Literacy and the quality of index insurance decisions

Glenn Harrison[1,2] · **Karlijn Morsink[3]** · **Mark Schneider[4]**

Abstract

There is widespread concern in developing countries with the expansion of formal insurance products to help manage significant risks. These concerns arise primarily from a lack of understanding of insurance products, general failures of financial literacy and the need to use relatively exotic products in order to keep costs down for poor households. We investigate the importance of incentivized measures for general understanding, as well as domain-specific knowledge of the decision context on the purchase and the quality of index insurance decisions. We evaluate the quality of financial decisions by comparing the individual expected welfare outcomes of a number of decisions each individual makes to purchase index insurance or not. We find that excess purchase is an important driver of welfare losses, and that our incentivized measure of domain-specific literacy plays a critical role in bringing about better quality index insurance decisions.

Keywords Index insurance · Behavioural welfare economics · Financial literacy · Insurance literacy

JEL classification D81 · O2 · D90 · D78

✉ Glenn Harrison
gharrison@gsu.edu

Karlijn Morsink
k.morsink@uu.nl
https://www.karlijnmorsink.com

Mark Schneider
mschneider@gsu.edu

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

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

3   School of Economics, Utrecht University, Utrecht, The Netherlands

4   Center for the Economic Analysis of Risk, Georgia State University, Atlanta, USA

1 Introduction

Financial products are becoming increasingly complex, at the same time that decision-making is speeding up and taking place through digital channels. This raises serious concerns about the extent to which individuals understand the financial decisions they make. Despite the fact that considerable amounts of money are spent on efforts to increase general financial literacy as well as the understanding of specific financial decisions, most of these efforts are evaluated in terms of their potential to increase purchases rather than enhance understanding or lead to *better* financial decisions.[1] Furthermore, even if understanding is explicitly tested, it is often in terms of general financial literacy[2] and not the domain-specific understanding of the decision context in question.

We investigate the importance of general financial understanding, as well as domain-specific knowledge of the decision context on the quality of index insurance decisions. We evaluate the quality of financial decisions by analysing the expected consumer welfare of each individual's decisions to purchase or not purchase a variety of index insurance product. We focus on index insurance because it offers great potential for managing risks in a cost-efficient manner, but, due to its design features, can be a complex product to understand and is not welfare-enhancing for all (Clarke 2016). We provide rigorous incentivized metrics for evaluating both general financial literacy and domain-specific index insurance literacy, as well as complementary unincentivized metrics for behavioral traits such as "cognitive reflection" and "fluid intelligence." We then investigate who are winners and losers in terms of the expected consumer welfare of decisions made by each individual.

We conduct lab experiments with 150 subjects who each make 54 decisions to purchase index insurance or not. We elicit the bias and confidence of each subject with respect to general financial literacy and domain-specific index insurance literacy using incentivized experiments. We also survey their general financial literacy, cognitive reflection, and fluid intelligence using hypothetical surveys. To estimate the expected consumer welfare from insurance decisions, we elicit the risk preferences of each individual separately, from a risky lottery task. We use the risk preferences that were estimated at the individual level to assess the expected aggregate consumer welfare of all the insurance decisions each individual made. We can then separate individuals that achieved the highest level of expected consumer welfare and individuals that achieved the lowest level of expected consumer welfare, and investigate to what extent our measures of understanding determine the level of expected consumer welfare achieved.

We deliberately start with a laboratory experiment, to cost-effectively identify the measures of comprehension that can be used in the field, to develop our measures of welfare, and to observe those that do or do not purchase the product. Of course we understand the importance of field context, and have long been advocates for

---

[1] See, for example, Gaurav et al. (2011), Cole et al. (2013), Dercon et al. (2014) and Hill et al. (2016).

[2] This is typically done through hypothetical survey questions, where individuals have no incentive to correctly respond. See, for example, Lusardi and Mitchell (2014) and Mitchell et al. (2011).

the importance of field experiments (Harrison and List 2004). Field experiments are, however, expensive at interesting scale and are often constrained to study few decisions.

We find that higher domain-specific literacy about the index insurance decision context, as well as greater fluid intelligence, significantly increase the likelihood that an individual achieves the highest level of expected consumer welfare, while significantly reducing the likelihood that they achieve the lower level of expected consumer welfare. These results indicate that domain-specific literacy plays a critical role in ensuring that individuals avoid welfare-reducing insurance decisions. For those that achieve the lowest level of consumer welfare, predominantly because of *excess* take-up, greater financial literacy does not seem to be correlated with take-up. For those individuals that achieve the highest level of expected welfare, predominantly by more cautious take-up, an increase in domain-specific literacy, fluid intelligence, an indication of greater "cognitive reflection," and the hypothetical financial literacy score, significantly increases the likelihood of purchase. The asymmetry of these results when comparing each individual in either tail of the distribution of consumer welfare suggests that the decision-making processes driving their insurance decisions are different. Finally, hypothetical survey responses for general financial literacy are not correlated with the expected consumer welfare of insurance decisions.

Index insurance has great potential as a risk management contract but there are serious marketing problems in the field. The core problem is that index insurance entails a compound risk that implies that the insured faces a basis risk. The canonical contract is built around some public index that is transparent and applies equally to all covered by the contract. The index has a trigger threshold that determines if payment is to be made or not to the insured, irrespective of any idiosyncratic loss being realized. Index insurance greatly reduces moral hazard because actions by the insured do not effect the index level or payout decision, whether or not those actions could be observed by the insurance company. Similarly, adverse selection is reduced because there are no potential insured agents that have any influence whatsoever on the index level or payout decisions, again whether or not their idiosyncratic risk could be observed.

The simplest examples of index insurance are built around physical measures of the level of rainfall using a standard rain gauge, or an average of rain gauges in some region. Based on historical data on aggregate rainfall in the region, actuaries can determine the risks of levels of rainfall that can lead to crop or livestock distress, where the connection between rainfall and distress can be determined separately.[3] In principle the use of a public index can be extended to many other settings, and this makes index insurance attractive for global risks such as climate change or pandemics which can be correlated with indices such as average temperature change or aggregate mortality change.

---

[3] For example, agricultural research stations might determine the link to distress in different crops or livestock in a controlled manner. The point is that data on idiosyncratic crop or livestock distress from potential claimants is not needed.

There are many variants on this canonical contract, and it is precisely the gulf between the abstract, canonical index insurance contract and the contracts that are marketed in the field that motivates our focus on assessing the quality of index insurance purchase decisions. The implication of this gulf is that "literacy," "cognition" and "intelligence" likely play a critical role in whether individuals understand index insurance contracts. If they do not understand the contract, it follows immediately that welfare-reducing decisions might be made with respect to insurance purchases.[4]

We make two general contributions. The first is to the literature on the measurement of literacy, cognitive reflection, and fluid intelligence. In terms of *literacy*, the literature has been dominated by the use of the general financial literacy questions of Lusardi and Mitchell (2014), indeed often just their Big 3 or Big 5 questions. These survey questions are not incentivized, and refer to aspects of formal financial products (savings accounts, mortgages and stocks versus bonds). Following Di Girolamo et al. (2015) we measure the confidence that an individual has in their knowledge of these concepts. In terms of *cognitive reflection*, what started as an intuitive attempt, also with just a Big 3 set of questions by Frederick (2005), has been widely portrayed as a measure of cognition or even intelligence. Even if it is unclear if these questions capture cognitive reflection or are just "good math problems," we demonstrate that they play a role in good quality insurance decisions. Finally, in terms of *intelligence*, there are measures of fluid intelligence and measures of crystallized intelligence, where the former measures the ability to reason well in novel settings and the latter measures knowledge of facts and how to apply reasoning *methods* (such as formal logic or mathematics). For new financial products, fluid intelligence would be the measure of most interest when studying the quality of decisions to adopt the product or not. We contribute to this literature in two ways. Firstly, we evaluate to what extent these existing measures actually correlate with welfare of insurance decisions, and we show that the majority do not. Secondly, we develop an incentivized domain-specific measure of knowledge about index insurance decisions and show that this has an important role to play in enhancing the likelihood that an individual makes welfare-enhancing decisions.

---

[4] By implication, explained later, we reject reliance on direct revealed preference as the basis for assessing the welfare of decisions to purchase insurance or not.

   The second contribution is to the literature on index insurance, specifically the literature that focuses on the role of the understanding of the product on insurance decisions. Carter et al. (2008) used an artefactual field experiment, much like our own laboratory experiment, as a financial education treatment.[5] No results were reported on correlates with actual take-up in the experiment or field. Gaurav et al. (2011) evaluated a financial literacy survey that had been extended to consider some insurance-specific issues, as well as a randomized intervention with an extended insurance education treatment. Their measures of take-up refer to an actual, field product, and all behavioral determinants are elicited with non-incentivized surveys.[6] They find that the education intervention increases take-up from 8 to 16%, but do not evaluate if these were welfare-improving increases in take-up.[7] Finally, Hill et al. (2016) find that households in their field experiment with index insurance that receive "more intense" training in insurance products have a 5 percentage point increase in take-up in the short-run. On the other hand, this effect disappears in the medium-run. They also show that negative loadings, leading to greater take-up, appear to increase understanding of the product, and hence longer-term take-up. The extent of "understanding" was based (p. 1258) on hypothetical survey questions, although the questions did canvass domain-specific knowledge of the index insurance product. (p. 1261, fn. 19). Once again, the focus is on take-up, and not the quality of the decision. We contribute to this literature by assessing the effect of understanding on the expected consumer welfare of insurance decisions, rather than on take-up, and we show that, actually, excess take-up of index insurance is an important driver of welfare losses caused by insurance decisions. Furthermore, we show that domain-specific index insurance literacy is an important driver of the likelihood that individuals have positive welfare outcomes when making insurance decisions.

   We present our conceptual framework (Sect. 2) and our experimental design in detail (Sect. 3). Then we review descriptive evidence from the experiments spanning 150 subjects who each make 54 decisions to purchase different index insurance contracts and provide evidence of the correlations of our literacy, cognition and intelligence measures on expected consumer welfare and take-up, with emphasis on the differences between good quality decisions and bad quality decisions in terms of welfare (Sect. 4). Conclusions and implications for policy are provided (Sect. 5).

---

[5] Their experiment was much more closely tied to the decision-making environment of the field setting in which their sample was drawn, which is appropriate for a financial education treatment in the field.

[6] Examples include Chantarat and Mude (2010) for Kenya and Aina et al. (2018) for Nigeria. They elicit willingness to pay for hypothetical index insurance contracts, which is, in principle, the Certainty Equivalent of the contract to the individual. If one also knew the Certainty Equivalent for that individual at the hypothetical status quo risk, it would be possible to infer Expected Consumer Surplus just as we do here in Section 2.3. Chantarat et al. (2013, p. 209) illustrate how this information might be used to design index insurance contracts to encourage take-up. The limitations of the "contingent valuation" surveys for eliciting reliable responses are well-documented: see Harrison (2006, 2014).

