The Determinants of Good and Bad Insurance Decisions

Glenn Harrison*, Karlijn Morsink† and Mark Schneider‡

February 12, 2020

[DRAFT]

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, the need to use relatively exotic products in order to keep costs down for poor households, and the extent to which insurance products might crowd out pre-existing informal risk management arrangements. We focus on the determinants of insurance product demand, taking care to evaluate rigorously the understanding of the insurance product offered, as well as general literacy, cognition and intelligence. We further focus on index insurance products, which promise so much for risk management in developing countries. Our approach is to deliberately utilize the control of laboratory experiments to identify the actuarial and behavioral determinants of the demand for insurance. A feature of our approach is the identification of welfare-improving and welfare-worsening decisions, and the comparative determinants of each type of decision. We find that literacy plays a critical role, and that there are important differences in the way in which general financial literacy and domain-specific literacy about index insurance help to bring about better quality decisions.

JEL classification: D81, O2, D90, D78
Keywords: Index insurance, behavioural welfare economics, financial literacy, insurance literacy

*Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA. Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail: gharrison@gsu.edu

†Department of Economics, Utrecht University. Morsink is also affiliated with the Blavatnik School of Government, University of Oxford; Vrije Universiteit Amsterdam; and Center for the Economic Analysis of Risk, Georgia State University. Email: k.morsink@uu.nl

‡Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University. Email: mschneider@gsu.edu

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 are able to make purchase decisions that improve their welfare.[1] One would expect that a necessary condition for this improvement is that individuals understand the products that are marketed to them.[2] Despite the fact that understanding is critical for welfare, and 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 purchase rather than enhanced understanding or better financial decisions.[3] Furthermore, if understanding is explicitly tested, it is often through general financial literacy[4] and not the domain-specific understanding of the decision context in question.[5]

We investigate the importance of proxies for general understanding as well as domain-specific knowledge of the decision context on the purchase of index insurance, and benchmark their effects by comparing it to the standard actuarial determinants of insurance purchase. We speak to the quality of financial decisions by comparing purchase decisions that substantially *increase* welfare, to purchase decisions that substantially *reduce* welfare. 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 fairly complex product to understand and may not be welfare-enhancing for all (Clarke, 2016). We provide rigorous incentivized metrics for evaluating 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

[1] Many studies suggest that households under-save, engage in excessive and expensive borrowing, and underinsure (Stango and Zinman (2009), Gaurav et al. (2011), Karlan et al. (2016) and Casaburi and Willis (2018).

[2] We say that it is only an expectation that comprehension is a necessary condition to allow for the use of appropriate, domain-specific heuristics as a guide to decision-making. One such heuristic that we know to be important in the field is the trust that individuals place in certain institutions. Of course, this heuristic is not always applied appropriately, and can indeed be exploited for gain over time, as nicely illustrated by de Haan and Jona (2018).

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

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

[5] See, for example, Harrison and Phillips (2014) and Di Girolamo et al. (2015).

and losers, measured by the "quality" of their purchase in terms of consumer welfare. In effect, we recognize that there is a widespread concern with the levels of comprehension behind index insurance take-up, both from practitioners and academics, and propose measures to systematically address them.

We conduct lab experiments with 119 subjects who each make 54 decisions to purchase index insurance or not. We elicit subjects' bias and confidence in general financial literacy and domain-specific index insurance literacy through incentivized experiments. We also elicit their general financial literacy, cognitive reflection, and fluid intelligence using hypothetical surveys. We then separate purchase decisions that are welfare-enhancing in terms of consumer surplus from those that are welfare-reducing and investigate to what extent our measures of understanding as well as standard actuarial determinants effect the quality of purchase decisions. We deliberately start with a laboratory experiment, to cost-effectively identify the measures of comprehension that can be used in the field and to observe those that do or do not purchase the product. Of course we understand the importance of field context, and have been fierce advocates for the methodological importance of field experiments (Harrison and List, 2004). But field experiments are expensive at interesting scale and are often constrained to study very few, if any, variants on one contract, limiting the extent to which one can investigate the effect of standard actuarial determinants, and use these as a benchmark to compare to behavioural variables.

We find that for both the decisions that increase consumer surplus the most, as well as for decisions that reduce consumer surplus the most, individuals who score higher on our incentivized general financial literacy measures make better, or less poor, decisions. For those that make the worst welfare-reducing decisions, predominantly by *excess* take-up, greater financial literacy increases take-up significantly when the product should *not* have been purchased. For those that make the best welfare-enhancing decisions, predominantly by cautious take-up, greater financial literacy increases take-up significantly, and greater domain-specific incentivized insurance literacy decreases take-up significantly. In this case the domain-specific literacy makes sure that individuals do not make a welfare-reducing purchase decision. We benchmark these effects to the effects of standard actuarial variables (loss probability, loss amount, loading, and basis risk) which have the expected effect on purchase for the most welfare-enhancing purchase decisions, but opposite effects for the worst welfare-reducing purchase decisions. The effects of our incentivized literacy measures on purchase are of the same magnitude as the effects of actuarial variables. Finally, hypothetical survey responses for literacy do not provide any insight into the quality of purchase decisions.

