levels of wealth have the highest take-up. There is no strong evidence of schooling,

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					understanding of the decision problems or financial literacy significantly increasing take-up. Background risk however significantly affects take-up. Parametric assumptions matter when estimating determinants of risk aversion.
Hill et al. (2013)	Reduced adverse consequence of shocks on income and consumption	Average	Survey	DBDC contingent valuation method (Hypothetical); Binswanger (one Hypothetical, and one Real)	Those who faced higher rainfall risk, were less risk averse, more educated, more proactive, and richer, were more likely to purchase insurance. Offering insurance through a risk-sharing group increases demand for less educated females, but is constrained by lack of trust among neighbors.
Barberis et al. (2006)	Take-up of health insurance, through bundling with renewed loans	Average	Field survey and admin data		Adverse selection was not detected in take-up of bundled product because there was no demand for the product. Low demand could have been due to consumers' pessimism on how insurance would be implemented.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Cole et al. (2014)	Take-up of rainfall index insurance	Average	Actual insurance sold and survey		Households in villages that have experienced insurance payouts are more likely to purchase in the following season, but this effect decreases over time. Households that have experienced payouts themselves are more likely to purchase two and three seasons later, than the first.
Dercon, Hill, Clarke, Outes-Leon, and Taffesse (2014)	Take-up of rainfall insurance	Average	Actual insurance sold and survey		Insurance demand increased when groups were exposed to training that encouraged sharing of insurance within groups. A suggested reason is that risk-sharing and index insurance can be shown to be complementary.
Norton et al. (2014)	Optimal allocation of endowment between risk management options	Average	Field experiment and actual insurance sold	Allocation of endowment between risk management options (Real)	Participants exhibited clear preferences for insurance contracts with higher frequency payouts and for insurance over other risk management options, including high interest savings. The preference for higher frequency payouts is mirrored in commercial sales of the product, with commercial purchasers paying substantially higher premiums than the minimal, low frequency

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					option available. Commercial insurance also has the option for premiums to be paid through labor.
Jin, Wang, and Wang (2016)	Take-up of weather-based index insurance	Average	Household survey and field experiment	MPL (real)	More than half of the farmers surveyed purchased the weather index insurance. Their main stated reasons were the support and subsidy from the government, and the belief that the probability of future crop losses due to weather events is high. The main reasons for not participating are farmer's low income, low trust in local insurers, and lack of understanding of the policy. The average farmer is moderately risk averse, and risk aversion has a positive effect on farmer's weather index insurance participation decisions.
Jensen, Mude, and Barrett (2018)	Demand for IBLI	Average	Actual insurance sold and survey		Basis risk and spatial adverse selection associated with division average basis risk dampen demand for IBLI. Households in divisions with greater average idiosyncratic risk are much less likely to purchase insurance. There is also strong evidence of intertemporal adverse selection as households

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					purchase less coverage, conditional on purchasing, before seasons for which they expect good conditions.
B. Welfare measured by willingness to pay					
Donfouet, Makaudze, Mahieu, and Malin (2011)	WTP for health insurance	Average	Surveys and field experiment	DBDC (Hypothetical)	Age, religion, usual means of seeking treatment when getting sick, profession, knowledge of insurance, income, and involvement in associations or health policies are key determinants of WTP. There is a demand for this insurance in the studied region.
Schram and Sonnemans (2011)	MEU	MEU	Lab experiment	MPL (real)	A positive effect on behaviour from having more alternatives to choose from (because the best available policy will be at least as good), is compensated by the negative effects of higher search costs, more switching costs and lower decision quality. Switching costs decrease switching; however, switching costs did *not* decrease information