[7] In addition they consider six "marketing interventions" for households randomly assigned to the insurance education treatment. The only intervention to have a significant impact on take-up is a money-back guarantee of a premium refund if there is no payout. This is not a "message" (p. 153), but a massive negative loading treatment. It is certainly a marketing intervention, since "loss leading pricing" is a common marketing tool for many new products when first introduced. Hence their results show that clients are indeed sensitive to loading.

2 Conceptual framework

We first review some features of real-world index insurance contracts that are marketed, some observations on how they are marketed, and efforts within the industry to assess and mitigate comprehension problems. We then explain what we mean by literacy, cognition and intelligence. Finally we discuss how we identify the "quality" of purchase decisions.

2.1 The marketing of index insurance

For many years we have been conversing with insurance companies, non-governmental organizations, government bodies and insurance regulators charged with marketing, or evaluating the marketing, of index insurance products in developing countries.[8] Several issues arise, some more serious than others.

First, field products are more *complex* than the canonical product described above. The index might not be as physically transparent as a rain gauge. In many cases it has evolved into complicated algorithmic transformations of satellite imagery, historical data, selected "ground-proofing" validations, and so on. All of these additional data sources are intended to reduce basis risk, and are often cost-effective, but they inevitably render the index harder to explain and even harder to validate to clients. In a related vein, the index generated by these data could be some precise, formal "vegetation index," which may be an appropriate or superior statistical proxy for generalized distress of livestock or crops, but which cannot be as easily understood as levels of rainfall by clients. Mitigating these concerns are efforts to display index outcomes in the form of contour maps or choropleth maps.

Index products may also have two or more threshold points, flagging different fixed levels of indemnification. One threshold might be "serious," leading to a modest payout, and another threshold might be "catastrophic," leading to a much higher payout. More common is to find index insurance contracts that have a single threshold which then triggers a linear payoff depending on the level of the continuous index. For example, a threshold might be set at the 20th percentile of a distribution defined over the index, such that any index value above that threshold generates no payoff; and where index values further below that threshold generate a larger share of the level of indemnification. For example, an index at the 15th percentile might trigger a payoff that is $(20 - 15)/20 = 5/20 = 25\%$ of the maximum payout, and an index at the 10th percentile might trigger a payoff that is $(20 - 10)/20 = 10/20 = 50\%$ of the maximum payout. Of course, this remains an index or parametric product, since everyone covered by the product receives that payout. But the function which maps index values below the threshold and *payouts* might not reflect the function which maps index values below the threshold and *losses*: one might be linear, and often is, and the other highly non-linear. This

---

[8] For example, at meetings of the International Microinsurance Conference since 2010 (http://www.munichre-foundation.org/en/Inclusive_insurance/International_Conference_on_Inclusive_Insurance.html). This conference is now re-named as the International Conference on Inclusive Insurance.

function is also, in the statistical nature of extremes, typically less precisely estimated than other parts of the distribution.[9]

Second, it is very hard to find out *what information is actually provided* to potential clients. This derives largely from the understandable decentralization necessary to contact potential clients in small groups, leading to primarily oral communication between the salesperson and the potential client. This level of decentralization often means that "partners" of the insurance company are the ones actually charged with explaining and selling the product, leading to even less knowledge by the insurance company of what is communicated. Even if standardized scripts are prepared, as they often are, there is little or no enforcement of the conversation that actually occurs. In many settings literal literacy is a barrier, forcing even more reliance on oral and pictorial representations. Even if insurance regulators require some formal presentation of scripts, these are not readily available (in any language), and often regulators also rely on oral presentations of intent. In many instances the potential client is being presented with an insurance product that is bundled with some other product, often a loan or access to credit, that they really want, and trust the broker of that primary product.[10] Finally, details of marketing are often proprietary, for understandable reasons.

Third, some contracts *do not have basis risk explained*. This is a more sensitive issue, of course. In many cases basis risk is implied, by statements that the trigger for a payout is generated by some index, and nothing else. In a formal sense this leaves the client to draw the conclusion that the payout does not depend on what happens to them idiosyncratically, but if the default belief is that it does depend on what happens to them idiosyncratically then there is some presumption that it should be clearly stated.[11] We have encountered some marketing materials in which positive basis risk (where the client does not suffer a loss but the index pays out) is mentioned, but downside basis risk (where the client suffers a loss but the index does not pay out) is not mentioned. This asymmetry is clearly problematic. We have also encountered comments that questions about basis risk are sometimes dismissed as not statistically important, implying that the correlation between index outcome and idiosyncratic outcome is presumed to be close to 1. As one experienced industry hand put it, you do not *start* selling a product by leading with the biggest limitation of the product.

Fourth, there is widespread *misunderstanding of the nature* of an insurance product in general. The definition of a canonical indemnity insurance product alerts us to how unusual it must seem in the abstract.

---

[9] And, of course, from the perspective of the client, it is their subjective beliefs that matter for purchase decisions

[10] See McPeak et al. (2010) and Shee et al. (2015) for examples of an artefactual field experiment used to explain basis risk in the context of naturally-occurring insurance purchases.

[11] The law on deception by omission is complicated. We are making a simpler ethical statement. Ortmann and Hertwig (2002, p. 113) discuss this form of deception in experiments as a violation of default assumptions.

> You give me some money now, and I promise to come back in the next year if something bad happens to you and offset some or all of your losses with some money. The money you pay me up front is mine to keep if nothing bad happens to you. But the amount of money I provide to you if something bad happens will typically be many times larger than the amount of money you give me in any year.

Put aside the reasonable concerns with "non-performance" and trust that the insurance contract will be honored. Many people considering the purchase of this type of contract view it as an investment product rather than a risk management product.[12] If the product does not payout a claim in the first year or two, it is then viewed as a worthless investment.

This misunderstanding is not unique to developing countries. It is still found in many settings in developed countries.[13] However, it is important to understand the (recent) history of basic insurance education in the United States to get some perspective on how hard it is for the general population to understand insurance as a risk management product. Around 1905 the insurance industry in general, and life insurance companies in particular, were widely held in low repute by the general public, as a result of the Armstrong Investigation of financial misconduct between insurance companies and investment banks reviewed by North (1954). This was slowly overcome by a partnership between industry and academe, documented by Stone (1960). Ernest Clark was a general agent for a life insurance company, and held senior positions in the National Association of Life Underwriters. And in 1913 Solomon Huebner was a Professor in the newly-formed Department of Insurance at Wharton. Together they spent many years establishing formal academic accreditation for life insurance sales agents, stressing fundamental concepts of insurance as the way to sell the product and obtain renewals year after year.[14] Huebner repeatedly toured the United States to lecture to collections of insurance agents about the principles of life insurance. By roughly 1935 these efforts had generated a stable industry with agents schooled in selling a product that stressed the principles of insurance.

The point of this "history lesson" is to realize that insurance literacy for a general population, and perhaps even a targeted sub-population, is not likely to be something

---

[12] Sadly, some researchers also view it this way. For example, Matsuda et al. (2019) claim (p. 482) that the *direct* effect of an index insurance scheme with respect to weather risk management is whether it pays out *ex post* in a given period, and that the *indirect* effect is the *ex ante* "peace of mind" from knowing that it would, in expectation, reduce risks to income. This is exactly backwards from the perspective of insurance economics.

[13] For example, Brown et al. (2017) exhaustively examine the extent to which Americans understand realistic variations on the annuitization decision in a series of hypothetical surveys, and find significant mistakes that they attribute to limited literacy and cognitive constraints.

[14] Stone (1960, p. 127ff) reproduces a major speech he delivered repeatedly to such associations. It stressed the ideas that a human life of the income-earner had value for a family, just as surely as property had value; that insurance should not be viewed as an investment to be judged by the pecuniary return it might provide, but for the *ex ante* "peace of mind" it provided; and that life insurance complemented savings by hedging against the possibility that the time period to undertake savings is cut short.

that any one-time intervention can possibly inculcate. At the level of research, one really needs longitudinal studies of literacy interventions, several "years" of experience with the product in good times and bad times, and the opportunity for repeat purchases.[15] The argument that Solomon Huebner made was that attrition was often a signal of a poor quality product that was "poor quality" only because it was a misunderstood product.

Fifth, the *manner in which sales agents are rewarded* does not incentivize them to ensure that the product is properly understood as a risk management product. In many cases the compensation is based on the number and premium value of initial sales. In most developed countries the sales agent is also compensated for repeat purchases, which of course come, in part, from understanding that it is not an investment product.[16]

We stress that there are many "good actors" marketing index insurance products in developing countries. We know of many insurance companies that pro-actively undertake assessment of comprehension using small surveys at key points in the process. In large part these are, correctly, justified by wanting to maintain good customer relations, to ensure that attrition does not set in as clients become disenchanted with products that they did not properly comprehend in the first instance.[17] Equally, we have heard stories of some insurance regulators holding applicants with new products to the fire, requiring them to explain how they explain different contingencies to potential clients.

2.2 Literacy, cognition and intelligence

The comprehension of any insurance product depends on a number of factors, which we bundle under the headings of literacy, cognition and intelligence. We employ explicit measures for these concepts. Our perspective is that the responses we see

---

[15] There are exceptions. Many people attribute the spectacular and sustained growth in the U.S. life insurance market to the harsh "demonstration effect" of the Spanish flu pandemic in 1918. A striking feature of the mortality incidence of that pandemic was the high rate among young men, making the benefits of life insurance immediately visible to all because of their historical importance as income earners for households (Milevsky 2019).

[16] In the United States the salary structure for insurance agents can vary with the policy line and the connection to an insurance company. So-called "captive" agents are tied to one insurance company, and receive lower commissions, but have the benefit of a base salary and office infrastructure. Agents that are "independents" can sell products from a wide range of insurers, and generally receive larger commissions than captives. For captives the follow-on commissions from renewals typically declines over a 3-year period, and can even be zero thereafter. Captives often have the opportunity of a "performance bonus," depending on total sales in a given period. For independents, renewals typically generate a constant percentage commission over time. Sales of life insurance and health insurance tend to have commissions front-loaded more than other product lines, for both captives and independents, but the commissions on an initial sale can be very high. Sales of automobile and home insurance tend to attract lower commissions, but have smaller reductions for renewals.

[17] We also know of some consulting companies that superficially purport to do this for clients, but have a scientific gloss on their methods that makes it hard to believe that they are serious. The fact that someone *says* that they are evaluating comprehension does not mean that they are doing so at all, let alone rigorously.

in insurance purchase decisions come from a "cognitive production function" that pools effort, prior knowledge, heuristics, logic, and time.[18] Prior knowledge, in turn, comes largely from what we refer to as literacy. We also view intelligence as broader than just logic, and allow it to include a stage in which the individual can use some heuristic, whether to use a considered heuristic, or to use a convenient heuristic. The fact that there is always some opportunity cost to applying effort or time is enough for us to see that financial incentives might matter, even if they could "crowd out" and counteract intrinsic incentives at some point. Of course, just because we assume the existence of some cognitive production function does not mean that we are assuming that it is employed efficiently.