Index insurance has great potential as a contract but there are serious marketing problems. The core problem is that it 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 that 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. Moral hazard plays no role because there are no actions by the insured that can affect the index level or payout decision, whether or not those actions could be observed by the insurance company. Similarly, adverse selection plays no role 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.[6] 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 which can be correlated with indices such as average temperature change.

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 the various determinants of demand. 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. And here we stress both the decision to purchase the

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

product as well as the decision not to purchase the product.[7]

We make three general contributions. The first is to the literature on financial literacy and education. In all cases that we are aware, the measurement of the educational effects of literacy interventions has been direct but hypothetical, and the measurement of the behavioral effects has been indirect but incentivized. Hypothetical survey questions, sometimes abstract and sometimes built around naturally-occurring settings, are used to measure literacy. And the effects are measured by proxies for welfare with real consequences, simply assumed to be measures of welfare in the absence of better information (e.g., increased savings for retirement). We seek to tighten these loose connections in the evaluation of what might be causing what.

The second is to the literature on the measurement of literacy, cognitive reflection, and fluid intelligence. In each case there are popular measures that are perhaps useful starting points, but which are often claimed to be much more than they are (or were originally claimed to be). In terms of **literacy**, the literature has been dominated by the use of the general financial literacy questions of Lusardi and Mitchell (2014), indeed typically 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 want to measure the confidence that an individual has in their knowledge of these concepts, as well as concepts that are domain-specific to insurance decisions. 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 portrayed as a measure of cognition or even intelligence. In terms of economic behavior, their correlation with equally hypothetical measures of risk preferences and time preferences, in Frederick (2005)[8], is what spurred their popularity as quick add-ons to survey batteries. As we discuss later, it is not clear that they are in fact just "good math problems," and have nothing *per se* to do with cognitive reflection (or so-called System 1 and System 2 thinking). In terms of

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

[8]One of the 17 measures of time preferences involved real rewards (Table 2, p.31), but did not correct for the effects of diminishing marginal utility that is required to infer time preferences (see Andersen et al. (2008)). None of the 18 measures of risk preferences involved real rewards (Table 3a, p.34), despite overwhelming evidence of hypothetical bias in these tasks (Holt and Laury (2002) and Harrison (2007)).

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.

The third is to the literature on index insurance, where we provide a rich characterization of the determinants of take-up, with special attention to the role of behavior traits. There have been several important attempts to characterize these determinants of purchase. Carter et al. (2008) proposed using an artefactual field experiment, much like our own laboratory experiment, as a financial education treatment.[9] 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. 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.[10] Giné et al. (2008) consider the role of risk aversion as a determinant of demand for a field index insurance product, and arrive at a much-referenced conclusion that demand is lower for those that they deem to be more risk averse. Building on the model of non-performance risk due to Schlesinger and Schulenburg (1987) and Doherty and Schlesinger (1990), Clarke (2016) established that one might not expect to see the traditional Insurance Economics 101 result for full indemnity products when dealing with basis risk from index insurance products and EUT decision-makers. Moreover, the measure of risk is of limited power for inferring risk preferences, although it was incentivized (p.550).[11]. Again, their focus is on take-up, with no evaluation of the

[9] 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.

[10] 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.

[11] The elicitation method was developed by Binswanger (1980), and discussed in relation to other elicitation methods by Harrison and Rutström (2008, §3.1)

quality of take-up decisions. 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 present our experimental design in detail (section 3), and review descriptive evidence from the experiments spanning 119 subjects and 54 choices over contracts per subject (section 4). We then provide evidence of the effect of our literacy, cognition and intelligence measures on take-up, with emphasis on the differences between good quality decisions and bad quality decisions. Conclusions, and implications for policy, are then provided (section 6).

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 in Section 2.1. We then explain what we mean by literacy, cognition and intelligence in Section 2.2. Finally we discuss how we identify the "quality" of purchase decisions in Section 2.3

2.1 The Marketing of Index Insurance

For many years we have been conversing with insurance companies, non-government organizations, government bodies and insurance regulators charged with marketing, or evaluating the marketing, of index insurance products in developing countries.[12] 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. 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. Mitigating these concerns are efforts to display index outcomes in the form of contour maps or choropleth maps.

There might be 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 best 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 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. Of course, this remains an index or parametric product, since everyone covered by the product receives that payout. But the function mapping index values below the threshold and payouts might not reflect the function mapping the index values below the threshold and losses: one might be linear, and often is, and the other highly non-linear. This function is also, in the statistical nature of extremes, typically less precisely estimated than other parts of the distribution.[13]

[12] For example, at meetings of the International Microinsurance Conference since 2010 (https://www.munichre-foundation.org/home/InclusiveInsurance/common/Microinsurance_Archive/2010IMC.html). This conference is now re-named as the International Conference on Inclusive Insurance (https://www.munichre-foundation.org/home/InclusiveInsurance.html).