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					retrieval, decision time or decision quality. If health profiles change gradually, subjects remain inactive and do not change to a new policy (the "boiling frog" effect). However, a sudden change in health profile increases the probability of a switch, which increases realized earnings, because subjects think more carefully about their decisions.
Carriquiry and Osgood (2012)	MEU	Average, variance	Theory		If contracts are appropriately designed there are important synergies between forecasts and insurance and effective input use.
Bryan (2018)	MEU	MEU	Field experiment	Binswanger (hypothetical)	Provides a theory that implies ambiguity may decrease the adoption of novel technologies and limit the value of insurance. The effect of ambiguity aversion decreases with experience, a policy of short-term subsidization, and long-term insurance may help to alleviate low demand.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Elabed, Bellemare, Carter, and Guirkinger (2013)	WTP for agricultural index insurance	CE	Field experiment	Binswanger and MPL (real)	Index insurance demand decreases under both risk aversion and compound risk aversion as basis risk increases. Multiscale contracts that depend on triggers at the district and village level may allow for lower triggers to reduce basis risk to the farmer, while avoiding moral hazard problem.
De Janvry, Dequiedt, and Sadoulet (2014)	WTP for insurance against common shocks	Average	Theory		Insurance exacerbates free-riding when covering common shocks. Insurance against a common shock may be unprofitable to an individual if he anticipates others in the group not participating in it.
Gerking, Dickie, and Veronesi (2014)	WTP for reduction of mortality and morbidity risk	Average	Survey		Develops and applies an integrated model of human mortality and morbidity in an EU framework, extended to incorporate a sick state of illness, allows parents to make choices about risk exposure Average and variance for herself and for a child, and a multiperiod framework.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Koufopoulos and Kozhan (2014)	MEU	Average	Theory		Full-insurance pooling equilibrium can exist when accounting for average and variance asymmetric information and ambiguity. An increase in ambiguity may also lead to a strict pareto improvement.
Leblois, Quirion, and Sultan (2014)	MEU	CE	Field experiment and actual insurance sold	MPL (Real)	Length of cotton growing cycle is the best performing index considered. Gain from weather-index based insurance is lower than that of hedging against cotton price fluctuations provided by the national cotton company.
Jaramillo, Kempf, and Moizeau (2015)	Risk reduction through informal insurance schemes	CE, variance	Theory		Heterogeneity within groups reduces risk sharing, redistribution schemes could counter social exclusion, norms of reciprocity and social capital are key determinants of insurance arrangements.
Elabed and Carter (2015)	WTP for agricultural index insurance	CE	Field experiment	Binswanger and sMPL (real)	Allowing for compound risk aversion would significantly decrease the expected demand for insurance with a downside basis risk.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
McIntosh et al. (2015)	WTP for probabilistic insurance	CE	Field experiment	Choices to purchase insurance (real)	Average WTP for insurance increases as the loss severity increases, even if the payout is constant, which causes the insurance coverage to be more partial. Average WTP decreases, however, when payouts are more probabilistic, so that the probability the insurance fails to pay for an adverse event increases. Offering insurance on a group level does not increase demand for index insurance.
Clarke (2016)	MEU	MEU	Theory		A model for rational demand for index insurance products is presented which explains two puzzles regarding index insurance demand: why demand for index insurance is lower than expected and why demand is low for more risk averse individuals.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Chantarat, Mude, Barrett, and Turvey (2017)	CE of herd growth rate	CE	Simulation		Household initial herd size is the key determinant of the product's performance, more so than household risk preferences or basis risk exposure. The product works least well for the poorest. The product is most valuable for the vulnerable nonpoor, for whom insurance can stem collapses in herd size following predictable shocks. Demand appears to be highly price elastic, and willingness to pay is, on average, much lower than commercially viable rates.
Kairies-Schwarz, Kokot, Vomhof, and Weßling (2017)	MEU	MEU	Lab experiment	Trade-off method (real)	Majority of subjects are classified as CPT types over EUT types. About 14% of all participants show inconsistent behaviour with regards to their insurance contract choices and their risk preferences, which is interpreted as an indication of low decision quality. However, the majority show consistent behaviour and good decision quality. One proposed explanation for inconsistent choices is the application of heuristics.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
C. Welfare measured by risk reduction proxies					
Townsend (1994)	Smoothness of consumption as a result of risk sharing	MEU	Household survey		By using a general equilibrum framework, the results on consumption and income are mixed for the complete market hypothesis.
Skees, Gober, Varangis, Lester, and Kalavakonda (2001)	Reduced revenue volatility of rainfall insurance	CV of expected revenue	Simulation on past data		A drought insurance program based on rainfall contracts would have reduced relative risk in Morocco.
Chou, Liu, and Hammitt (2003)	Reduced precautionary savings or risk reduction against unexpected health expenditures	Average	Government survey		Households significantly reduced their saving and increased their consumption when the comprehensive health insurance became available, with the largest effects on savings for households with the smallest savings.
Hess (2003)	Allowing risky farmers to maintain access to credit during drought and smooth income	VaR	Simulation on past data		Integrated scheme can help banks reduce their lending volume while bringing down default rates and transaction costs. It can also help farmers stabilize their incomes and possible access to greater credit line from enhanced collateral.
Jowett (2003)	Risk sharing (through informal networks or voluntary health insurance)	Average	Household survey		Individuals in highly cohesive communities are far less likely to purchase public voluntary health insurance.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Vedenov and Barnett (2004)	Efficiency: reducing exposure to yield risk	MSR of loss, VaR and CER	Simulation on past data		Weather derivatives may reduce risk, but complicated combinations of derivatives are needed to achieve reasonable fits (basis risk is not transparent). Results from in-sample do not translate to out-sample data.
Giné et al. (2007)	Reduced exposure to rainfall risk	Variance	Household survey		There are large diversification benefits from holding a portfolio of insurance contracts, even though all insurance payouts are driven by rainfall in the same Indian state.
Breustedt, Bokusheva, and Heidelbach (2008)	Risk reduction on farm level yields (vs. regional level)	MV and SSD	Simulation on past data		Out of weather index, area yield index and farm yield insurance, none provide statistically significant risk reduction for every farm.
Clarke and Dercon (2009)	Vulnerability to poverty	Variance	Review		Insurance (average and variance formal and informal), credit and safety nets can work together to reduce poverty.
Giné and Yang (2009)	Take-up of loan to adopt new technology	Average	Actual insurance sold		Packaging rainfall insurance with loan to purchase high-yielding seed decreases take-up of loan for Maize and groundnut farmers in Malawi. This could be due to implicit insurance from limited liability in loan contract.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Hill and Viceisza (2012)	Take-up of fertilizer (input)	Average	Actual insurance sold		Presence of (mandated) insurance increases take-up of fertilizer. Take-up also depends on initial wealth and previous weather realizations that affect subjective beliefs of weather outcomes.
Carter and Janzen (2012)	Less costly risk management behavior	Average	Actual insurance sold and survey		Insured households anticipate making cash flow choices which will increase welfare over uninsured households anticipated cash flow choices. These decisions include maintaining consumption levels, and less reliance on assistance.
Cole et al. (2013)	Improved risk sharing of weather shocks—which should affect income variability	Average	Actual insurance sold and survey	Binswanger (real)	Insurance demand is significantly price sensitive, with an elasticity of around unity. There is evidence that limited trust and understanding of the product, product salience, and liquidity constraints also limit insurance take-up and demand.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Mobarak and Rosenzweig (2013)	Take-up of risky technologies and wage risk reduction for landless population	Average	Actual insurance sold and survey		As basis risk increases, index insurance take-up increases if there is also informal risk sharing. Although informal risk sharing in caste groups reduces the sensitivity of profit and output to rainfall, relative to index insurance, it also reduces average returns. Landless households are more likely to purchase index insurance if cultivators are also offered weather insurance.
McIntosh, Sarris, and Papadopolous (2013)	Take-up of fertilizer input	Average	Actual insurance sold and survey		Farmers in Ethiopia are subject to credit constraints that limit their fertilizer use, which is also sensitive to risk-related variables. Actual weather index insurance take-up is not correlated with hypothetically stated WTP, and is sensitive to vouchers for insurance purchase.
De Brauw and Eozenou (2014)	EU from sweet potatoes yield	Average	Field experiment	MPL (hypothetical)	Farmers' preferences better follow the more flexible power risk aversion preferences over CRRA, and RDU over EUT. Assuming CRRA would poorly predict risk preferences among those who are less risk averse for their sample.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Karlan et al. (2014)	Increase in investments in risky input	Average	Actual insurance sold and survey		Uninsured risk is a binding constraint on farmer ex ante investment, but the liquidity constraints are not as binding as typically thought, meaning that credit markets alone are not sufficient to generate higher farm investments. They also find that there is sufficient demand for rainfall insurance, but factors such as basis risk, trust in the insurance company, and farmer's recent experience affected their demand for insurance.
Lin, Liu, and Meng (2014)	Overall increase in risk coverage (crowding out effect)	Average	Lab experiment		Formal partial insurance significantly crowds out public transfers, but without significant decline in risk coverage. Average altruistic preferences and fixed income inequalities contribute to the crowding-out effect
Cai et al. (2015)	Number of sows raised (input)	Average	Actual insurance sold and government data		Providing access to formal insurance significantly increases farmers' tendency to raise sows. These short-run effects seem to have some persistence in the longer run. This increase is not in response to a substitution of other livestock.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
					Lack of trust for government-sponsored insurance products acts a significant barrier for farmers' willingness to participate in the insurance program.
Jensen, Barrett, and Mude (2016)	Reduction of basis risk of IBLI	Stochastic dominance, mean-variance, OLS, semivariance	Actual insurance sold and survey		Covariate risk is spatially sensitive to the covariate region, resulting in spatial adverse selection. Basis risk, mainly idiosyncratic risk, is substantial, so insurance reduces risk but offers partial coverage.
Chantarat et al. (2017)	Reduction of livestock mortality risk	Average	Survey and household data		By addressing serious problems of covariate risk, asymmetric information, and high transactions costs that have precluded the emergence of commercial insurance in these areas to date, IBLI offers a novel opportunity to use financial risk transfer mechanisms to address a key driver of persistent poverty.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Zhang, Tan, and Weng (2018)	Reducing basis risk	MEU	Theory and simulation on past data		Studies the problem of optimal index insurance design under an EU maximization framework. They show that the indemnity as a highly nonlinear and even nonmonotonic function of the index variable can significantly outperform linear-type index insurance contracts in reducing basis risk.
D. Welfare measured by other metric					
Bone, Hey, and Suckling (2004)	*Ex ante* efficient risk sharing	Average	Lab experiment	Binswanger (real)	When sharing a risky financial prospect, the results indicate that fairness is not a significant consideration, but rather that having to choose between prospects diverts partners from allocating the chosen prospect efficiently.
Wagstaff and Pradhan (2005)	Improvement of health outcomes and expansion of household consumption	Average	Survey		The program led to increased use of healthcare, reduction in out-of-pocket health expenditures and increase in nonmedical household consumption.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Franco et al. (2008)	Impact of community-based MHO intervention on health utilization and financial protection	Average and variance	Survey		Members were more likely to seek treatment, make prenatal visits and use insecticide-treated nets. Distance significantly affects utilization, but enrollment not significantly associated with socioeconomic status, except for highest quintile. Members also have lower percentage of expenditures on health, lower out-of-pocket payments, and lower mean-to-median expenditures.
Rao et al. (2009)	Performance of CHF	Average	Actual insurance sold and survey		Members had significantly higher utilization of healthcare services, but no evidence of reduced out-of-pocket spending. The main reasons for not enrolling were being unaware of the program, high premiums, and perceived low quality of services.
Charness and Genicot (2009)	Optimal equilibrium from risk sharing	Average	Lab experiment	Binswanger (real)	Risk sharing exists, even without commitment, and depends on continuation probability of experiment, level of risk aversion of subject, reciprocity, prior expectations of subject, and relative initial income.