2.2.1 Literacy

Literacy measures are typically constructed based on multiple-choice questions where each individual is deemed literate with respect to the topic of that question. Indices of how literate the individual is, constructed by a simple sum of the correctly answered questions, are then used in estimations to correlate with downstream behavior.[19] We utilize methods of quantitative subjective belief elicitation to assess the full distribution of a decision-maker's knowledge with respect to an objectively true answer. We use an incentivized approach to increase accuracy and ascertain how precise a decision-maker's knowledge is.[20]

2.2.2 Cognitive reflection

The popular Cognitive Reflection Test (CRT) seeks to measure a person's unincentivized tendency to override an incorrect "gut" response and engage in further reflection to find a correct answer. A typical example of a CRT question is: "A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?" The intuitive answer or "gut" response that people typically give to this question is ten cents, while the correct answer is five cents. The gut response is also referred to as a false lure, and by Baron (2019) as a test for "actively open-minded thinking" that values questioning initially-favored conclusions. The CRT questions we use are taken from Frederick (2005) and Primi et al. (2016), extending the battery to avoid the risk of subject familiarity with the original questions.

2.2.3 Intelligence

The specific measure of intelligence that we consider is the Raven Advanced Progressive Matrices (RAPM) test, documented by Raven et al. (1998). The RAPM is widely viewed as a major test for "fluid intelligence" or "analytic intelligence"

---

[18] Camerer and Hogarth (1999) provide an explicit statement of this perspective on "produced" cognition.

[19] See, for example, Lusardi and Mitchell (2014).

[20] See Di Girolamo et al. (2015) and Harrison et al. (2017) for details.

(Penrose and Raven 1936). The RAPM test we employed consists of a set of 12 problems called Set I. The RAPM consists of a block of 9 images, arrayed as a 3 × 3 matrix, and with one image deleted. The subject is presented with 8 possible solutions to fill in the deleted image, and is asked to select the correct image.

2.3 The quality of insurance decisions

The objective of insurance is, ex ante any actual loss, to reduce the expected variability of consumption, by paying an insurance premium now, in exchange for a claim payment later, in case the future state is realised where the insured experiences a loss. The most obvious approach to evaluate the value of insurance is to assess the expected variability of consumption in the light of the risk preferences and beliefs of an individual. Therefore we focus on an assessment of the expected welfare of buying insurance to an individual compared to the expected welfare of not buying insurance to the same individual and we term this measure the Expected Consumer Surplus (ECS) of insurance decisions. We use this measure because insurance is an ex ante risk management product, and should be evaluated in terms of the value of its ex ante protection, rather than in terms of its Realised Consumer Surplus conditional on the realisation of states, and thus potential losses.

The calculation of the ECS from the choice to purchase insurance is, by itself, not controversial. What is perhaps controversial, to some, is how one takes the logic of that calculation to do more than just *describe* when someone would be expected to purchase insurance and say *normatively* whether they should purchase insurance. To explain the two steps, which are central to our evaluation of the effects of literacy on the quality of financial decisions, we can use one of the specific examples given to our subjects (displayed later as Fig. 1).

Fig. 1 Example screen of index insurance purchase decision

Assume that there is an endowment of $60, and a loss probability to the individual of 20%. If there is a loss, it has value $39, leaving the individual with $21 = $60 - $39. This is a lottery {$21, 0.2; $60, 0.8} in the usual notation, for an individual that *does not purchase* insurance. This lottery has an Expected Value (EV) of $52.20 = $21 × 0.2 + $60 × 0.8.

Assume for now that the individual behaves consistently with EUT, and has a Power utility function $u(x) = x^r$ with parameter r = 0.507, the estimated average of our sample. Then the Expected Utility (EU) of this lottery in which the individual *does not purchase* insurance is equal to $u(\$21) \times 0.2 + u(\$60) \times 0.8$. With a Power utility function we know that the Certainty Equivalent (CE) of the lottery is the CE value that solves $u(CE) = EU$. Hence the $CE = EU^{(1/r)} = 4.6813 \times 0.2 + 7.9712 \times 0.8 = 7.3132^{(1/0.507)} = 7.3132^{1.9724} = \$50.62$.

We now repeat the logic of this calculation to obtain the CE when *purchasing* the insurance. The premium for the insurance is $12.50, which happens to be the actuarially-fair premium. This insurance contract provides full indemnification of all losses, and there is no risk of non-performance. The probability that the index matches the outcome of the individual is 0.8. This matching probability implies a correlation between the individual's loss and the index outcome of $0.6 = 1 - (2 \times (1 - 0.8))$. In this case there are 4 possible outcomes, which we spell out in full:

1. The individual pays the $12.50 premium, the individual *experiences a loss* of $39, and *the index matches* the outcome of the individual. Hence the insurance pays out $39 and the individual ends up with $47.50 = $60 − $12.50 − $39 + $39. The compound risk of the individual having a loss and the index matching is $0.2 \times 0.8 = 0.16$.
2. The individual pays the $12.50 premium, the individual *experiences a loss* of $39, and *the index differs* from the outcome of the individual. Hence the insurance does not pay out and the individual ends up with $8.50 = $60 − $12.50 − $39. The compound risk of the individual having a loss and the index differing is $0.2 \times (1 - 0.8) = 0.2 \times 0.2 = 0.04$.
3. The individual pays the $12.50 premium, the individual *experiences no loss*, and *the index matches* the outcome of the individual. Hence the insurance does not pay out and the individual ends up with $47.50 = $60 − $12.50. The compound risk of the individual having a loss and the index matching is $(1 - 0.2) \times 0.8 = 0.8 \times 0.8 = 0.64$.
4. The individual pays the $12.50 premium, the individual *experiences no loss*, and *the index differs* from the outcome of the individual. Hence the insurance pays out $39 and the individual ends up with $86.50 = $60 − $12.50 + $39. The compound risk of the individual having a loss and the index differing is $(1 - 0.2) \times (1 - 0.8) = 0.8 \times 0.2 = 0.16$.

The upshot of these calculations is that the lottery involved in the purchase decision is {$47.50, 0.16; $8.50, 0.04; $47.50, 0.64; $86.50, 0.16}. This lottery has an EV of $52.18, slightly lower than the EV of the lottery in which no insurance is purchased. With the same degree of risk aversion for the individual making the decision to

purchase insurance or not, the EU of the lottery to purchase the insurance is 7.3183, which has a CE of $50.69.

Comparing the two lotteries, we can now see that with this degree of risk aversion the individual gains an ECS of $0.07 = $50.69 − $50.62 from purchasing the insurance. This is the *Expected* CS, since it refers to all payouts from all possible events when they are weighted by their probabilities of occurring. It has nothing whatsoever to do with whether the individual actually had a loss or the index insurance actually matched that individual outcome.

The same calculation can be repeated for different levels of risk aversion. We already know from the EV calculations that the ECS for a risk-neutral individual would be slightly negative, and equal to −$0.02 = $52.18 − $52.20. If the individual was more risk averse than the average, and $r = 0.3$ say, then the ECS would increase to $0.12; and if the individual was less risk averse than the average, and $r = 0.7$ say, then the ECS would decrease to $0.02. Hence if we know the risk aversion of each subject, as we do, we can calculate the ECS from making either the purchase decision or the non-purchase decision.

As we vary the actuarial parameters facing the individual, the ECS from making the correct decision varies. *Ceteris paribus*, a lower premium means a higher ECS if purchasing the product was the correct thing to do. For this reason, that some product offerings are better than others, we also consider the percentage of the total ECS that the individual realizes over all decision compared to the total ECS that the same individual would have realized over all decisions if all decisions were correct. We term this Efficiency, and it effectively normalizes across subjects for the different product offerings, since each individual faces the same set of 54 product offerings by design.

Everything to this point is standard *Insurance 101* arithmetic. Two complications arise when we recognize that a substantial fraction of subjects behave consistently with Rank Dependent Utility (RDU) rather than EUT, and that many subjects violate the Reduction of Compound Lotteries (ROCL) axiom. Both are significant for the proper structural evaluation of the ECS of observed decisions with respect to index insurance, but do not change the general logic of our approach: see Harrison et al. (2020a) for details. Violations of ROCL obviously matter for understanding the basis risk of index insurance, since basis risk is just compound risk. And RDU will always matter for insurance defined over probabilistic outcomes.

Now assume, however, that we observe our average risk averse EUT individual, who has $r = 0.507$, decides to *not* purchase the insurance. The arithmetic tells us that this individual has foregone $0.07 in expectation that could have been gained by purchasing the insurance. Nothing controversial with that statement.

But now switch from *describing* what we observe with our model to making *normative evaluations* of what we observe. What if we know that the risk aversion of the individual is 0.507, and observe the wrong choice for that risk preference? One response might be to say that we must have the wrong risk aversion for the individual, and that it must be 0.86 or higher, since that is the risk aversion parameter that can explain *not* purchasing insurance as the correct decision (the ECS gain then is $0.01 or higher). If, however, we recognize that individuals make mistakes, misunderstand the decision-context, or are not sophisticated about their preferences, we

cannot rely on direct revealed preference to infer preferences from financial decisions. This would, by definition of revealed preference, *never* show that someone made the wrong decision (in expectation). Hence, if our objective is to evaluate the extent to which the financial decisions of each individual enhance or reduce their welfare, we need to have some experimental task that is separate from the financial decision of normative interest with which to estimate risk preferences.

This is a key step in our normative evaluation of index insurance. Solely for normative reasons, we must assume that the risk aversion of the agent can be measured independently of the observed insurance choice. We can certainly debate the best way to measure risk aversion, but this basic methodological point is required *if we are to say that someone made an error and to be able to quantify it.*[21] We appreciate that this argument might seem odd to some economists, steeped in descriptive economics and the desire to test every assumption. We completely share that desire when it comes to descriptive economics; but not when it comes to normative economics. In that case the best we can do is measure the preferences of each individual independently and in a separate task, in the best way we know how, and evaluate their insurance decisions in terms of these "best estimates" of their preferences. Further details on this normative approach to insurance purchase behavior are provided by Harrison and Ng (2016) and Harrison (2019).

To see the significance of this normative position, consider the same individual and another choice in our battery. This choice provided the same actuarial parameters, but added a loading of just 8% to the premium, which becomes $13.50. In this instance the correct decision for this individual is to *not* purchase the insurance, generating an ECS of $0.98. In this instance what would we make normatively of a policy intervention that was motivated solely to increase take-up, with no regard for the precise risk aversion of the individual in the intervention? We would have to conclude that such an intervention simply had the wrong welfare proxy. If it had managed to increase take-up, with these changes in the actuarial parameters and risk preferences, it would have actually done harm in terms of individual welfare measured by ECS.