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

Second, it is very hard to find out **what is actually said** 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 said. Even if standardized scripts are prepared, as they often are, there is little or no enforcement of the conversation that actually occurs. In many setting 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, that they really want, and trust the broker of that primary product. 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 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.[14] 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 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 close to 1. A variant on this comment is that basis risk is something academics and actuaries worry about, not those that sell the product. And there might be some kernel of truth in that claim: even statistical properties that actuaries write out with convincing precision rarely come with equally formal statements of

[14]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.

"actuarial prudence." Hence, so the argument goes, there is a myriad of contractual uncertainties that one could attend to, such as non–performance risk due to inadequate reserving, and basis risk pales in the shadows of those factors. We certainly understand, as one experienced industry hand put it, that you do not start selling by leading with the biggest limitation of the product.

We stress that there are many "good actors" marketing index insurance products. 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.[15] Equally, we have heard stories of some insurance regulators holding applicants with new products to the fire, requiring them to explain how they cover different contingencies. How much of that *ex ante* "torching" prior to approval finds its way into the marketing trenches is an open question, of course. And we do not know of any insurance regulator that undertakes evaluations of client marketing processes ex post approval of index contracts in developing countries.

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. In the end we devise explicit measures for these concepts. Our perspective is that the responses we see in insurance purchase decisions come from a "cognitive production function" that pools effort, prior knowledge, heuristics, logic, and time.[16] 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 determines if best to use some heuristic, whether

[15] 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

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

to use a considered heuristic or 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 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 individuals are deemed literate with respect to the topic of that question. Indices of how literate the individual is, by a simple sum of the correctly answered questions, are then often constructed and used in estimations to correlate with downstream behavior (Lusardi and Mitchell, 2011, 2014; Mitchell et al., 2011). Rather than utilize multiple-choice or fill-in-the-blank survey instruments that measure literacy noisily as a point estimate, 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 follow the methods of Harrison and Phillips (2014), Di Girolamo et al. (2015) and Harrison et al. (2017) as an incentivized approach to ascertain how precise a decision-maker's knowledge is. By eliciting the subjective belief distribution, we can also construct measures of literacy and welfare, test for bias, and measure the level of confidence a decision-makers has in their subjective beliefs. Harrison and Phillips (2014) and Harrison et al. (2017) document in detail the subjective belief elicitation procedure that will be followed to test literacy.

2.2.2 Cognitive Reflection

The popular Cognitive Reflection Test (CRT), due to Frederick (2005), seeks to measure a person's unincentivised 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 measure of "actively open-minded thinking" that values questioning initially-favored conclusions.

All existing CRT problems are essentially math problems, albeit mathematical word problems. So they test the ability to think as if one can discern the mathematical structure at play, and then solve it for the correct answer. These two steps are important and distinct, and when a false lure is in play the meta-cognitive stress is on even starting this thought process, let alone executing it reliably. Despite being recognized as math problems, there appears to be a consensus, expressed by Campitelli and Gerrans (2014, p. 435) that "CRT problems, unlike other mathematical problems, trigger an automatic response, which is then inhibited or not, and only if inhibition is successful would individuals use their mathematical knowledge to solve the problem."

A growing literature, reviewed by Baron et al. (2015) and Attali and Bar-Hillel (2020), investigates the extent to which the CRT questions are "just good math problems," or if they have some special attributes beyond that. The emerging conclusion is that they are just good math problems. Attali and Bar-Hillel (2020, p. 105ff) summarize:

> CRT items are very appealing in some stylistic, aesthetic sense. Many of them can be found in collections of riddles. Frederick found a version of bat-&-ball in a 1919 Jewish puzzle book. Bar-Hillel knows variants of the other two from her childhood days, almost 70 years ago. It is no wonder that people are inclined to spread them among their friends and colleagues. [...] It has come to be regarded as a kind of instant IQ test, of almost magical predictive power. And the existence of an immediately available, seemingly apt, but wrong answer — the fast lure — that needs to be rejected in spite of its allure, is itself part of the CRT's allure. Alas, that allure is false. The main claim of this paper is a downer compared to this mystique. We claim that insofar as we tested, CRT items are essentially just ordinary math word problems. [... So, what is it ...] about such math problems that affords the kind of predictive power that is found. Is it math ability? Is it *g* – general intelligence? Is it reflective cognitive style? We can state only one conclusion with any confidence: it does not rely on the existence of a lure, and the need to overcome it *en route* to a correct solution. [...] Existing theories in the literature that explain the CRT's achievements whilst relying on its most characteristic feature — the existence of a lure — now need to acknowledge that having a lure is not a necessary condition for the CRT's power.

We therefore evaluate the *type* of CRT response to our questions, and also complement the evaluation with a well-regarded measure of general fluid intelligence. 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 intelligence test we consider is the Raven Advanced Progressive Matrices (RAPM) test, documented by Raven et al. (1998). The RAPM consists of a set of 12 problems called Set I, which are generally easier than the 36 problems in Set II. The Set I problems are often used as a fast, coarse measure of intelligence, to determine if the more discriminatory Set II are needed for some subjects. 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 should select the correct image. The original RAPM was administered in a "paper and pen" fashion, with prepared answer sheets, and relying solely on intrinsic rewards. Our subjects completed Set I of the RAPM as a series of hypothetical survey questions.