(Continues)

TABLE 1 (Continued)

Study	Metric of welfare	Measure	Data	Elicitation method for experiments (hypothetical or real)	Result
Aggarwal (2010)	Increase in health utilization and financial protection through health insurance program	Average	Survey		Utilization of outpatient care and surgeries was greater in the insured group. Borrowings, and payments made from savings, incomes and other sources reduced.
Hamid, Roberts, and Mosley (2011)	Impact of addition of health insurance to microcredit on poverty indicators	Average	Actual insurance sold and survey		Adding MHI to microcredit has a significant beneficial effect only on food sufficiency, which could possibly be due to the short time frame of the study.
Chou et al. (2014)	Improved infant health and postneonatal mortality rate	Average	Government data		NHI in Taiwan in 1995 led to reductions in the postneonatal mortality rate of infants born in farm households (previously uninsured, less education, low-weight births) but not to infants born in private sector households.

Abbreviations: CE, certainty equivalence; CER, certainty equivalent of revenue; CHF, community health fund; CPT, Cumulative Prospect Theory; CRRA, constant relative risk aversion; CV, coefficient of variation; DBDC, double-bounded dichotomous choice; EUT, Expected Utility Theory; IBLI, index-based livestock insurance; MEU, maximising expected utility; MFI; microfinance institution; MHO, mutual health organizations; MPL, multiple price list; MSR, mean root square; MV, mean variance; NHI, National Health Insurance; OLS, ordinary least squares; RDU, rank-dependent utility; sMPL, switching multiple price list; SSD, second-degree stochastic dominance; VaR, value-at-risk; WTP, willingness to pay benefits from weather securities, purchasing securities does not equate to purchasing securities that correctly hedge risk.

Hill, Hoddinott, and Kumar (2013) used survey questions on take-up to measure WTP for weather-index insurance among Ethiopian households. Their survey asks questions to see how characteristics such as risk and time preferences, initial wealth, ability to borrow money, and price of insurance affect whether or not the household would purchase a hypothetical insurance product. The survey uses a methodology from Binswanger (1980) to elicit risk preferences. However, rather than assuming a parametric form for utility to calculate the risk coefficient that corresponds with the subject's lottery choice, they use the direct relationship between the subject's preferred choice and take-up to draw their conclusions of how risk aversion affects take-up. The impact from basis risk is measured by using distance from the closest weather station as a proxy for basis risk.

They claim that using a probit model allows them to calculate the change in WTP brought about by each determinant of demand, and indeod to generate an estimated WTP *for each individual*. All that they mean by this, however, is that the average coefficients estimated across the population can be applied to the individual's specific characteristics to estimate the impact these *characteristics* have on WTP. For instance, their data shows increasing the distance from the nearest weather station from 5 to 15 km reduces the demand for insurance by 8.6 percentage points, which corresponds to a reduction in WTP of 10.75 Birr. Using a probit model, the relationship between level of risk aversion and insurance demand is limited to a linear or quadratic relationship.

Hill et al. (2013) set out to examine how individual household characteristics impact weather index insurance demand, and their study shows that educated, rich and proactive individuals are more willing to purchase insurance. However, through these results, they imply that an increase in insurance take-up reflects a welfare gain for the household. In their introduction they explain that the welfare gain from insurance is from the reduction in adverse consequences from shocks, which include the loss of livelihood through loss of assets, slower income growth, reduced investment in human capital, and discouragement against risky actions which could potentially lead to higher yields. They do not measure how households benefit from insurance in these ways. Since they use a probit model, they can only tell if a certain characteristic, risk aversion for example, impacts insurance demand on average for the entire sample. They cannot determine if the insurance product would benefit a specific individual based on his specific risk preferences, and how much that benefit is.

Cole, Stein, and Tobacman (2014) study the long-term impact of payouts of insurance claims on future take-up of index insurance. Their data are based on a rainfall insurance product sold by an NGO called SEWA in Gujarat India. They used randomized marketing packages as an exogenous variation in insurance coverage to households. These packages included discounts, targeted marketing messages, and special offers on multiple policy purchases. Using instrumental variable (IV) specifications, they instrument for the lag of number of insurance policies purchased and the amount of payouts received using variables characterizing the lagged marketing packages with lagged insurance payouts. Their results show that an increase in payout by Rs. 1,000 in the *village* as a whole results in a 29% average increase in the probability of purchasing insurance the following year, which is significantly positive. The coefficient of the *individual* payout received in the previous year, though positive, is not statistically significant. As the lag time increases, for 2 and 3 year lags, the estimated effect of the village payout decreases, while the effect of the individual payout increases.

5.2 | WTP for insurance

In their field experiment, **Elabed and Carter (2015)** use WTP for a weather index insurance product to measure welfare benefit of the insurance for cotton farmers in Mali. They take into account risk preferences when measuring welfare. However, they assume that all the farmers evaluate risk using EUT. Their study looks into the impact of compound risk preferences from basis risk on WTP for weather index insurance. They make use of the Smooth Model of Ambiguity Aversion formalized by Klibanoff, Marinacci and Mukerji (2005) to separate preferences on simple risk and on compound risk. The premium for the compound lottery is approximated by the formula derived by Maccheroni et al. (2013), which breaks the premium down into a compound-risk premium and the classical Pratt risk premium, allowing the CE to be derived as the expected value of the lottery less the risk premium. WTP for the index insurance contract is then calculated as the difference between the CE of the index insurance contract and the CE of the simple lottery faced in the autarkic situation.

Their experiment is divided into two tasks, where one of the tasks is randomly selected to actually be played out for real money. The first task presents insurance contracts with no basis risk using a methodology similar to Binswanger (1980), where the menu of insurance options is presented to the subject, and they select their preferred choice. The options are presented to the subjects as blocks of insurance: six discrete yield levels are specified with a probability assigned to each level, and subjects were asked to select how much insurance coverage they wanted such that they would be guaranteed a minimum of that yield level. The probability, revenue and premium for each yield level were determined beforehand and shown to subjects. Premia were set at 20% above the actuarially fair price. The actual yield outcome was then randomly selected based on the probabilities shown to subjects. Assuming CRRA preferences, the subject's CRRA risk parameter was then inferred from the range consistent with the selected insurance contract. This experiment frames the risk parameter elicitation question in the context of insurance. Although the parameters of this experiment were set up to reflect scenarios in the field, with a 50% chance of a highest yield, this does not allow one to reliably identify non-EUT models. Furthermore, the range of CRRA risk parameter that can be captured only spans 0.08 to 0.55. Lastly, with this methodology only one data point, the mid-point of the interval that corresponds to the subject's preference, is used to estimate the risk preferences for each individual subject; hence there is no standard error for that estimate.