This standard logic has important implications. First, even if we assume an individual is an EUT decision-maker, we need to know *how risk averse* she is to say if her decision to "take-up" the product is the right one or not. The same point applies generally to the case in which she is an RDU decision maker. Second, we need to know which *type* of risk preferences best characterizes her. It is easy to find examples where we get the *sign* of the realized ECS wrong unless we know the type of decision-maker. Third, we see ECS numbers in dollars, reflecting the equivalent variation in income from old-fashioned welfare economics. We can therefore distinguish "small" welfare effects from "large" welfare effects.

---

[21] There are several concerns that one might have with this approach. First, if risk preferences are not stable over time, is there a risk of normative evaluation being based on "stale" preferences? We agree. Second, if risk preferences are elicited in one domain, how do we know that they are appropriate for another domain? In the (agreed) absence of perfect markets, making the two domains fungible, we also agree that this is an open question. We discuss these methodological issues in Harrison et al. (2020b).

3 Experiments

We conduct our experiments with 150 student subjects at the Experimental Economics Center (ExCEN) at Georgia State University. Subjects take part in an incentivized experiment where they make 54 choices to purchase index insurance or not, where actuarial parameters are varied. Before this task they complete a battery of 100 incentivized choices designed to allow us to infer their risk preferences so that we can calculate the consumer surplus from observed insurance choices. To better understand the manner in which subjects' understanding of the decision task effects their insurance decisions, subjects engage in several incentivized and hypothetical measures to assess their literacy, cognitive reflection, and fluid intelligence. Specifically, we use incentivized experiments where subjects' bias and confidence is elicited in 10 financial literacy and 10 index insurance literacy questions. Hypothetical survey questions are used to measure "cognitive reflection" and "fluid intelligence." We also use hypothetical general financial literacy questions as a benchmark for the incentivized financial literacy measure. Complete instructions and parameters for all tasks, as well as data and statistical code to replicate results, may be obtained from https://cear.gsu.edu/gwh.

3.1 Index insurance purchase decisions

Our main financial decision task is the index insurance experiment where subjects make 54 choices in which they receive an endowment that is at risk of a loss from a personal risk event. In each of the 54 decisions subjects can choose to purchase an index insurance contract or not, and at the end of the experiment one choice is randomly selected for payment. In each choice a random personal event determines losses, and a correlated random index event determines insurance claim payments if the subject chooses to purchase insurance. We consider an endowment of $60 for all choices. Loss amounts are either $39 or $30. Loss probabilities are either 0.1 or 0.2. Premium loadings on actuarially-fair premia are -50%, 0% or +8%. Finally, the correlation of the index event and the idiosyncratic loss event is 100%, 80%, 60%, 40%, 20% or 0%. The choice of a severely negative loading of -50% corresponded to our experience in the field: when subsidies are provided, they are not trivial. On the other hand, for the most widespread index insurance contracts positive loadings are modest, so we selected +8% in this case. The actuarially fair contract, with a 0% loading, is an obvious control. Our choice of a large negative loading also reflects our a priori concern to avoid low take-up, specifically zero take-up over all 54 choices by any given subject.

Before the subjects make these 54 insurance choices they receive basic instructions about the insurance. On the computer screens (see Fig. 1, which is also the decision screen for subjects of the worked example in Sect. 2.3) the probability of experiencing a loss in the personal event, and the probability of the personal event matching the outcome of the index, are presented separately to the subjects. The monetary payouts are also presented over all outcomes for each draw of the personal event and the index matching event for the decision to buy and not buy insurance.

There are several important components of the logic of this task and the interface. The first is the use of a matching probability between the personal loss and the index loss. One might have assumed that a simpler implementation would have been to specify a correlation of the two, and randomly generate a personal and index realization consistent with that correlation. The difficulty with this approach is that one would need many such realizations in order for the subject to "experience" the correlation, and this is a key actuarial parameter for this product. The method used allows us to induce specific values for the correlation, and indeed to vary it from choice to choice.

The second component of the task, and the interface, is the use of distinct colors for the personal event (red and blue) and for the index (green and black). These colors are used to illustrate the urns on the left, as well as to explain the payoffs on the right.

A final component of this task is the clear display of the two possible outcomes if the insurance is not purchased, and more significantly the four possible outcomes if the insurance is purchased. The four outcomes translate into only three distinct payoffs, but that redundancy is, we believe, valuable to fully convey the operation of the product.[22]

3.2 Elicitation of risk preferences

Before subjects participate in the index insurance purchase decisions, they participate in a risk elicitation task that allows us to characterise the risk preferences of the subjects. This characterisation allows us to predict the ECS from the optimal choice for each of the 54 insurance decisions, compare the subjects' actual decision to the optimal decision, and calculate the welfare gain or loss. Each subject was asked to make choices for 100 pairs of lotteries. Each subject faced a randomized sequence of choices from this battery of 100. The analysis of risk attitudes given these choices follows Harrison et al. (2008), and is undertaken at the level of the individual following Harrison and Ng (2016).

We always use the estimated utility function from the preferred RDU estimates for every subject when evaluating the ECS for each subject. The statistical reason, stressed by Monroe (2021), is that those subjects that are characterized as EUT by the test for "no probability weighting" still have standard errors around the probability weighting parameters, and potentially large ones.[23] And, perhaps surprisingly, these standard errors can make a substantive difference in precisely the normative

---

[22] The only other experimental interface focused on index insurance that we are aware of was used in artefactual field experiments in Peru and Kenya by Carter et al. (2008). Their design and interface was deliberately structured to mimic the field setting it was applied in, as a literacy treatment, whereas ours is deliberately structured to be more abstract, to allow evaluation of theoretical propositions. Each emphasis has a valid, and complementary, inferential role to play, as stressed by Harrison and List (2004). They do not report results of the use of their literacy intervention.

[23] Indeed, larger standard errors makes it easier to reject the null hypothesis.

evaluations undertaken here.[24] Hence there is no formal need to differentiate EUT and RDU decision makers for these calculations, because EUT is nested within RDU.

3.3 Hypothetical literacy measures

Subjects first undertake the RAPM fluid intelligence; the *CRT* survey questions; and finally 9 survey questions about general financial literacy, spanning concepts including interest, inflation, stock returns, bond prices, risk diversification, and mortgages.

3.4 Incentivized financial and index insurance literacy

Subjects participate in two *incentivized task* tasks that assesses subjective beliefs about their own answers to a set of standard financial literacy questions (*Financial Literacy Beliefs*) due to Lusardi and Mitchell (2014) and a set of ten insurance decisions randomly selected from the 54 insurance choices (*Index Insurance Beliefs*). We elicit these beliefs using an incentivized Quadratic Scoring Rule (QSR) for payment: for each question subjects' responses are elicited over a continuous range of possible answers presented in terms of ten intervals or 'bins,' where one bin represents the correct answer. For each set of incentivized questions, financial literacy questions or insurance decision questions, one question is selected for payment.

Figure 2 is an example of the display response screen, and illustrates an example using one of the financial literacy questions the subject received. The question reads *"Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow?"*

The participant has 10 sliders to adjust, shown at the bottom of the screen, and has 100 tokens to allocate across the sliders. Each slider corresponds to a bin labeled first as $102, and then up to $120 in $2 increments. Each slider allows the subject to allocate tokens that reflects their belief about the answer to the question displayed at the top. They must allocate all 100 tokens, and in this example they start with 0 tokens allocated to each slider. As they allocate tokens, by adjusting sliders, the payoffs displayed on the screen change as illustrated in Fig. 2.

The earnings of each participant are based on the payoffs, which are generated by a discrete version of a QSR developed by Matheson and Winkler (1976) for eliciting beliefs about non-binary events. The QSR is applied to a participant's token allocation and displayed in real time by the software as they allocate all 100 tokens across the bins. A participant is paid the displayed amount above an interval if and only if that interval contains the true answer. Using Fig. 2 as an example, in this instance a subject believes that they are fairly confident they know where the true answer lies to the question. They assign zero tokens, thus zero probability, that the true answer

---

[24] Monroe (2021) shows that they can make a difference to the analyses of full indemnity insurance contracts by Harrison and Ng (2016), and our approach follows theirs.
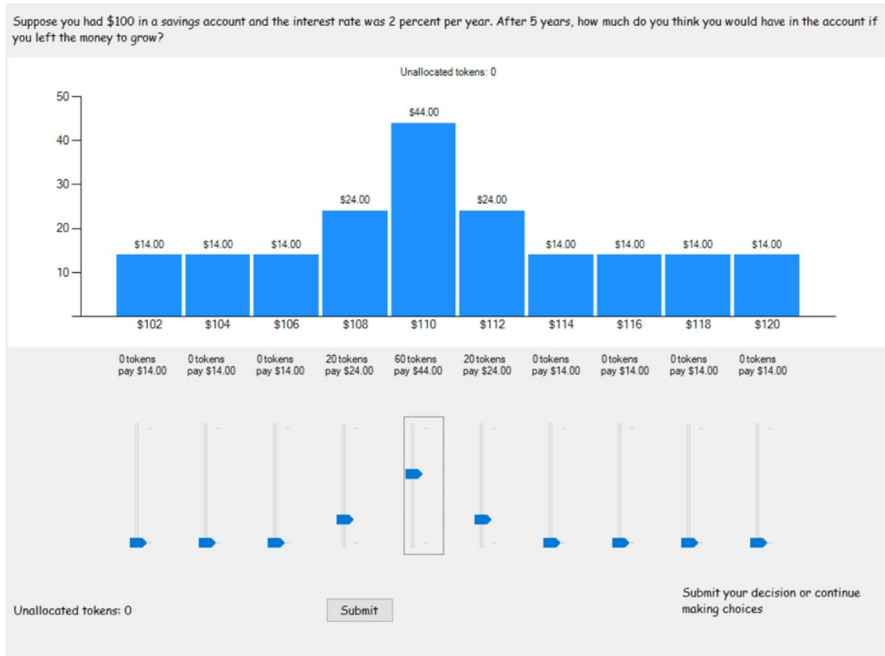
Fig. 2 Example screen of incentivized financial literacy question

lies in bins 1, 2, and 3, or in bins 7, 8, 9, and 10. But they allocate 20 tokens to bin 4, 60 tokens to bin 5, and 20 tokens to bin 6.

Since the correct answer to this question is $110.41 and lies in bin 5, their payment would have been $44 dollars out of the maximum $50 they could have earned. The instructions explained that the subject could earn up to $50 dollars, but only by allocating all 100 tokens to one interval *and* when that interval contains the true answer.