The RAPM is widely viewed as a major test for "fluid intelligence" or "analytic intelligence," where these terms are understandably debated. It is apparent from the original development of the test, in Penrose and Raven (1936), that it was seeking to measure fluid intelligence. Focusing on the RAPM, Carpenter et al. (1990) evaluated the "... cognitive processes in a widely used, nonverbal test of analytic intelligence ... [and] analyzed [them] in terms of which processes distinguish between higher scoring and lower scoring subjects and which processes are common to all subjects and all items on the test" (p. 404). Analytic intelligence is defined as the "ability to deal with novelty, to adapt one's thinking to a new cognitive problem" (p.404). They developed one computer simulation model that could perform akin to the median score of a sample of college students, and another model that could perform akin to the scores of the very best. The differences in the two models served to suggest the nature of the analytic intelligence required to perform the RAPM at different levels.

One sense in which the RAPM is attractive as a test of fluid intelligence is that it the primary stimuli for a response do not rely on any formal linguistic skills. Obviously the instructions do, but they are typically brief and often delivered orally. Indeed, it is proposed that the RAPM be accompanied by a test of vocabulary to "round out" the assessment of intelligence. The specific test used as part of the original Raven suite of progressive matrices is known as the Mill Hill Vocabulary Scale.[17] The Raven Progressive Matrices are viewed as measuring fluid intelligence, and vocabuilary tests for as the Mill Hill test as measuring aspects of "crystallized intelligence."[18]

2.3 The Quality of Purchase Decisions

Harrison and Ng (2016) provided an explicit welfare analysis of the simplest full indemnity insurance contract in controlled laboratory experiments.[19] They used the risk preferences for each individual estimated from a risk aversion task to infer if the individual was an Expected Utility Theory (EUT) or Rank-Dependent Utility (RDU) decision maker, and to provide parameter estimates for their specific risk preferences. Armed with estimates

[17]It is described by the American Psychological Association as "a measure of crystallized abilities (see Cattell–Horn theory of intelligence) designed to be administered in conjunction with Raven's Progressive Matrices. It consists of 88 words and is available in three forms: one requiring participants to define all words, one requiring participants to recognize the meanings of all words by choosing the correct synonym for each from among six options, and one requiring participants to define half of the words and choose synonyms for the rest. [The test was ....] originally developed in 1943 by British psychologist John C. Raven (1902–1970) at Mill Hill Emergency Hospital, London" (`https://dictionary.apa.org/mill-hill-vocabulary-scale`).

[18]The American Psychological Association explains that the "Cattell–Horn theory of intelligence [is] a theory proposing that there are two main kinds of intellectual abilities nested under general intelligence: *g-c*, or crystallized intelligence (or ability), which is the sum of one's knowledge and is measured by tests of vocabulary, general information, and so forth; and *g-f*, or fluid intelligence (or ability), which is the set of mental processes that is used in dealing with relatively novel tasks and is used in the acquisition of *g-c*. In later versions of the theory, other abilities were added, such as *g-v*, or visual intelligence (or ability), which is the set of mental processes used in handling visual-spatial tasks (e.g., mentally rotating a geometric figure or visualizing what pieces of paper would look like if folded). [The theory was] originated by Raymond B. Cattell in the 1940s [and] subsequently developed by U.S. psychologist John L. Horn (1928–2006) beginning in the 1960s." (`https://dictionary.apa.org/cattell-horn-theory-of-intelligence`). Carpenter et al. (1990) initiated a long literature evaluating the sense in which the RAPM stressed elements of *g-v*, which is believed to drive some gender differences in RAPM scores.

[19]An extensive discussion of the methodological assumptions and implications of this approach is provided by Harrison and Ng (2016, p. 111-116), Harrison and Ng (2018, 2019) and Harrison (2019, §5), all in the specific context of insurance. Harrison and Ross (2017) provide a similar discussion in the context of portfolio choice, and also offer a general philosophical exposition of what they characterize as the "quantitative intentional stance" towards behavioral welfare economics.

of the utility function of each subject, they were able to directly calculate the expected[20] Consumer Surplus (CS) of purchasing insurance or not purchasing insurance, in each case using the Certainty Equivalent (CE) difference between the two actions.[21]

Consider the calculation for a specific subject and for a specific contract.[22] If this subject is classified as an EUT decision-maker, we know her risk preference parameters from a prior risk aversion task and can use her estimated utility function to calculate the CE of purchasing the insurance contract and the CE of not purchasing the insurance contract. For this we need to calculate the EUT of the decision to purchase the contract, and the EUT of the decision to not purchase the contract. The former depends on the actuarial parameters of the contract.

The difference between these CE is her *ex ante* CS from purchasing. It is possible that if this subject was an EUT decision-maker she should have purchased up to and including a premium of $1.60, say, and then not purchased for higher premia. If this subject is classified as an RDU decision-maker, we would *in general* infer different CE for each possible choice, and different CS from purchasing. For example, the same subject, modeled as an RDU decision-maker, might have purchased up to and including the premium of $2.20.