The second task presents the subjects with the index insurance contract, where there is a 20% chance the insurance will not pay out even though the subject has a low yield. Only downward basis risk is considered here. Given the price of the index insurance contract, a switching multiple price list (sMPL), following Andersen, Harrison, Lau, and Rutström (2006), was used to elicit the minimum price of the "fail-safe" insurance where the index insurance would start being preferred over the "fail-safe" insurance contract. Such a set-up might frame the questions such that it leads subjects to select a switch-over price in the middle of the prices offered. Only compound risk aversion, and not risk loving, is considered. WTP to avoid basis risk is defined as the difference between the price the subject is willing to pay to avoid switching to index insurance and the market price of the "fail-safe" insurance, which was determined in the previous task as 120% of the actuarially fair premium.

Using the CRRA risk parameter elicited from the first task, and assuming constant compound risk aversion, the compound risk parameter was also estimated, and 57% of subjects were found to be compound risk averse to varying degrees. They use the estimated risk parameter and compound risk parameter to calculate the WTP of index insurance, and demand for the insurance product is defined as whether WTP lies above or below the market price, which is defined as 120% of actuarially fair premium. Taking into consideration compound risk aversion when calculating WTP would reflect a demand that is only slightly over half of the demand estimated when only simple risk aversion parameters are used to calculate WTP.

Elabed and Carter (2015) states that the welfare benefits from insurance are from the expected impact on the improved well-being of households exposed to risk. They implicitly estimate this expected improved well-being by measuring WTP of the individual subjects, and determine that there is a positive welfare gain from purchasing insurance if WTP is greater than 120% of the actuarially fair premium, and a negative welfare gain from purchasing insurance if WTP is below that market price.

5.3 | Risk reduction proxies

Mobarak and Rosenzweig (2013) used a randomized control trial (RCT) to examine the relationships between informal risk sharing, index insurance, and risk-taking behaviors in India. They made use of pre-existing census data, offers of rainfall insurance contracts that provided a cash payment if rainfall was delayed beyond a predetermined date at randomized discounted prices, and knowledge of the extent of informal risk sharing within readily identifiable, exogenously formed networks: the subcaste, or *jati*. *Jatis* were their natural risk-sharing network: the data indicated that the majority of loans and transfers to the households were from family and fellow caste members, but also they were from fellow caste members originating from outside the village. This meant that this informal framework could also indemnify rainfall risk which was on a village-level, as well as household-specific idiosyncratic risk. Another feature of their design is that they randomly placed weather stations in some of the project villages, and proxied basis risk of the household as their distance from these weather stations. This allows them to explore how basis risk affects take-up of index insurance, and how informal risk sharing affects the impact of basis risk on the index insurance take-up.

Using their results Mobarak and Rosenzweig (2013) measure welfare gain in three ways. First, they examined whether and how caste-based risk sharing affects the demand for formal insurance. Second, they compared the effects of index insurance provision and informal risk sharing on farmers' willingness to invest in risky production methods and technologies which could lead to higher yield and profits, which was measured by adaption of these methods and technologies. Third, they assessed the general equilibrium effects of offering insurance to both cultivators of the land as well as to agricultural laborers on wage levels and volatility of the wage levels. This was done by estimating labor supply and labor demand effects.

To answer the first question regarding index insurance demand, **Mobarak and Rosenzweig (2012)** embed a model of index insurance with basis risk in the cooperative risk-sharing model developed by Arnott and Stiglitz (1991). This model predicts that (a) when there is no basis risk, index insurance demand is independent of whether or not there is informal risk sharing, and (b) as basis risk increases, it can decrease index insurance take-up, but having an informal risk sharing network can increase that demand as it can still cover the idiosyncratic loss when the index contract fails. The results from the RCT corroborate those predictions. For the second question regarding welfare gain from willingness to invest in riskier production techniques and new technologies, the modified Arnott-Stiglitz model predicts that higher informal coverage may be associated with less risk taking. The level of risk taken by farmers was proxied for by using sensitivity of their crop yield and profits to rainfall. This was measured by how much their crop yields and profits vary according to rainfall levels, and is based on the assumption that the larger the risk the farmers take, the more their yields and profits are exposed and dependent on rainfall. Once again the results are consistent with the theory. Farmers who depended more on index insurance had profits and yields that were more sensitive to rainfall, relative to farmers who depended more in informal risk sharing. The impact of this welfare gain, although qualitatively clear, could not be quantified.

Lastly, welfare gain was measured as a reduction in wage risk for landless agricultural laborers. Mobarak and Rosenzweig (2013) were able to measure this because they offered the index insurance to landless laborers as well as to cultivators, whereas most index insurance products are only marketed to landowners. The take-up rate of index insurance among the agricultural laborers was similar to the take-up rate of cultivators. A general equilibrium model was used to assess the impact of index insurance on agricultural labor demand and supply. They assume workers work have to work more when rainfall levels are low in order to smooth income, and are able to take more leisure time when rain is plentiful, which would result in higher equilibrium wage rates in the good times and lower equilibrium wage rates when rainfall levels are lower. Regarding supply, number of days of agricultural work completed, for those with index insurance, was much less sensitive to rainfall than those without index insurance. Similarly the probability of temporary migration as an *ex post* means to income smoothing was significantly less sensitive to rainfall for those who purchased insurance. On the demand side, more male harvest labor was hired as rainfall levels increased, although the increase in demand for laborers was much steeper for farmers who were offered insurance. This indicates that when farmers purchase index insurance, their increased risk taking will increase wage levels, but labor demand volatility will also increase, which will increase wage risk. The welfare gain for laborers from purchasing index insurance should therefore increase if they know that the farmers are also purchasing index insurance, and this is reflected in laborer insurance take-up being higher when cultivators are also offered insurance.

DeBrauw and Eozenou (2014) conduct a hypothetical field experiment to measure risk preferences of Mozambican farmers regarding sweet potato production. Although their study does not consider insurance, they consider heterogeneity in risk preferences for farming inputs given uncertain weather conditions, and do not just assume that subjects are all EUT or CRRA. The results and methodology of this study could be applied to designing a weather insurance product that would match their objectives, which is to encourage people in rural Mozambique to grow and consume a more nutritional variety of sweet potato. The risk preference elicitation experiment was modelled after Holt and Laury (2002). Respondents were given a series of 10 scenarios where they had to choose between two varieties of sweet potatoes which, depending on rainfall conditions, would produce different yields. The first variety would produce only average yields that vary less with rainfall, and the second variety would produce much higher yields under good weather conditions, but much lower yields under bad weather conditions. The probability of good rainfall increased across the scenarios from 10% to 100%.

DeBrauw and Eozenou (2014) used the MPL methodology, and could only estimate the average risk preferences of the sample. They are not clear on how risk preferences were estimated. They found that they can strongly reject CRRA preferences in favor of a more flexible utility function they call "power risk aversion" that nests the conventional CRRA utility function. Regardless of utility function, they reject the hypothesis of EUT preferences for the pooled sample, in favor of RDU with S-shaped probability weighting functions where respondents on average underweight small probabilities and overweight larger probabilities. Their study focuses only on estimating the average risk preferences of the sample, and does not use the risk preferences to go one step further to estimate the WTP of insurance that would reduce the exposure of the subjects to risk.