Subjects were rewarded for one of these belief elicitation tasks, with the task selected at random by the subject's rolling of a die. The question they picked was called back up on the display, then the correct answer revealed, and a participant's earnings recorded. It is therefore up to the participant to balance the strength of their personal beliefs with the possibility of them being wrong. Their subjective belief about the correct answer to each question is a judgment that depends on the information they have about the topic of the question. The subject is also told that their choices may depend on their willingness to take risks.

For the index insurance literacy experiment the beliefs interface is the same as in the previous task. It asks the participants, however, about their bias and confidence in their answers to ten questions about potential outcomes, initial stakes, and personal and index event probabilities. Recall from Fig. 1 that the insurance task display has information on the initial stake at risk, the amount at risk for personal loss, the premium that insurance can be purchased to insure against the personal loss, the probability of a bad personal event, the probability that the index matches, as well as

the possible outcomes calculated if insurance was purchased or not over the various states of the personal event and the index matching.

To answer the 10 insurance literacy questions, subjects are handed reference materials containing five figures labeled Figure A through Figure E, with 2 questions relating to each of the figures. Figure 3 is an example of reference material
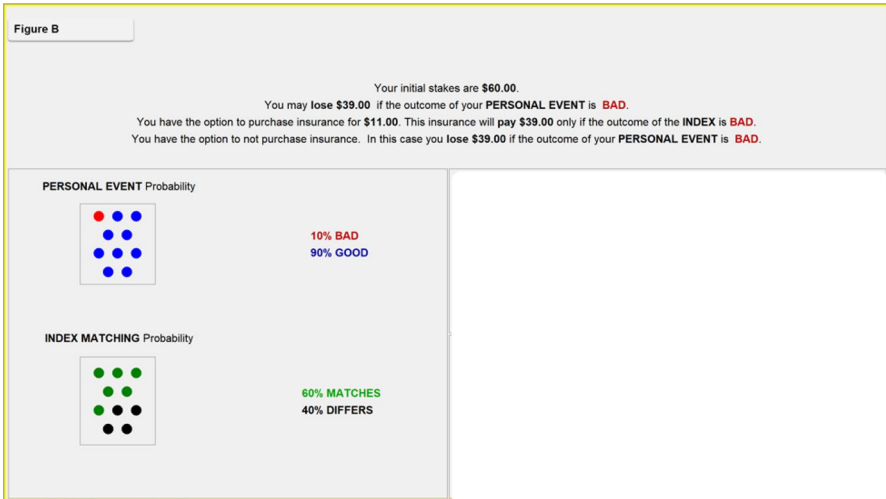

Fig. 3 Example reference material for index insurance literacy question
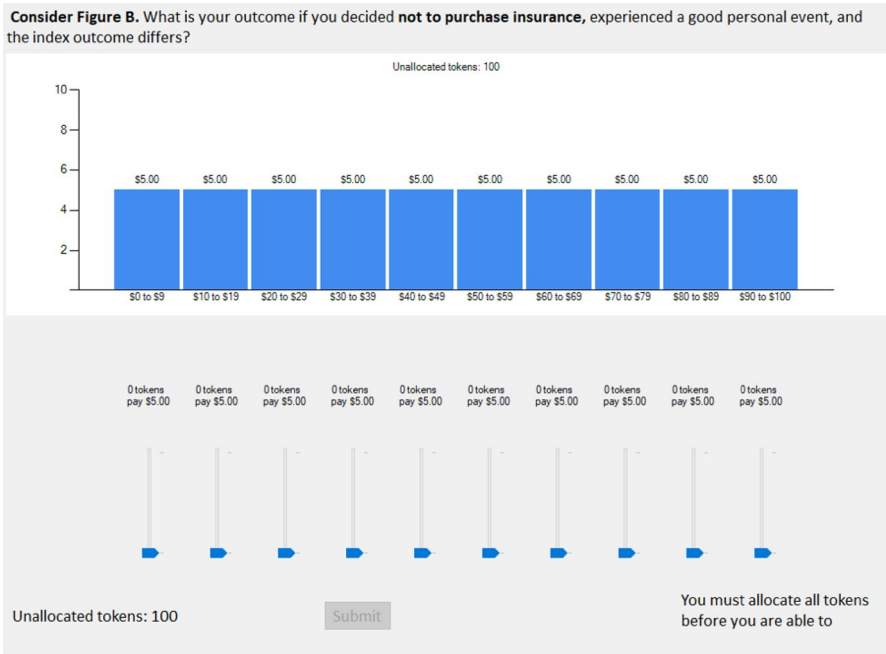

Fig. 4 Example screen for index insurance literacy question

provided to participants relating to the insurance literacy task: "Figure B" appears in the top-left corner. By contrast, note that the information that was displayed in the lower-right panel of Fig. 1 is now omitted, so the participant will have to calculate the outcome for each insurance question.

Figure 4 is an example of an insurance literacy question asked of the subject. It reads: *Consider Figure B. What is your outcome if you decided not to purchase insurance, experienced a good personal event, and the index outcome differs?* Here the subject refers to Figure B in the reference materials in order to try to calculate the payoff, and then place some bets on their beliefs about the answers to the question being asked.

Working through the arithmetic, and using the reference material, we see that the initial stake at risk is $60. According to the question, if we decided not to purchase insurance and experienced a good personal event, then we do not pay the premium and would receive $60 irrespective of the index matching or differing from the personal event. The bin containing the correct response, $60, is bin 7 in Fig. 4. Subjects could earn up to $10 in this task, by being 100% confident about the answer and placing all 100 tokens in the correct bin.

We measure literacy from these responses by calculating the Financial Literacy Index Score and the Index Insurance Literacy Score. In each case we calculate a measure between 0 and 1. This measure, denoted L, is defined as the fraction of the raw token allocation that is placed into the true bin, which is equal to the number of tokens allocated to the bin with the true answer divided by 100. If all tokens were allocated to the correct bin, then $L = 1.0 (= 100/100)$. If 35 tokens were allocated to the correct bin, then $L = 0.35 (= 35/100)$. This constructed L is data that we use directly when undertaking estimation (i.e., it can be used as a covariate since it is data and not a random variable). In addition the literacy measure L reflects the joint effects of bias *and* confidence, albeit in the simplest possible way.

4 Results

4.1 Descriptives

Figure 5 displays the distributions of the various measures of literacy, fluid intelligence and cognitive reflection. Panels A and C are the measures here that reflect incentivized responses to belief elicitation questions. One of the issues we are interested in is the comparative reliability of these incentivized responses and hypothetical survey responses that are widely used to assess decision-making quality.[25] Comparing the incentivized and hypothetical financial literacy scores we observe a much tighter distribution of scores in the incentivized measure (panel A), and a virtually uniform distribution of scores in the hypothetical measure (panel B). This is, on its

---

[25] Of course, we are less interested in whether the hypothetical and incentivized responses provide the same scores, so much as whether they are correlated similarly with the outcomes of interest.
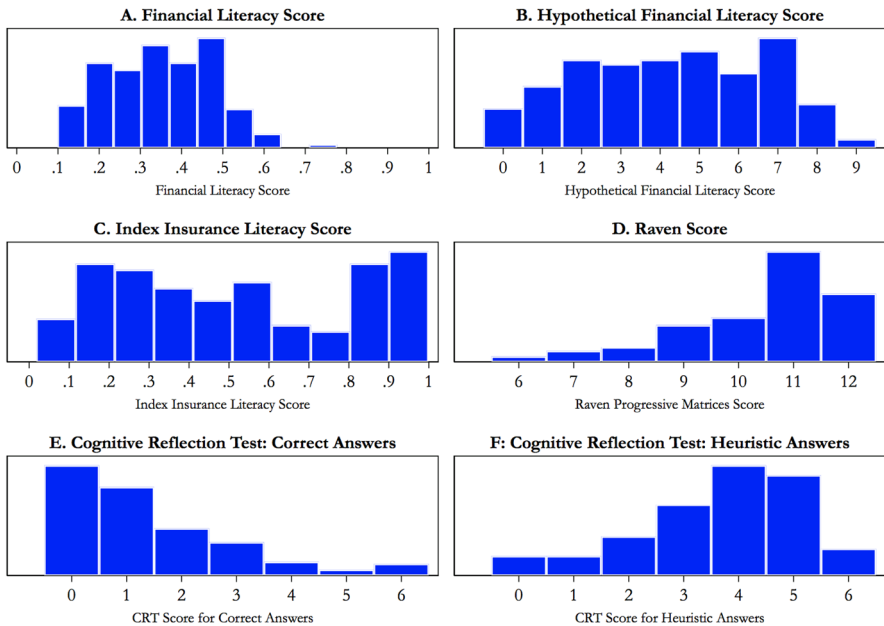
Fig. 5 Distributions of the literacy and cognitive reflection variables. *Note* For the Index Insurance Literacy Score distribution we have 112 subjects and for the distributions for the other variables we have 150 subjects. "Financial Literacy Score" is a score from 0 to 1 and is the average of the fraction of tokens out of a 100 allocated to the bin with the true answer over all 10 financial literacy questions. "Hypothetical Financial Literacy Score" is a count of the number of questions out of nine where the subject gave the correct answer to the financial literacy survey question. "Index Insurance Literacy Score is a score from 0 to 1 and is the average of the fraction of tokens out of a 100 allocated to the bin with the true answer over all 10 index insurance literacy questions. "Raven Progressive Matrices Score" is a score from 0 to 12 and is the count of the number of questions where the subject gave the correct answer. "CRT Score for Correct Answers" is a count of the number of questions out of six where the subject gave the correct answer to the CRT question. "CRT Score for Heuristic Answers" is a count of the number of questions out of six where the subject gave the heuristic answer to the CRT question

face, suggestive of a familiar hypothetical bias in experimental economics, where incentives lower the noise in responses.

The Index Insurance Literacy Score in panel C of Fig. 5 displays a much wider distribution than the general Financial Literacy Score in panel A. The latter is intended as a general financial literacy measure, and the former as a tightly domain-specific literacy measure. These differences tell us three things. First, that we have much more heterogeneity across the sample in their domain-specific literacy. Second, that the average domain-specific literacy is higher than the average general financial literacy. And third, that the domain-specific literacy tends to be bimodal, with one mode exhibiting very high literacy and another mode exhibiting very low literacy. The general financial literacy scores, on the other hand, are unimodal. Hence these comparisons already tell us that we might expect to see different insights from the effects of the two literacy measures, and that this would be informative as to the type of literacy, general *versus* domain-specific, that matters.

The Raven scores in panel D of Fig. 5 are generally very high, which is what one would expect *a priori* from a sample from a population that has selected into a university education. Hence we might not see much inferential power from these scores, with most subjects bunched towards the top possible scores.