This simple example has important implications. First, even if we assume an individual is an EUT decision-maker, we need to know *how risk averse* she is to say if her decision to "take-up" the product is the right one or not. The same point applies generally to the case in which she is an RDU decision maker. A reminder from the *Economics of Risk 101* that there should be a pox on unconditional nudges to take up the insurance product that ignore risk preferences! Second, we need to know which *type* of risk preferences best characterizes her. It is easy to construct realistic examples in which we get the *sign* of the

[20] The word "expected" refers here to the outcome, whether or not a loss was realized, and whether or not a claim paid. Some non-economists misunderstand the word "expected" to refer to whether a claim was paid, and incorrectly use that as a metric for evaluating if an insurance product was good for the individual considering purchasing the contract. That approach confuses a risk management product with an investment product.

[21] Carter and Chiu (2018) propose essentially the same approach to setting "quality standards" for the design of index insurance contracts, proposing that plausible simulations of EUT risk preferences be used to evaluate the CS for various contracts before approval by regulators. We heartily endorse their proposal, and would extend it to consider non-EUT risk preferences (particularly RDU) and subjective beliefs (about loss probabilities and basis risk, in particular).

[22] A worked numerical example is provided by Harrison (2019, Table 1).

welfare effect wrong unless we know the type of decision-maker. Third, we have CS numbers in dollars, reflecting the equivalent variation in income. We can distinguish what are "small" welfare effects from what are "large" welfare effects. This insight will be important for our analysis, since it provides a metric to judge the quality of the purchase decision. Finally, we have calculated the CS from purchasing or not from purchasing. This might seem trivial, until one realizes that many observational data sets, not all, suffer from the selection bias of only seeing those that purchased insurance.

Armed with these calculations, we can then immediately calculate realized welfare gains and welfare losses for this subject. The *realized* CS is defined as the CS that comes from the observed decision: the CS *gained* from making the right decision or the CS *foregone* by making the wrong decision. To stress, the right decision here is often *not* to purchase the product (e.g., relatively risk averse subjects facing a high loading factor). Hence over the many choices by the same subject, with varying actuarial paremeters, we might see welfare gains and losses by the same subject. It is not as if a given subject always makes the right decision or always makes the right decision, although we rejoice if we see the former and despair if we see the latter. Again, some of these losses arise from taking up the insurance contract, to warn us from automatically associating take up with a welfare gain (i.e., to warn us from a naive application of revealed preference).

3 Experiments

We conduct our experiments with 119 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 in 10 financial literacy and 10 index insurance literacy questions are elicited. 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.

3.1 Index Insurance Purchase Decisions

The index insurance decisions are the focal point of our experiment. In this task 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%. Before subjects make their insurance purchase decisions they receive basic instructions about the insurance. We now present the core part of these instructions while the full instructions are presented in Appendix B.1. All references to a figure made in this extract refer to Figure 1:

> In this task you will make choices about whether to insure against possible monetary loss. In each choice you will start out with an initial amount of money and, in the event of a loss, the loss amount will be taken from this initial stake. In each choice you will have the option to buy insurance to protect you against the possible loss, although you are not required to buy the insurance.
>
> You will make 54 choices in this task. You will actually get the chance to play one of the choices you make, and you will be paid in cash according to the outcome of that choice. So you should think carefully about how much each insurance choice is worth to you.
>
> Each choice has two random events: a **Personal Event** and an **Index Event**. Each event has two possible outcomes: Good or Bad. If the Personal Event outcome is Bad, then you will suffer a loss. Before you know the outcome of the Personal Event, you must decide whether to purchase insurance against this possible loss. However, the insurance only pays a claim if the Index Event outcome is Bad.

If you do not purchase insurance, then only the outcome of the Personal Event will decide your earnings:

Personal Event	Your Earnings
Bad	Initial stake - Loss
Good	Initial stake

If you do purchase insurance, it is important for you to understand that an insurance claim is not paid according to whether you actually suffer a loss. Instead, an insurance claim is paid only according to the Index Event. Both events will decide your earnings:

Personal Event	Index Event	Your Earnings
Bad	*Bad*	Initial stake - Insurance cost – Loss + Insurance coverage
Bad	*Good*	Initial stake - Insurance cost – Loss
Good	*Good*	Initial stake - Insurance cost
Good	*Bad*	Initial stake - Insurance cost + Insurance coverage

So there are four possible outcomes if you purchase insurance. You might suffer a loss and receive an insurance claim payment. Or you might suffer a loss but not receive an insurance claim payment. You might not suffer a loss and also receive no insurance claim payment. Finally, you might receive an insurance claim payment even when you do not suffer a loss.

Each event is determined by randomly drawing a colored chip from a bag. In general, each draw will involve two colors, and each decision you make will involve different amounts and mixtures of two colors. When making each decision, you will know the exact amounts and mixtures of colored chips associated with the decision. After you have decided whether or not to purchase insurance, the two events will be determined as follows.