Karlan, Osei, Osei-Akoto, and Udry (2014) assert that the welfare gains from improving financial markets through weather index insurance are threefold. First, uninsured risk and limited access to credit could discourage risky investments that could produce higher yields. Second, weather risk is worth managing, as agriculture in northern Ghana, where the study is conducted, is almost exclusively rain-fed. Third, index insurance can help smooth consumption. Karlan et al. (2014) test the impact of insurance and credit on investment decisions by using a 2×2 treatment of either offering a cash grant or not, and offering insurance at varying prices or not. Using ordinary least square (OLS) they find that uninsured risk is a binding constraint on farmer *ex ante* investment (land investment costs and acres cultivated), but the liquidity constraints are not as binding as typically thought, implying that credit markets alone are not sufficient to generate higher farm investments. They also find that there is sufficient demand for rainfall insurance. At actuarially fair prices, 40–50% of farmers demanded insurance, purchasing coverage for more than 60% of their cultivated acreage. Factors such as basis risk, trust in the insurance company, and farmer's recent experience affected their demand for insurance. Since OLS was used, the methodology can only give the sign and size of the welfare gain for the average of the sample. They are unable to quantify welfare effects, or even tell if there is an expected welfare gain or loss for the individual given the individual's characteristics.

Cai, Chen, Fang, and Zhou (2015) considered welfare gain as an increase in the number of sows produced by pig farmers in Southwest China. Pig farmers have to decide if they raise their female piglets as sows for breeding purposes or if they spay them and raise them for their meat. A high mortality rate of sows (2%) deters farmers from choosing to not spay their female piglets, which leaves pork production numbers lower, and pork prices more sensitive to pork shortages. Cai et al. (2015) examine the effect sow insurance would have on the number of sows bred. The insurance is offered by the government, and pays out a lump sum of 1,000 yuan should the sow die through disease, natural disaster, or accident. To further encourage take-up of sow insurance, the government subsidized 80% of the annual premium of 60 yuan, so the farmers only pay 20% or 12 yuan.

One cannot directly use OLS to directly measure the causal impact of having sow insurance on number of sows in the village, since there is a problem of unobserved heterogeneity. There could be confounders that exist that would affect both insurance decisions and production decisions, and the regression analysis does not account for that. For instance, risk preferences, which are not considered in this study, might affect both the farmers' preference for insurance as well as preference in other activities that might prolong the life of the sows. Cai et al. (2015) therefore use the incentives for animal husbandry workers (AHWs) as an IV to counter this unobserved heterogeneity. AHWs serve as the bridge between the formal institutions and the rural villages for matters involving animal husbandry, and are responsible for checking and marking the sow for insurance, as well as initiating the claim process in the event of a sow death. The AHWs are randomly assigned one of three incentive packages: the control group is given a higher base pay of 50 yuan, but is not given any additional incentive dependent on number of sows insured by the villages they go to. The low-incentive group was given a lower base pay of 20 yuan, but an additional small financial incentive of 2 yuan for every sow insured. The high-incentive group was given the same lower base pay of 20 yuan, but was given an additional higher financial incentive of 4 yuan for every sow insured. AHW incentives should be significantly and positively correlated with the number of insured sows, while only affecting number of total sows produced through the number of insured sows. This would make it suitable as an IV for this regression.

The results show that having insured sows significantly increases the number of sows. On average one additional insured sow increases the number of sows in the village by about 7.5 after 3 months, and 9.4 after 6 months. The study estimates the results using OLS, hence it is able to show if the insurance actually provided a negative welfare gain on average. Welfare gain in this experiment can only be measured as an average on the village level, and not on a household level.

The basic idea of "peace of mind," a phrase that is often heard from insurance salesmen, is indeed exactly what welfare reflects in this insurance setting. However, we have no interest in surveys of "well being" as measures of "peace of mind," let alone welfare: see Tafere, Barrett, and Lentz (2019).

5.4 | Other metrics

Chou, Grossman, and Liu (2014) state that the welfare gain from health insurance is the resultant improvement in infant and child health. Having (subsidized) health insurance should lower the effective price of medical care services such as prenatal and neonatal care, delivery, vaccinations and immunizations, and this price reduction should increase demand for these services. Supply is also encouraged, since insurance would guarantee payment for these services. They were interested in the effect of the National Health Insurance (NHI) coverage in Taiwan, which was introduced to all employees in 1995 when it was previously only offered to government employees. NHI was the only employee-based health coverage that provided benefits for infants of employees, and the premium was subsidized by the government. The nongovernment employed households were assigned as the treatment group and the control group was the government-employed households that were already receiving NHI coverage. They tested the impact of introducing NHI on postneonatal deaths, and found that there was a significant reduction in postneonatal deaths among farm households, but not among households who work in the private sector.

Chou et al. (2014) used difference-in-difference analysis to remove effects from unobserved trends while measuring the impact of insurance on postneonatal deaths. Using this methodology they are only able to estimate the average impact, and whether or not it was a positive or negative welfare gain on the sample level, and not for the individual.

McIntosh, Povel, and Sadoulet (2015) define basis risk as risk that is not covered by the insurance product, and test the impact of basis risk on insurance demand when it is expressed in two ways. The first is when insurance is partial, in the sense that the insurance will pay out when there is a shock but it might not completely cover the loss. The second is when insurance is probabilistic, in the sense that the insurance may fail to pay out when there is a shock. They used a field experiment with coffee farmers in Guatemala to understand the demand for index-based rainfall insurance. Insurance demand is calculated using a flexible utility function at the individual level to evaluate WTP for insurance. The risk parameters of the utility function were estimated from actual insurance choices using a nonlinear least squares estimator. They find that the average WTP for insurance increases as loss severity increases. This result holds even if the insurance payout remains constant, regardless of loss severity, which causes the insurance coverage to be even more partial as loss severity increases. Average WTP decreases, however, when payouts are more probabilistic: as the probability the insurance fails to pay for a shock increases, insurance demand decreases.

McIntosh et al. (2015) use the same insurance choices that they estimate risk parameters from to calculate the WTP of insurance. Applying the estimated risk parameters to the same data set that they were estimated from would result in the WTP for insurance to be biased, in the sense that these risk parameters are selected in order to maximize the likelihood that the observed insurance choice is the correct thing to do (by naïve revealed preference). There is no allowance for mistaken choices, in the behavioral sense, and for the estimated WTP estimated to be negative (in statistical expectation). They also have less than 10 data points per subject to use to estimate risk parameters, which makes their results very noisy statistically. Also, their results only apply to villages that self-report in a survey that they are vulnerable to excess rainfall risk. Since the survey is hypothetical, this adds another layer of uncertainty to the validity of their results.

6 | CONCLUSIONS

It is quite astonishing, in review, to see how many descriptive and normative evaluations of insurance have been undertaken with minimal attention to basic economic theory. To take two final examples, virtually no attempt is made to *design* products that reflect the risk preferences of individuals. One example of the casual nature of judgments in this area comes from Giné et al. (2008), p. 544, italics added), describing how the premium was set: "The policy premium was initially benchmarked on projected payouts using historical rainfall data (at least 25 years of data for each rain gauge were used). The premium was calculated as the sum of the expected payout, 25% of its standard deviation, 1% of the maximum sum insured in a year, plus a 25% administrative charge and 10.2% government service tax. In some cases the premium dictated by this formula was then reduced, because it was believed to exceed farmers' willingness to pay." After all of the formal actuarial arithmetic, we scratch our heads and just change things based on some hunch. To justify being puzzled by low take-up, Barberis, Huang, and Thaler (2006, p. 292) refer to "evidence of a strong need for health insurance," but by this all they mean is evidence that *average* health expenditures exceed the typical premium by a factor of $8.9 = 4670/525$. The *average* outcome for an insurance product bears no theoretical connection to the core risk management rationale for insurance: to mitigate the effects of outcome *variability*.