The CRT scores in panel E of Fig. 5 display a familiar pattern from previous research, where the "false lure" of the heuristic response dominates. The comparison of panels E and F display a complementarity that tells us that indeed the low scores in panel E are due to "high" scores in panel F, rather than just random responses. Hence we would expect either the CRT Correct Score or CRT Heuristic Score to pick up the same patterns, albeit with opposite sign.

Figure 6 displays the realized Efficiency from all of the 54 decisions that all of the 150 subjects made. The measure for Efficiency ranges from 0 to 1, which represents the fraction of the total ECS that the individual realizes over all decision compared to the total ECS that the same individual would have realized over all decisions if all decisions were correct. The red dashed lines present the 25th, 50th and 75th centiles. We see from Fig. 6 that, over all 150 subjects, there were many inefficient decisions. From the median Efficiency, we find that 50% of subjects achieved less that 50% Efficiency over all 54 choices. We observe two modes, around 35% and 55% Efficiency, suggesting that there are at least two types of decision-makers in terms of the quality of decisions. Very few individuals had more than 65% Efficiency, highlighting the potential welfare benefits of better decision-making for every individual. To focus on essentials, in Fig. 6 we separate the individuals into those individuals that did Great in terms of their welfare outcomes, generating an Efficiency gain in excess of the 75th centile of individuals, and those that did Terrible in terms of their welfare outcomes, generating an Efficiency loss below the 25th centile. This separation helps us avoid comparing lots of decisions that generated de minimus Efficiency gains and losses around zero: by definition these decisions do not matter as much for
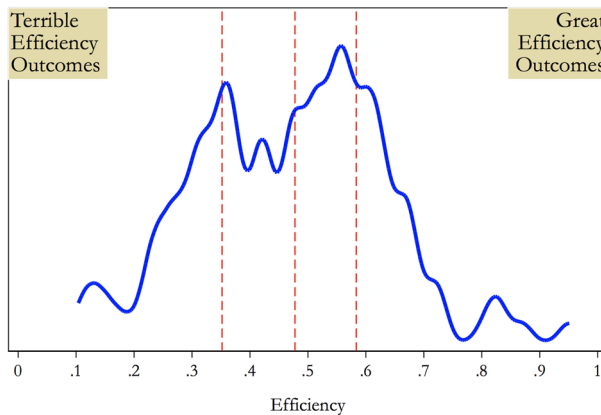

Fig. 6 The quality of insurance decisions. *Note* Kernel density of the realized Efficiency for each of the 150 subjects making 54 choices. Dashed red vertical lines show the 25th, 50th and 75th centiles. We classify terrible outcomes as those individuals that realized a level of Efficiency below the 25th centile ($N = 37$ individuals) and great outcomes as those individuals that realized a level of Efficiency above the 75th centile ($N = 45$ individuals)

welfare, and are more likely to be driven by noise. Our formal statistical analyses of these Efficiency results in Sect. 4.2 uses data from all subjects, even if we highlight the tails for ease of interpretation.

We stress, again and for good cause, that Fig. 6 provides the first sighting of the right policy target for regulators and policy-makers. To help implement those policies, economists should be intensely interested in what factors generate the tails of this distribution. There is no a priori reason to think that the same factors that determine the terrible decisions are the same, with different values, as the factors that determine the great decisions. For example, those making terrible decisions might be those that are lured *solely* by the use of inappropriate heuristics, and hence be identified by the CRT Heuristic Score. On the other hand, those making great decisions are likely those that are not lured by inappropriate heuristics, but may also need to have some high level of domain-specific insurance literacy.

We do have one very useful aggregate statistic that can help guide intuition here: average purchase decisions. For all subjects the purchase decision was made 56% of the time. But individuals with terrible welfare outcomes made a purchase decision 71% of the time, and individuals with great welfare outcomes made a purchase decision only 46% of the time. So the stylized fact is that the terrible welfare outcomes tended to reflect excessive take-up, and the great welfare outcomes tended to reflect caution when deciding to purchase the product. Recalling the loading of the products, the suggestion is that the contracts with a negative loading of 50% below the actuarially fair premium would be the ones that should have been taken up by most subjects, and that the contracts with a positive loading of 8% above the actuarially fair premium would be the ones to scrutinize more carefully.[26]

Figure 7 displays the same distributions shown earlier in Fig. 5, but stratifying according to whether the welfare outcomes were terrible or great in terms of Efficiency. For Financial Literacy in panel A we observe that those that made great decisions had slightly higher scores than those that made terrible decisions, but the two were not significantly different. For Index Insurance Literacy in panel C those that made great decisions were, indeed, more heavily represented in the upper modes, as conjectured above. For those making terrible decisions, however, we see a strong bimodal distribution, where the majority have a low score and a large share of individuals with terrible outcomes actually have the highest index insurance literacy score. When we look at this group of individuals more carefully they also score highest on the general financial literacy score and the Raven score for fluid intelligence, suggesting that, even though they seem to understand the decision-context perfectly, some other decision-process appears to drive their decisions. The mean purchase rate amongst this group of individuals is 84%. With respect to Raven scores of fluid intelligence a larger share of individuals having great welfare outcomes seem to have a high score compared to those with terrible outcomes.

---

[26] A reminder again that for index insurance products the familiar *Insurance Economics 101* theorem, that any risk averse EUT agent should purchase any full indemnity product for actuarially fair premiums, or lower, does not apply: see Clarke (2016).

4.2 Statistical analysis

The descriptive insights from our data can be evaluated more carefully with conditional regression analyses. Our focus is on the determinants of welfare, stratified by whether outcomes were great or terrible in terms of the Efficiency they led to.

The core statistical model we use is an ordered probit regression of Efficiency where the lowest category consists of individuals with terrible welfare outcomes (below the 25th percentile, $N = 37$), the middle category consists of individuals with welfare outcomes around zero (between the 25th and the 75th percentile, $N = 68$) and the highest category consists of individual with great welfare outcomes (above the 75th percentile, $N = 45$). This model allows a test of the hypothesis that the decision-making process for those with terrible and great welfare outcomes is different, as suggested by our descriptives. Even though we do not focus on the middle category because their average welfare outcomes are approximately zero, and more likely driven by random noise, the ordered probit model allows us to evaluate the upper and lower tails of the welfare distribution as part of a statistical analysis of the full sample.[27] We cluster standard errors at the session level, to correct for potential session-level heteroskedasticity. In each regression we control for a long list of demographics, and focus on the predicted marginal effects of the covariates of interest for each category of the outcome variable.[28]

Figure 8 presents the average marginal effects and the 95% confidence intervals of the ordered probit regressions of Efficiency on our literacy variables, for the upper category (great decisions) and lower category (terrible decisions). For ease of comparison the literacy variables have been normalized to fall in the interval between 0 and 1. The effect here, shown on the horizontal axis, is on the predicted probability of realizing welfare outcomes from the 54 insurance decisions that are great or terrible relative to the other individuals.

---

[27] Harrison et al. (2020a, Sect. 4.4) evaluate ECS for index insurance decisions using panel regressions with individual random effects. Their experiment is comparable to ours, but used different subjects drawn from the same population and different actuarial parameters. They randomized roughly half of their subjects into a treatment in which the effect of applying ROCL to the index insurance contract was literally displayed to subjects at the time of the choice. By pooling the data from this treatment with their baseline, they were able to show statistically that violations of ROCL played an important role in explaining the observed variation in ECS from choice to choice. These conclusions also controlled for the choice-specific actuarial characteristics, as well as for subject specific demographic characteristics. Their analysis also provided support for the broad classification of *individual choice decisions* into the upper and lower tails of the ECS distribution, as well as the classification of *individuals* into the extreme tails of the Efficiency distribution. We build on these insights, and simplify the statistical analysis, by using the ordered probit model.

[28] The demographics we control for are the age of the respondent, the number of household members living with them, the amount of money in USD the respondent typically spends each day in cash or via debit card, and binary indicators of whether the respondent is female, whether the respondent expects to complete a bachelor (versus higher than bachelor), whether the respondent is black, owns a business, is single, has a full-time or part-time job, has a high or very high income (versus low or very low), whether the parents of the respondent have a high or very high income (versus low or very low), whether the respondent is Christian, whether the respondent is "Junior," "Senior" or "Postbaccalaureate" as compared to "Freshman" or "Sophomore," and whether the respondent has a self-reported "high GPA" of 3.25 or higher.

Fig. 7 Distributions of the literacy variables, for welfare outcomes that are terrible or great. *Note* We classify terrible outcomes (red bars) as those that led to a realized Efficiency below the 25th centile and great outcomes (green bars) as those that led to a realized Efficiency above the 75th) centile. "Financial Literacy Score" is a score from 0 to 1 and is the average of the fraction of tokens out of a 100 allocated to the bin with the true answer over all 10 financial literacy questions. "Hypothetical Financial Literacy Score" is a count of the number of questions out of nine where the subject gave the correct answer to the financial literacy survey question. "Index Insurance Literacy Score" is a score from 0 to 1 and is the average of the fraction of tokens out of a 100 allocated to the bin with the true answer over all 10 index insurance literacy questions. "Raven Progressive Matrices Score" is a score from 0 to 12 and is the count of the number of questions where the subject gave the correct answer. "CRT Score for Correct Answers" is a count of the number of questions out of six where the subject gave the correct answer to the CRT question. "CRT Score for Heuristic Answers" is a count of the number of questions out of six where the subject gave the heuristic answer to the CRT question

We find that, for those who make terrible decisions in terms of welfare, an increase in the incentivized domain-specific index insurance literacy score is associated with a decreased likelihood of a terrible welfare outcome (9 percentage points, *p*-value 0.049), but an increased likelihood of a great welfare outcome (10 percentage points, *p*-value 0.015). We observe a similar pattern for the Raven measure of fluid intelligence, significant at the 5% level, with effect sizes of 18 to 20 percentage points. The effect for incentivized financial literacy is very imprecisely estimated. None of the other hypothetical measures (the CRT heuristics score, the CRT correct score, and the hypothetical financial literacy) have an effect on the quality of welfare outcomes from the insurance decisions.