First, the Personal Event will be determined with blue and red chips.

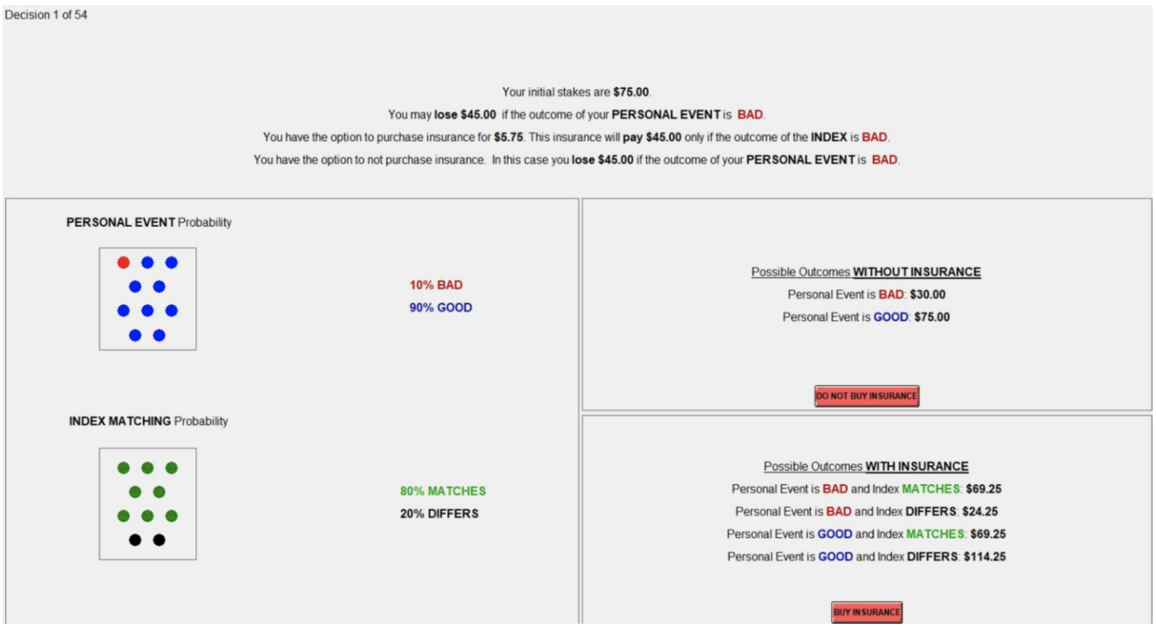
- If you draw a **blue** chip, then the Personal Event outcome is Good and you do not suffer a loss.
- If you draw a **red** chip, then the Personal Event outcome is Bad and you suffer a loss.

Next, if you purchased insurance, the Index Event will be determined with green and black chips.

- If you draw a **green** chip, then the Index Event outcome Matches the Personal Event outcome
- If you draw a **black** chip, then the Index event outcome Differs from the Personal Event outcome

Here is an example of what your decision would look like on the computer screen. The display on your screen will be bigger and easier to read.

Figure 1: Example Screen of Index Insurance Purchase Decision



In this example you start out with an initial stake of $75. If the outcome of the Personal Event is Bad you will lose $45, and if the outcome of the Personal Event is Good you will not lose any money. If you faced the choice in this example and chose to purchase insurance, you would pay $5.75 from your initial stake. You would pay this $5.75 before you drew any chips, so you would pay it regardless of the outcomes of your draws.

You will be drawing colored chips from bags to determine the outcomes of both events. First, you will draw a chip to determine the Personal Event outcome. The image on the left shows that there is a 10% chance that the Personal Event outcome is Bad, and a 90% chance that the Personal Event outcome is Good. This means there will be 9 blue chips and 1 red chip in a bag, and the color of the chip you randomly draw from the bag represents the outcome of the Personal Event. If a blue chip is drawn, the Personal Event outcome is Good, and if a red chip is drawn the Personal Event outcome is Bad.

Next, you will draw a chip to determine the Index Event outcome. There is an 80% chance that the Index Event outcome Matches the Personal Event outcome and a 20% chance that the Index Event outcome Differs from the Personal Event outcome. This means there will be 8 green chips and 2 black chips in a bag. If a green chip is drawn the Index Event outcome Matches the Personal Event outcome, and if a black chip is drawn the Index Event outcome Differs from the Personal Event outcome.

You will indicate your choice to purchase, or not purchase, the insurance by clicking on your preferred option on the computer screen.

There are 54 decisions like this one to be made, each shown on a separate screen on the computer. Each decision might have different chances for the Personal Event outcome, the Index Event outcome, the initial stake, or the cost of insurance, so pay attention to each screen.

There are some 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 correaltion of the two, and randomly generate personal and index realization consistent with that correlation. The logical difficulty is that one would need many such realization 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 that 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 not 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. [23]

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 subjects as behaving according to

[23] The only other experimental interface focused on index insurance that we are aware of was used in artefactual field experiments in Peru 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.