We can, and must, do better.

REFERENCES

Aggarwal, A. (2010). Impact evaluation of India's 'Yeshasvini' community-based health insurance programme. *Health Economics*, *19*(S1), 5–35.

Andersen, S., Cox, J. C., Harrison, G. W., Lau, M. I., Rutström, E. E., & Sadiraj, V. (2018). Asset integration and attitudes to risk: Theory and evidence. *Review of Economics and Statistics*, *100*(5), 816–830.

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

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583–618.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2013). Discounting behavior and the magnitude effect: Evidence from a field experiment in Denmark. *Economica*, *80*, 670–697.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2014). Discounting behavior: A reconsideration. *European Economic Review*, *71*, 15–33.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2018). Multiattribute utility theory, intertemporal utility, and correlation aversion. *International Economic Review*, *59*(2), 537–555.

Andreoni, J., & Sprenger, C. (2012a). Estimating time preferences from convex budgets. *American Economic Review*, *102*(7), 3333–3356.

Andreoni, J., & Sprenger, C. (2012b). Risk preferences are not time preferences. *American Economic Review*, *102*(7), 3357–3376.

Ang, A., Bekaert, G., & Liu, J. (2005). Why stocks may disappoint. *Journal of Financial Economics*, *76*(3), 471–508.

Arnott, R., & Stiglitz, J. E. (1991). Moral hazard and nonmarket institutions: Dysfunctional crowding out or peer monitoring? *American Economic Review*, *91*(1), 179–190.

Augenblick, N., Niederle, M., & Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, *130*(3), 1067–1115.

Barberis, N., Huang, M., & Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, *96*(4), 1069–1090.

Barberis, N., & Thaler, R. (2005). A survey of behavioral finance, chapter 1. *Advances in Behavioral Economics*, *2*, 1–75.

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, *103*(6), 2499–2529.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*, 226–232.

Biener, C., Landmann, A., & Santana, M. I. (2017). *Contract nonperformance risk and ambiguity in insurance markets* (Working Papers on Finance No. 2017/1). University of St. Gallen. Retrieved from SSRN: http://dx.doi.org/10.2139/ssrn.2908072.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, *62*(3), 395–407.

Blackburn, M. K., Harrison, G. W., & Rutström, E. E. (1994). Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics*, *76*(5), 1084–1088.

Bone, J., Hey, J., & Suckling, J. (2004). A simple risk-sharing experiment. *Journal of Risk and Uncertainty*, *28*(1), 23–38.

Breustedt, G., Bokusheva, R., & Heidelbach, O. (2008). Evaluating the potential of index insurance schemes to reduce crop yield risk in an arid region. *Journal of Agricultural Economics*, *59*(2), 312–328.

Bryan, G. (2018). Ambiguity aversion decreases the demand for partial insurance: Evidence from african farmers. *Journal of the European Economic Association*, forthcoming. https://www.eeassoc.org/doc/upload/Bryan20181205161428.pdf

Cai, H., Chen, Y., Fang, H., & Zhou, L.-A. (2015). The effect of microinsurance on economic activities: Evidence from a randomized field experiment. *Review of Economics and Statistics*, *97*(2), 287–300.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, *15*(3), 355–379.

Carriquiry, M. A., & Osgood, D. E. (2012). Index insurance, probabilistic climate forecasts, and production. *Journal of Risk and Insurance*, *79*(1), 287–300.

Carter, M. R., & Janzen, S. A. (2012). *Coping with drought: Assessing the Impacts of Livestock Insurance in Kenya* (Index Insurance Innovation Initiative Brief No. 2012-1), University of California at Davis. Retrieved from http://basis.ucdavis.edu/projects/i4-index-info/

Casaburi, L., & Willis, J. (2018). Time *vs.* state in insurance: Experimental evidence from contract farming in Kenya. *American Economic Review*, *108*(12), 3778–3813.

Chantarat, S., Mude, A. G., Barrett, C. B., & Turvey, C. G. (2017). Welfare impacts of index insurance in the presence of a poverty trap. *World Development*, *94*, 119–138.

Charness, G., & Genicot, G. (2009). Informal risk sharing in an infinite-horizon experiment. *Economic Journal*, *119*(537), 796–825.

Chou, S.-Y., Grossman, M., & Liu, J.-T. (2014). The impact of national health insurance on birth outcomes: A natural experiment in Taiwan. *Journal of Development Economics*, *111*, 75–91.

Chou, S.-Y., Liu, J.-T., & Hammitt, J. K. (2003). National health insurance and precautionary saving: evidence from Taiwan. *Journal of Public Economics*, *87*(9), 1873–1894.

Clarke, D., & Kalani, G. (2012). *Microinsurance decisions: Evidence from Ethiopia* (ILO Microinsurance Innovation Facility Research Paper No. 19). Geneva: International Labour Organization.

Clarke, D. (2016). A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, *8*(1), 283–306.

Clarke, D., & Dercon, S. (2009). *Insurance, credit and safety nets for the poor in a world of risk* (Working Papers 81), Department of Economics and Social Affairs, United Nations.

Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745–788.

Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., & Vickrey, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, *5*(1), 104–135.

Cole, S., Stein, D., & Tobacman, J. (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review (Papers and Proceedings)*, *104*(5), 284–290.

Coller, M., & Williams, M. B. (1999). Eliciting individual discount rates. *Experimental Economics*, *2*, 107–127.

Cox, J. C., & Sadiraj, V. (2006). Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior*, *56*, 45–60.

Cox, J. C., & Sadiraj, V. (2008). Risky decisions in the large and in the small: Theory and experiment. In Cox, J. C., & Harrison, G. W. (Eds.), *Research in Experimental Economics* (Vol. 12). Bingley, UK: Emerald, Risk Aversion in Experiments.

Cummings, R. G., Elliot, S., Harrison, G. W., & Murphy, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy*, *105*, 609–621.

Cummings, R. G., Harrison, G. W., & Rutstrom, E. E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive compatible? *American Economic Review*, *85*, 260–266.

DeBrauw, A., & Eozenou, P. (2014). Measuring risk attitudes among Mozambican farmers. *Journal of Development Economics*, *111*, 61–74.

DeJanvry, A., Dequiedt, V., & Sadoulet, E. (2014). The demand for insurance against common shocks. *Journal of Development Economics*, *106*, 227–238.

Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., & Taffesse, A. S. (2014). Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia. *Journal of Development Economics*, *106*, 132–143.

Di Mauro, C., & Maffioletti, A. (1996). An experimental investigation of the impact of ambiguity on the valuation of self-insurance and self-protection. *Journal of Risk and Uncertainty*, *13*, 53–71.

Di Mauro, C., & Maffioletti, M. (2001). The valuation of insurance under uncertainty: Does information about probability matter? *Geneva Papers on Risk and Insurance Theory*, *26*, 195–224.

Doherty, N. A., & Schlesinger, H. (1990). Rational insurance purchasing: Consideration of contract nonperformance. *Quarterly Journal of Economics*, *105*(1), 243–253.

Donfouet, H. P. P., Makaudze, E., Mahieu, P. A., & Malin, E. (2011). The determinants of the willingness-to-pay for community-based prepayment scheme in rural cameroon. *International Journal of Health Care Finance and Economics*, *11*(3), 209–220.