Figure 9 presents the point estimates of the predicted marginal effects along with their 95% confidence intervals from two separate panel probit regressions of index insurance purchase on the literacy variables, recognizing that each subject
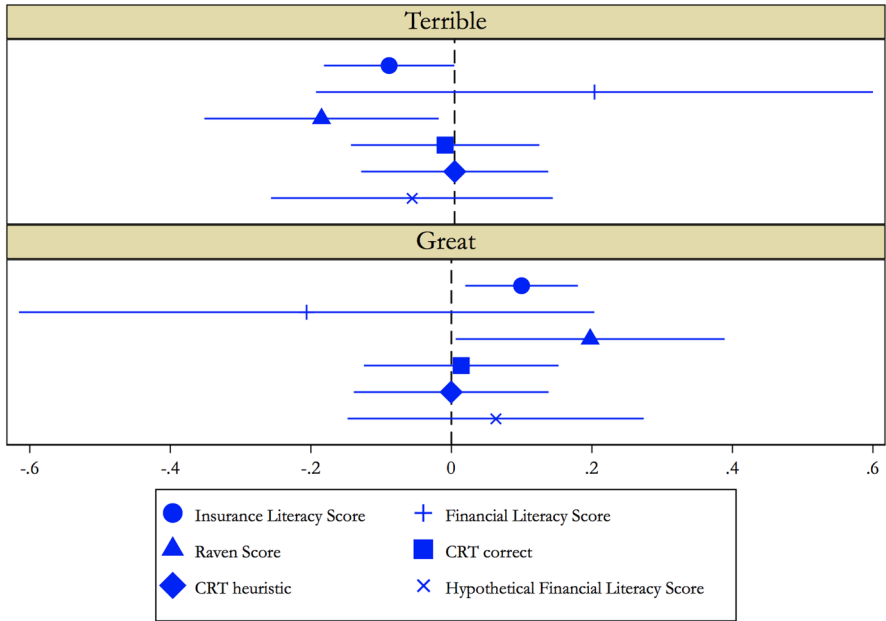
Fig. 8 Effects of Literacy on Likelihood of Terrible or Great Efficiency Outcomes *Ordered Probit Regressions on Full Sample*. *Note* For the regressions with the Index Insurance Literacy Score we have 112 subjects and for the regressions for the other variables we have 150 subjects. Average marginal effects, and their 95% confidence intervals, of literacy covariates in an ordered probit regression of Efficiency where the lowest category consists of individuals with terrible welfare outcomes (below the 25th percentile), the middle category consists of individuals with welfare outcomes around zero (between the 25th and the 75th percentile) and the highest category consists of individual with great welfare outcomes (above the 75th percentile). The welfare outcomes, termed Efficiency, are calculated based on 54 index insurance purchase decisions per individual by calculating the percentage of the total ECS that the individual realizes over all decision compared to the total ECS that the same individual would have realized over all decisions if all decisions were correct. Demographic controls are the age of the respondent, the number of household members living with them, the amount of money in USD the respondent typically spends each day in cash or via debit card, and binary indicators of whether the respondent is female, whether the respondent expects to complete a bachelor (versus higher than bachelor), whether the respondent is black, owns a business, is single, has a full-time or part-time job, has a high or very high income (versus low or very low), whether the parents of the respondent have a high or very high income (versus low or very low), whether the respondent is Christian, whether the respondent is "Junior," "Senior" or "Postbaccalaureate" as compared to "Freshman" or "Sophomore," and whether the respondent has a self-reported "high GPA" of 3.25 or higher

contributed 54 purchase decisions. One regression model considers the purchase decisions that led to terrible welfare outcomes, and the other regression model considers the purchase decisions that led to great welfare outcomes. The effect here, shown on the horizontal axis, is on the predicted probability of deciding to purchase the insurance product.

For those that make terrible decisions, predominantly by excess take-up, we observe that greater domain-specific insurance literacy is significantly associated with *lower* take-up. None of the other literacy variables, however, influence
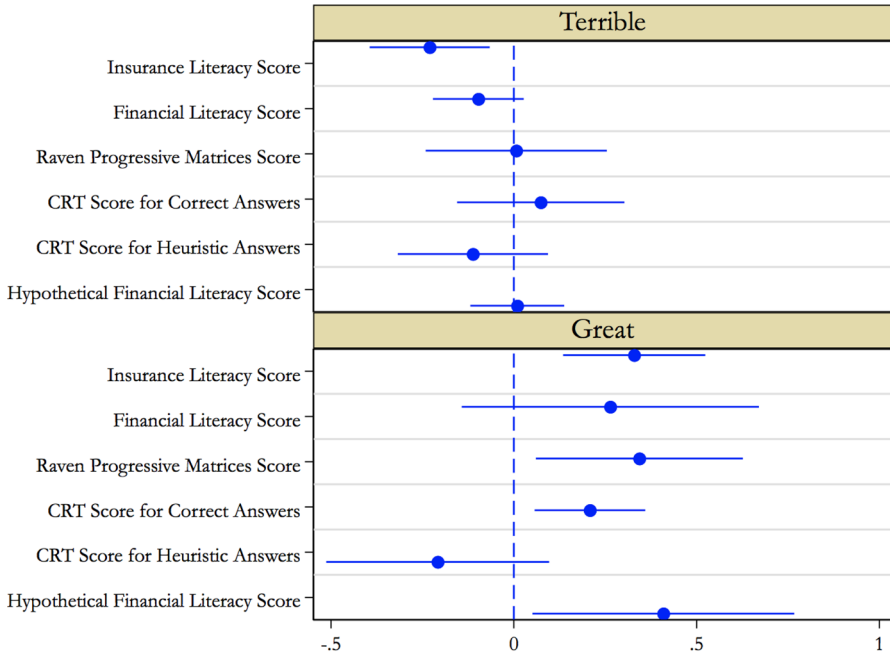
Fig. 9 Effects of Literacy on Likelihood of Purchase for Terrible and Great Efficiency Outcomes *Panel Probit Regressions on Sub-Samples. Note* Average marginal effects, and their 95% confidence intervals, of covariates of a panel probit regression of the purchase decision for a subsample of individuals who had terrible outcomes in terms of welfare (below the 25th percentile) and great outcomes in terms of welfare (above the 75th percentile). Each of the subjects made 54 purchase decisions. For the Insurance Literacy Score that means we have 1350 decisions and for the other variables we have 1998 decisions for the individuals with terrible outcomes. For the Insurance Literacy Score that means we have 1512 decisions and for the other variables we have 1998 decisions for the individuals with great outcomes. We cluster standard errors at the session level, to correct for potential session-level heteroskedasticity. Demographic controls are the age of the respondent, the number of household members living with them, the amount of money in USD the respondent typically spends each day in cash or via debit card, and binary indicators of whether the respondent is female, whether the respondent expects to complete a bachelor (versus higher than bachelor), whether the respondent is black, owns a business, is single, has a full-time or part-time job, has a high or very high income (versus low or very low), whether the parents of the respondent have a high or very high income (versus low or very low), whether the respondent is Christian, whether the respondent is "Junior," "Senior" or "Postbaccalaureate" as compared to "Freshman" or "Sophomore," and whether the respondent has a self-reported "high GPA" of 3.25 or higher

decisions to purchase or not purchase in the group of individuals who make terrible decisions. On the contrary though, for those that realise great welfare outcomes, predominantly by cautious take-up, five literacy variables show substantial positive effects on take-up, and four out of these five are significant at the 5% level. As expected, the CRT heuristics score has a negative effect, albeit not significant.

On the face of it the opposing effect on take-up of the Insurance literacy score might seem paradoxical, since "literacy" is pushing in different directions with respect to take-up. But "great outcomes" in our design derive from knowing the lyrics of a well-known song about strategy in poker: *You've got to know when to hold*

'em, Know when to fold 'em Know when to walk away, And know when to run.[29] In this context, the insurance contracts with a massive negative loading of 50% are the ones that the subject needs to know to hold (i.e., purchase), and the contracts with the modest positive loading of 8% are the ones that the subject needs to know to fold on (i.e., decline to purchase). So better *domain-specific* Insurance Literacy seems to help both.

5 Conclusion

The potential for index insurance to meet risk management needs is clear, not least from the mitigation of adverse selection and moral hazard problems. It also allows for the development of "predictive insurance," where an index is correlated with some higher risk of a natural disaster in the future, such as global warming or pandemic mortality. In principle, settlement can occur prior to any disaster, allowing mitigation of the personal loss and rapid response (Clarke and Dercon 2016). Quite apart from the risks of natural disasters on a major scale, predictive index products could be applied in innovative ways to manage health risks. One of the most significant risks facing the very poor in developing countries is the out-of-pocket cost and opportunity cost (of foregone employment) of health problems (Collins et al. 2009), and early settlements could make a major difference in that domain as well.

However, there are challenges marketing these products, and in some regulatory quarters they are viewed with the same suspicion as other exotic financial derivatives. Our approach is to use the tools of behavioral economics to rigorously measure the extent to which "understanding" of the risk management choice drives purchase decisions and expected consumer welfare from insurance. Our starting point was to understand how terrible welfare outcomes and great welfare outcomes are created, in a controlled setting in which we vary the actuarial variables and measure "understanding."

We find that our incentivized domain-specific literacy measure plays an important role in determining great welfare outcomes, as well as the hypothetical Raven score. General financial literacy and the other hypothetical survey measures of understanding do not influence the extent to which welfare outcomes are great or terrible. Our analysis also suggests that terrible welfare outcomes are realized through excess purchase, while great welfare outcomes seem to be the result of cautious purchase. We observe an asymmetric pattern in terms of the predictors of purchase for great and terrible outcomes. For great welfare outcomes an increase in five literacy measures seems to lead to welfare-enhancing purchase decisions, while no general relationship exists for individuals with terrible outcomes.[30] This pattern suggest that there

---

[29] The song is *The Gambler*, written by Don Shlitz, recorded by Johnny Cash in 1978, and then made famous by the 1978 recording by Kenny Rogers. See https://en.wikipedia.org/wiki/The_Gambler_(song) for the history and lyrics.

[30] The only measure of understanding that has a significant effect on purchase for individuals with terrible outcomes is the Insurance Literacy Score, reducing the likelihood of purchase by 25 percentage points.

are two distinct decision-making processes driving the decisions in both tails of the distribution.

Our central contribution is to demonstrate how one can rigorously quantify the Expected Consumer Surplus (ECS) of index insurance purchase decisions, and the effects that literacy has on that measure of individual welfare. To operationalize the theory and experimental design, we used the the "wind tunnel" of a lab setting, prior to costly implementation in a field setting. To measure ECS we need to develop methods to estimate individual risk preferences and apply them to infer ECS given the observed insurance purchase decisions of each of our subjects. We cannot rely on insurance take-up as a qualitative measure of ECS, since that biases inferences to only ever show ex ante welfare improving decisions. In order to measure literacy we needed to differentiate general financial literacy from domain-specific literacy with respect to the specific index insurance contract on offer. We also needed to elicit incentivized measures of literacy that tell us the confidence of the understanding that the subject has about these general financial concepts and then the index insurance contract. Many observers of field behavior with respect to index insurance have worried about how much individuals actually "understand" about insurance and the specific product. Our lab experiments, and analysis of results, demonstrate that it is feasible to develop an experimental design to measure these effects of literacy on the ex ante welfare of index insurance choices.