EUT and RDU (see Appendix B.2). This characterisation allows us to predict the consumer surplus 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.

The battery is based on designs from Loomes and Sugden (1998) to test the Independence Axiom, designs from Harrison and Swarthout (2016) to evaluate Cumulative Prospect Theory (CPT) models of risk preferences, designs from Harrison et al. (2015) to test the Reduction of Compound Lotteries (ROCL) axiom, and a series of lotteries that are actuarially-equivalent versions of some of our index insurance choices. Each subject faced a randomized sequence of choices from this battery of 100. The analysis of risk attitudes given these choices follows Harrison and Rutström (2008), and is undertaken at the level of the individual following Harrison and Ng (2016). The typical interface used is shown in the first figure in Appendix B.2, for instances of two simple lotteries. For compound lotteries, we used a simple "Double or Nothing" option, illustrated in the second figure in Appendix B.2.

We classify a subject as being characterized best as having EUT or RDU risk preferences, in order to make it clear that it is not just the *level* of risk aversion that matters, but the *type* of risk preference that matters as well. This characterization is operationalized by a subject needing to have RDU estimates that indicate statistically significant evidence of probability weighting. Specifically, we undertake a Wald test of the hypothesis that the probability weighting parameters (γ for the Power and Inverse-S specifications, and η and ϕ for the Prelec specification) take on values that imply no probability weighting ($\gamma = 1$ and $\eta = \phi = 1$). If we reject this null hypothesis of EUT risk preferences at the 5% significance level, we characterize an individual as having RDU risk preferences. It is a simple matter to vary this particular significance level.

However, we always use the estimated utility function from the preferred RDU estimates for every subject when evaluating the CS for each subject. Why do this, if we have characterized an individual as EUT? 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.[24] And, perhaps surprisingly, these standard errors can make a substantive difference in

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

precisely the normative evaluations undertaken here.[25] Hence there is no formal need to differentiate EUT and RDU decision makers for these calculations, because EUT is nested within RDU.[26]

3.3 Hypothetical Literacy Measures

Subjects first participate in a popular test of fluid intelligence based on Set I of the **RAPM**. In this task subjects are asked to complete a series of 12 matrices by identifying a missing element that completes the pattern in each. The instructions to this task are provided in Appendix B.5

Second, subjects participate in the **CRT**, which measures a person's unincentivised tendency to override an incorrect "gut" response and engage in further reflection to find a correct answer. In this task subjects answer 6 questions in total: 3 questions are from Frederick (2005), who first described the task, and a further 3 questions from Primi et al. (2016). Each question, along with their correct and heuristic answer, are listed in Appendix B.6.

Third, we ask subjects 9 survey questions about financial literacy. These questions spanned concepts including interest, inflation, stock returns, bond prices, risk diversification, and mortgages. The questions, along with their answers, are listed in Appendix B.7.

3.4 Incentivized Financial and Index Insurance Literacy

Subjects participate in two tasks that assesses subjective beliefs about their own answers to a set of standard financial literacy questions (**Financial Literacy Beliefs**) Lusardi and Mitchell (2007, 2008, 2014) and a set of ten insurance decisions randomly selected from the 54 insurance choices (**Index Insurance Beliefs**). We elicit these beliefs following the

[25] Monroe (2021) shows that they can make a difference to the analyses of full indemnity insurance contracts Harrison and Ng (2016), and our approach follows theirs.

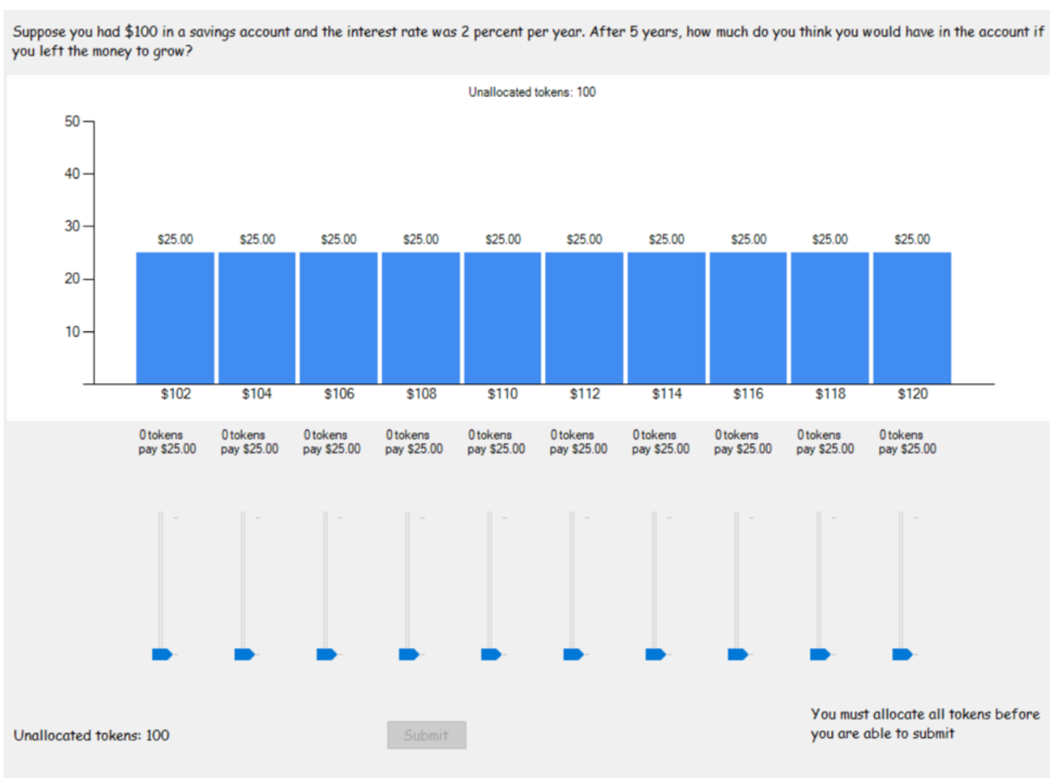
[26] The same point does *not* apply when considering models that are not nested. An important example is Cumulative Prospect Theory, which actually does not nest RDU or EUT, as explained by Harrison and Swarthout (2016). In this instance one can use non-nested hypothesis tests to categorize subjects for the calculation of their CS.