Elabed, G., & Carter, M. R. (2015). Compound-risk aversion, ambiguity and the willingness to pay for microinsurance. *Journal of Economic Behavior and Organization*, *118*, 150–166.

Elabed, G., Bellemare, M. F., Carter, M. R., & Guirkinger, C. (2013). Managing basis risk with multiscale index insurance. *Agricultural Economics*, *44*, 419–431.

Epstein, L. G., & Zin, S. E. (1990). 'First-order' risk aversion and the equity premium puzzle. *Journal of Monetary Economics*, *26*(3), 387–407.

Franco, L. M., Diop, F. P., Burgert, C. R., Kelley, A. G., Makinen, M., & Simpara, C. H. T. (2008). Effects of mutual health organizations on use of priority health-care services in urban and rural Mali: A case-control study. *Bulletin of the World Health Organization*, *86*(11), 830–838.

Ganderton, P. T., Brookshire, D. S., McKee, M., Stewart, S., & Thurston, H. (2000). Buying insurance for disaster-type risks: Experimental evidence. *Journal of Risk and Uncertainty*, *20*(3), 271–289.

Gerking, S., Dickie, M., & Veronesi, M. (2014). Valuation of human health: An integrated model of willingness to pay for mortality and morbidity risk reductions. *Journal of Environmental Economics and Management*, *68*(1), 20–45.

Giesbert, L. (2008). Demand for microinsurance in rural Ghana: A household survey report on the anidaso policy of the Gemini Life Insurance Company. *German Institute of Global and Area Studies*.

Giné, X., & Yang, D. (2009). Insurance, credit, and technology adoption: Field experimental evidence from Malawi. *Journal of Development Economics*, *89*(1), 1–11.

Giné, X., Townsend, R., & Vickery, J. (2007). Statistical analysis of rainfall insurance payouts in Southern India. *American Journal of Agricultural Economics*, *89*(5), 1248–1254.

Giné, X., Townsend, R., & Vickrey, J. (2008). Patterns of rainfall insurance participation in rural India. *World Bank Economic Review*, *22*(3), 539–566.

Gollier, C., & Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: A stochastic dominance approach. *Economic Theory*, *7*, 359–363.

Gumber, A. (2001). *Hedging the Health of the Poor: The Case for Community Financing in India*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/13663.

Hamid, S. A., Roberts, J., & Mosley, P. (2011). Can micro health insurance reduce poverty? Evidence from Bangladesh. *Journal of Risk and Insurance*, *78*(1), 57–82.

Hansen, J. V., Jacobsen, R. H., & Lau, Morten I. (2016). Willingness to pay for insurance in Denmark. *Journal of Risk and Insurance*, *83*(1), 49–76.

Hansson, B. (1988). Risk aversion as a problem of conjoint measurement. In Peter Gardenfors, & N.-E. Sahlin (Eds.), *Decisions, probability, and utility*. New York, NY: Cambridge University Press.

Harrison, G. W. (1992). Theory and misbehavior of first-price auctions: Reply. *American Economic Review*, *82*, 1426–1443.

Harrison, G. W. (1994). Expected utility theory and the experimentalists. *Empirical Economics*, *19*(2), 223–253.

Harrison, G. W. (2006a). Hypothetical bias over uncertain outcomes. In J. A. List (Ed.), *Using Experimental Methods in Environmental and Resource Economics*. Northampton, MA: Elgar.

Harrison, G. W. (2006b). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics*, *34*, 125–162.

Harrison, G. W. (2011). Experimental methods and the welfare evaluation of policy lotteries. *European Review of Agricultural Economics*, *38*(3), 335–360.

Harrison, G. W. (2013). Field experiments and methodological intolerance. *Journal of Economic Methodology*, *20*(2), 103–117.

Harrison, G. W. (2014a). Impact evaluation and welfare evaluation. *European Journal of Development Research*, *26*, 39–45.

Harrison, G. W. (2014b). Cautionary notes on the use of field experiments to address policy issues. *Oxford Review of Economic Policy*, *30*(4), 753–763.

Harrison, G. W. (2014c). Real choices and hypothetical choices. In S. Hess, & A. Daly (Eds.), *Handbook of Choice Modeling*. Northampton, MA: Edward Elgar.

Harrison, G. W., Lau, M. I., Ross, D., & Swarthout, J. T. (2017). Small stakes risk aversion in the laboratory: A reconsideration. *Economics Letters*, *160*, 24–28.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*, 1009–1055.

Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2015). Reduction of compound lotteries with objective probabilities: Theory and evidence. *Journal of Economic Behavior and Organization*, *119*, 32–55.

Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., & Ulm, E. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior and Organization*, *134*, 430–448.

Harrison, G. W., & Ng, J. M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance*, *83*(1), 91–120.

Harrison, G. W., & Ng, J. M. (2018). Welfare effects of insurance contract non-performance. *Geneva Risk and Insurance Review*, *43*, 39–76.

Harrison, G. W., & Ross, D. A. (2017). The empirical adequacy of cumulative prospect theory and its implications for normative assessment. *Journal of Economic Methodology*, *24*(2), 150–165.

Harrison, G., & Ross, D. (2018). Varieties of paternalism and the heterogeneity of utility structures. *Journal of Economic Methodology*, *25*, 42–67.

Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. In Cox, J. C., & Harrison, G. W. (Eds.), *Research in Experimental Economics* (Vol. 12). Bingley, UK: Emerald.

Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, *12*, 133–158.

Harrison, G. W., & Swarthout, J. T. (2016). *Cumulative prospect theory in the laboratory: a reconsideration* (CEAR Working Paper). Center for Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harstad, R. M. (2000). Dominant strategy adoption and bidders' experience with pricing rules. *Experimental Economics*, *3*, 261–280.

Herrero, C., Tomás, J., & Villar, A. (2006). Decision theories and probabilistic insurance: An experimental test. *Spanish Economic Review*, *8*(1), 35–52.

Hess, U. C. (2003). *Innovative financial services for rural India: Monsoon-indexed lending and insurance for small holders* (Agriculture and Rural Development Working Paper 9). World Bank, Washington DC.

Hill, R. V., Hoddinott, J., & Kumar, N. (2013). Adoption of weather-index insurance: Learning from willingness to pay among a panel of households in rural Ethiopia. *Agricultural Economics*, *44*, 385–398.

Hill, R. V., & Robles, M. (2011). *Flexible insurance for heterogeneous farmers: Results from a small scale pilot in Ethiopia* (IFPRI Discussion Paper 01092).

Hill, R. V., & Viceisza, A. (2012). A field experiment on the impact of weather shocks and insurance on risky investment. *Experimental Economics*, *15*(2), 341–371.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Irwin, J. R., McClelland, G. H., & Schulze, W. D. (1992). Hypothetical and real consequences in experimental auctions for insurance against low-probability risks. *Journal of Behavioral Decision Making*, *5*, 107–116.

Ito, S., & Kono, H. (2010). Why is the take-up of microinsurance so low? Evidence from a health insurance scheme in India. *The Developing Economies*, *48*(1), 74–101.

Jaramillo, F., Kempf, H., & Moizeau, F. (2015). Heterogeneity and the formation of risk-sharing coalitions. *Journal of Development Economics*, *114*, 79–96.

Jaspersen, J. G. (2016). Hypothetical surveys and experimental studies of insurance demand: A review. *Journal of Risk and Insurance*, *83*(1), 271–255.