As explained in our introduction we deliberately start with a laboratory experiment to identify, in a cost-effective way, the measures of comprehension that can be used in the field, to develop our welfare measures, and to observe those that do or do not purchase the product. We should be clear that there is a continuum between lab and field experiments, and a complementarity. This was stressed by Harrison and List (2004) for economics, but has been well-known in medicine. In a primer on efficacy trials and effectiveness trials for clinicians, Singal et al. (2014) noted that:

> Intervention studies can be placed on a continuum, with a progression from efficacy trials to effectiveness trials. Efficacy can be defined as the performance of an intervention under ideal and controlled circumstances, whereas effectiveness refers to its performance under 'real-world' conditions. However, the distinction between the two types of trial is a continuum rather than a dichotomy, as it is likely impossible to perform a pure efficacy study or pure effectiveness study.

The same applies in economics. And in the testing of drugs for humans, such as COVID-19 vaccines, we would never contemplate an effectiveness trial with first conducting rigorous efficacy trials.[31]

To apply this design in the field similar experiments need to be conducted, and adapted to the local context. For example, in field experiments in Ethiopia we will be using video instructions in the local language; familiar pictorial representations of key concepts such as risky lotteries, such as balls from a bag; using pictures of local currency rather than numbers; using images, such as healthy, sickly or dead

---

[31] The Russian *Sputnik V* vaccine is an exception, hopefully without tragic consequences.

cows, that provide rich field referents for our subjects to consider insuring; and so on.[32] Virtually every field experiment requires rich adaptation to the local decision context (Harrison and List 2004), and ours is no exception. Before we undertake those changes, however, we needed to know that the basic tasks would be able to evaluate the theoretically-motivated hypotheses we have. If they could not do that in the "clean beaker" of a laboratory experiment, we see no reason to believe that they would magically work in the field. We do not claim that inferences from the lab experiments generalize beyond the population from which the sample is drawn. Hence we must go to the field as well as the lab.

We appreciate that many researchers only care, these days, about field behavior. We only care about field behavior that can be rigorously linked to theory, and in our case that means *structural* theory. Specifically, structural models of individual risk preferences, structural models of how those risk preferences inform the ECS of insurance choices, and structural models of the way in which scoring rules elicit subjective belief distributions to measure literacy and understanding of the insurance product. In order to generate field behavior that can be linked to theory, we must know how to link behavior to theory, and it is efficient in terms of time, cost, and the patience of subjects to do that first in the laboratory.

Our approach takes a position on how one judges good and bad decisions, and there are other approaches that policy makers should be aware of.[33] Our approach repeatedly[34] comes to a major conclusion: that blindly encouraging take-up of insurance is an outright dangerous thing to do if individual welfare improvement is the goal. Blind watchmakers[35] typically end up making a *lot* of terrible watches before they produce *one* great watch.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1057/s10713-020-00060-1.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

---

[32] Gao et al. (2020) provide another example of using lab data to improve inferences in the field. They show that Bayesian hierarchical models of individual risk preferences do not require that every subject be asked to make a choice for every one of the full battery of risky lottery choices. In our case, we asked every subject to make 100 lottery choices. In the field, one is always concerned with the amount of time needed for subjects to make many choices, particularly if those choices are just part of a larger experimental session. Gao et al. (2020, Sect. 3.A) show how one can dramatically reduce this number for each subject to something like 20 or 40 choices, selected at random from the full battery, and draw statistical inferences about ECS that are comparable to the inferences drawn when every subject made all possible choices. We plan to use this valuable insight in the field, to save time and subject patience.

[33] See Harrison (2019, Sect. 6) for a full list, in the context of the behavioral welfare economics of insurance.

[34] See Harrison and Ng (2016, 2019).

[35] Dawkins (1986).

References

Aina, I. V., O. E. Ayinde, D. R. Thiam, E. Muchandondwa, and J. M. Miranda. 2018. Willingness to Pay for Index-Based Livestock Insurance: Perspectives from West Africa. Working Paper. Department of Agricultural Economics and Farm Management, University of Ilorin, Nigeria.

Baron, J. 2019. Actively Open-minded Thinking in Politics. *Cognition* 188: 8–18.

Brown, J., A. Kapteyn, E.F. Luttmer, and O.S. Mitchell. 2017. Cognitive Constrains on Valuing Annuities. *Journal of the European Economic Association* 15 (2): 429–462.

Camerer, C.F., and R.M. Hogarth. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-labor-production Framework. *Journal of Risk and Uncertainty* 19 (1–3): 7–42.

Carter, M. R., Barrett, C. B., Boucher, S., Chantarat, S., Galarza, F., McPeak, J., Mude, A. G., and Trivelli, C. (2008). Insuring the Never Before Insured: Explaining Index Insurance through Financial Education Games. BASIS Briefs BB2008-07, Department of Agricultural and Applied Economics, University of Wisconsin.

Chantarat, S. and Mude, A. G. (2010). Willingness to Pay for Index Based Livestock Insurance: Results from a Field Experiment in Northern Kenya. Working Paper. Cornell University.

Chantarat, S., A.G. Mude, C.B. Barrett, and M.R. Carter. 2013. Designing Index-based Livestock Insurance for Managing Asset Risk in Northern Kenya. *Journal of Risk and Insurance* 80 (1): 205–237.

Clarke, D.J. 2016. A Theory of Rational Demand for Index Insurance. *American Economic Journal: Microeconomics* 8 (1): 283–306.

Clarke, D.J., and S. Dercon. 2016. *Dull Disasters? How Planning Ahead Will Make a Difference*. New York: Oxford University Press.

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

Collins, D., J. Morduch, S. Rutherford, and O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton: Princeton University Press.

Dawkins, R. 1986. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. New York: WW Norton & Company.

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

Di Girolamo, A., G.W. Harrison, M.I. Lau, and J.T. Swarthout. 2015. Subjective Belief Distributions and the Characterization of Economic Literacy. *Journal of Behavioral and Experimental Economics* 59: 1–12.

Frederick, S. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19 (4): 25–42.

Gao, X. S., G. W. Harrison, and R. Tchernis. 2020. Behavioural Welfare Economics and Risk Preferences: A Bayesian Approach. NBER Working Paper 27685. National Bureau of Economic Research.

Gaurav, S., S. Cole, and J. Tobacman. 2011. Marketing Complex Financial Products in Emerging Markets: Evidence from Rainfall Insurance in India. *Journal of Marketing Research* 48 (SPL): S150–S162.

Harrison, G.W. 2006. Experimental Evidence on Alternative Environmental Valuation Methods. *Environmental and Resource Economics* 34: 125–162.

Harrison, G.W. 2014. Real Choices and Hypothetical Choices. In *Handbook of Choice Modelling*, ed. S. Hess, and A. Daly. Northampton, MA: Edward Elgar.

Harrison, G.W. 2019. The Behavioral Welfare Economics of Insurance. *The Geneva Risk and Insurance Review* 44 (2): 137–175.

Harrison, G.W., and J.A. List. 2004. Field Experiments. *Journal of Economic Literature* 42 (4): 1009–1055.

Harrison, G. W., J. Martínez-Correa, K. Morsink, J. M. Ng, and J. T. Swarthout. 2020a. Compound Risk and the Welfare Consequences of Insurance. CEAR Working Paper 2020-10. Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harrison, G.W., J. Martínez-Correa, J.T. Swarthout, and E.R. Ulm. 2017. Scoring Rules for Subjective Probability Distributions. *Journal of Economic Behavior & Organization* 134: 430–448.

Harrison, G. W., K. Morsink, and M. Schneider. 2020b. Do No Harm? The Welfare Consequences of Behavioural Interventions. CEAR Working Paper 2020-12. Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

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

Harrison, G.W., and J.M. Ng. 2019. Behavioral Insurance and Economic Theory: A Literature Review. *Risk Management and Insurance Review* 22 (2): 133–182.

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

Hill, R.V., M. Robles, and F. Ceballos. 2016. Demand for a Simple Weather Insurance Product in India: Theory and Evidence. *American Journal of Agricultural Economics* 98 (4): 1250–1270.

Lusardi, A., and O.S. Mitchell. 2014. The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature* 52 (1): 5–44.

Matheson, J.E., and R.L. Winkler. 1976. Scoring Rules for Continuous Probability Distributions. *Management Science* 22 (10): 1087–1096.

Matsuda, A., K. Takahashi, and M. Ikegami. 2019. Direct and Indirect Impact of Index-Based Livestock Insurance in Southern Ethiopia. *The Geneva Papers on Risk and Insurance-Issues and Practice* 44: 481–502.

McPeak, J., S. Chantarat, and A. Mude. 2010. Explaining Index-Based Livestock Insurance to Pastoralists. *Agricultural Finance Review* 70 (3): 333–352.

Milevsky, M.A. 2019. *Longevity Insurance for a Biological Age: Why Your Retirement Plan Shouldn't be Based on the Number of Times you Circled the Sun*. Toronto: PiLECo Ventures.

Mitchell, O.S., J. Hastings, and E. Chyn. 2011. Fees, Framing, and Financial Literacy in the Choice of Pension Manager. In *Financial Literacy: Implications for Retirement Security and the Financial Market Place*, ed. A. Lusardi, and O.S. Mitchell. New York: Oxford University Press.

Monroe, B. 2021. The Welfare Consequences of Individual-Level Risk Preference Estimation. In *Models of Risk Preferences: Descriptive and Normative Challenges*, ed. G.W. Harrison, and D. Ross. Bingley: Emerald, Research in Experimental Economics.

North, D.C. 1954. Life Insurance and Investment Banking at the Time of the Armstrong Investigation of 1905–1906. *Journal of Economic History* 14 (3): 209–228.

Ortmann, A., and R. Hertwig. 2002. The Costs of Deception: Evidence from Psychology. *Experimental Economics* 5 (2): 111–131.

Penrose, L.S., and J. Raven. 1936. A New Series of Perceptual Tests: Preliminary Communication. *British Journal of Medical Psychology* 16 (2): 97–104.

Primi, C., K. Morsanyi, F. Chiesi, M.A. Donati, and J. Hamilton. 2016. The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT). *Journal of Behavioral Decision Making* 29: 453–469.

Raven, J., J. Raven, and J. Court. 1998. *Manual for Raven's progressive Matrices and Vocabulary Scales: Section 4, Advanced Progressive Matrices, Sets I & II*. San Antonio, TX: Pearson.

Shee, A., C.G. Turvey, and J. Woodard. 2015. A Field Study for Assessing Risk-Contingent Credit for Kenyan Pastoralists and Dairy Farmers. *Agricultural Finance Review* 75 (3): 330–348.

Singal, A.G., P.D. Higgins, and A.K. Waljee. 2014. A Primer on Effectiveness and Efficacy Trials. *Clinical and Translational Gastroenterology* 5 (1): e45.

Stone, M.F. 1960. *The Teacher Who Changed An Industry: A Biography Of Solomon S. Huebner Of The University Of Pennsylvania*. Homewood, IL: Richard D. Irwin.