Jensen, N. D., Barrett, C. B., & Mude, A. G. (2016). Index insurance quality and basis risk: Evidence from northern Kenya. *American Journal of Agricultural Economics*, *95*(5), 1450–1469.

Jensen, N. D., Mude, A., & Barrett, C. B. (2018). How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya. *Food Policy*, *74*, 172–198.

Jin, J., Wang, W., & Wang, X. (2016). Farmers' risk preferences and agricultural weather index insurance uptake in rural China. *International Journal of Disaster Risk Science*, *7*(4), 366–373.

Jowett, M. (2003). Do informal risk sharing networks crowd out public voluntary health insurance? Evidence from Vietnam. *Applied Economics*, *35*(10), 1153–1161.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kairies-Schwarz, N., Kokot, J., Vomhof, M., & Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory—An experiment. *Journal of Economic Behavior and Organization*, *137*, 374–397.

Karlan, D., Osei, R., Osei-Akoto, I., & Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *Quarterly Journal of Economics*, *129*(2), 597–652.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, *73*(6), 1849–1892.

Koufopoulos, K., & Kozhan, R. (2014). Welfare-improving ambiguity in insurance markets with asymmetric information. *Journal of Economic Theory*, *151*, 551–560.

Kunreuther, H., Novemsky, N., & Kahneman, D. (2001). Making low probabilities useful. *Journal of Risk and Uncertainty*, *23*(2), 103–120.

Laury, S. K., & McInnes, M. M. (2003). The impact of insurance prices on decision making biases: An experimental analysis. *Journal of Risk and Insurance*, *70*(2), 219–233.

Laury, S. K., McInnes, M. M., & Swarthout, J. T. (2009). Insurance decisions for low-probability losses. *Journal of Risk and Uncertainty*, *39*, 17–44.

Leblois, A., Quirion, P., & Sultan, B. (2014). Price vs. weather shock hedging for cash crops: Ex ante evaluation for cotton producers in Cameroon. *Ecological Economics*, *101*, 67–80.

Lin, W., Liu, Y., & Meng, J. (2014). The crowding-out effect of formal insurance on informal risk sharing: An experimental study. *Games and Economic Behavior*, *86*, 184–211.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–824.

Loomes, G., & Sugden, R. (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory*, *41*, 270–287.

Maccheroni, F., Marinacci, M., & Ruffino, D. (2013). Alpha as ambiguity: Robust mean-variance portfolio analysis. *Econometrica*, *81*(3), 1075–1113.

McClelland, G. H., Schulze, W. D., & Coursey, D. L. (1993). Insurance for low-probability hazards: A bimodal response to unlikely events. *Journal of Risk and Uncertainty*, *7*, 95–116.

McIntosh, C., Povel, F., & Sadoulet, E. (2015). *Utility, risk and incomplete insurance: lab experiments with Guatemalan cooperatives*. Unpublished Manuscript, School of Global Policy and Strategy, University of California San Diego. Forthcoming, Economic Journal.

McIntosh, C., Sarris, A., & Papadopolous, F. (2013). Productivity, credit, risk, and the demand for weather index insurance in smallholder agriculture in Ethiopia. *Agricultural Economics*, *44*, 399–417.

Mobarak, A. M., & Rosenzweig, M. R. (2012). Selling formal insurance to the informally insured. *Yale University Economic Growth Center Discussion Paper*, *1007*.

Mobarak, A. M., & Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review: Papers and Proceedings*, *103*(3), 375–380.

Norton, M., Osgood, D., Madajewicz, M., Holthaus, E., Peterson, N., Diro, R., … Gebremichael, M. (2014). Evidence of demand for index insurance: experimental games and commercial transactions in Ethiopia. *Journal of Development Studies*, *50*(5), 630–648.

Plott, C. R., & Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, *95*(3), 530–545.

Rabin, M. (2000). Risk aversion and expected utility theory: A calibration theorem. *Econometrica*, *68*, 1281–1292.

Rao, K. D., Waters, H., Steinhardt, L., Alam, S., Hansen, P., & Naeem, A. J. (2009). An experiment with community health funds in Afghanistan. *Health Policy and Planning*, *24*(4), 301–311.

Rutström, E. E. (1998). Home-grown values and incentive compatible auction design. *International Journal of Game Theory*, *27*, 427–441.

Schade, C., Howard, K., & Koellinger, P. (2012). Protecting against low-probability disasters: The role of worry. *Journal of Behavioral Decision Making*, *25*, 534–543.

Schneider, P., & Diop, F. (2001). *Synopsis of results on the impact of community-based health insurance on financial accessibility to health care in Rwanda*. World Bank, Washington, DC.

Schram, A., & Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review*, *55*, 799–819.

Segal, U. (1988). Probabilistic insurance and anticipated utility. *Journal of Risk and Insurance*, *55*(2), 287–297.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica*, *58*(2), 349–377.

Segal, U., & Spivak, A. (1990). First order versus second order risk aversion. *Journal of Economic Theory*, *51*(1), 111–125.

Skees, J. R., & Gober, S. (2001). Varangis, Panos; Lester, Rodney, and Kalavakonda, Vijay. *Developing Rainfall-based Index Insurance in Morocco* (Vol. 2577). World Bank Publications.

Slovic, P., Fischhoff, B., Lichtenstein, S., Corrigan, B., & Combs, B. (1977). Preference for insuring against probable small losses: Insurance implications. *Journal of Risk and Insurance* , *44*(2), 237–258.

Spiegler, R. (2019). Behavioral economics and the atheoretical style. *American Economic Journal: Microeconomics*, *11*(2), 173–94.

Sydnor, J. (2010). (Over)insuring modest risks. *American Economic Journal: Applied Economics*, *2*(4), 177–199.

Tafere, K., Barrett, C. B., & Lentz, E. (2019). Insuring well-being? Buyer's remorse and peace of mind effects from insurance. *American Journal of Agricultural Economics*, *101*(3), 627–650.

Thornton, R. L., Hatt, L. E., Field, E. M., Islam, M., Diaz, F. S., & Gonzales, M. A. , (2010). Social security health insurance for the informal sector in Nicaragua: A randomized evaluation. *Health Economics*, *19*, 181–206.

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica*, *62*(3), 539–591.

Vedenov, D. V., & Barnett, B. J. (2004). Efficiency of weather derivatives as primary crop insurance instruments. *Journal of Agricultural and Resource Economics*, 387–403.

Wagstaff, A., & Pradhan, M. (2005). *Health insurance impacts on health and nonmedical consumption in a developing country* (3563). World Bank Publications.

Wakker, P. P., Thaler, R. H., & Tversky, A. (1997). Probabilistic insurance. *Journal of Risk and Uncertainty*, *15*(1), 7–28.

Zhang, J., Tan, K. S., & Weng, C. (2018). *Index insurance design* (Working paper), Nanyang Technological University,

Zimmer, A., Gründl, H., Schade, C. D., & Glenzer, F. (2018). An incentive-compatible experiment on probabilistic insurance and implications for an insurer's solvency level. *Journal of Risk and Insurance*, *85*, 245–273.

Zimmer, A., Schade, C. D., & Gründl, H. (2009). Is default risk acceptable when purchasing insurance? Experimental evidence for different probability representations, reasons for default, and framings. *Journal of Economic Psychology*, *30*(1), 11–23.

How to cite this article: Harrison GW, Ng JM. Behavioral insurance and economic theory: A literature review. *Risk Manag Insur Rev*. 2019;22:133–182. https://doi.org/10.1111/rmir.12119