method by Di Girolamo et al. (2015) and Harrison et al. (2017), who make use of the incentivized Quadratic Scoring Rule 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.

A computer interface is used to present the belief elicitation tasks to subjects and record their choices, allowing them to allocate tokens in accordance with their subjective beliefs. For each set of incentivized questions, financial literacy questions or insurance decision questions, one question is selected for payment (see Appendix B.3 and Appendix B.4).

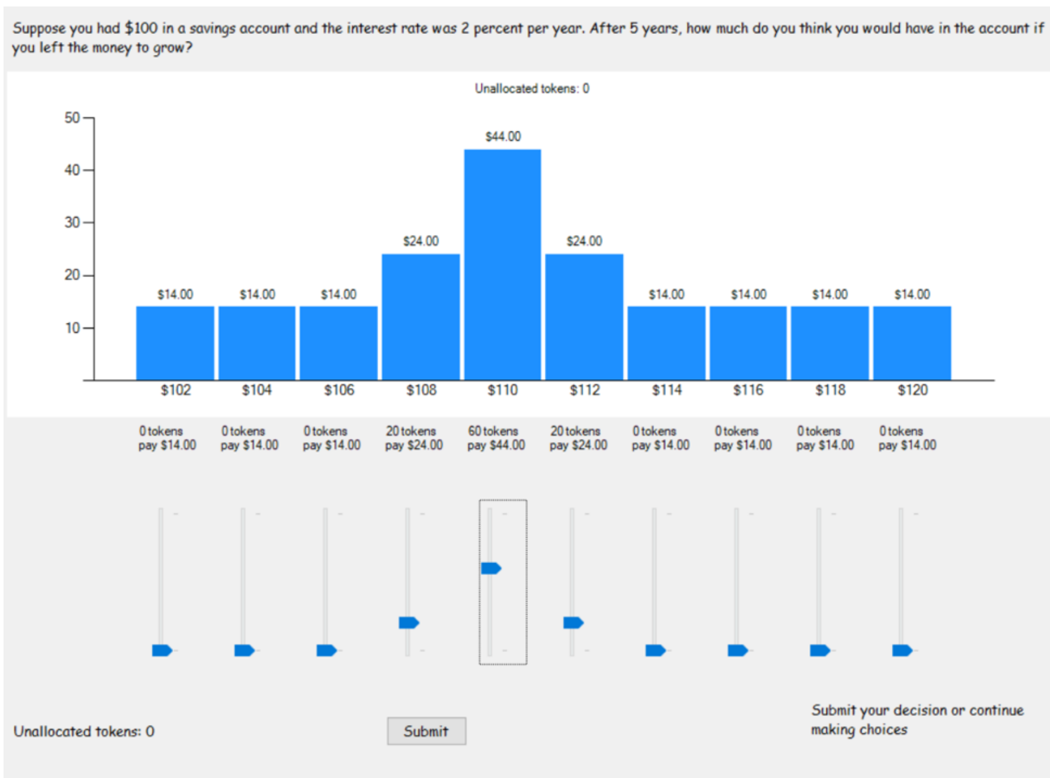
Figure 2 is an example of the display response screen, and illustrates an example using one of the financial literacy questions the subject received. The question reads *"Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would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

Figure 2: Example Screen of Incentivized Financial Literacy Question



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 reflect 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 shown in Figure 3.

Figure 3: Example Screen of Incentivized Financial Literacy Question



The earnings of each participant are based on the payoffs, which are generated by a discrete version of a Quadratic Scoring Rule (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 Figure 3 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 lies in bins 1, 2, and 3, or in bins 7, 8, 9, and 10. But they do 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 participants 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 Figure 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 person 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 4 is an example of reference material provided to participants relating to the insurance literacy task; note that "Figure B" appears in the top-left corner and the information in the lower-right panel of Figure 1 is now omitted, thus the participant will have to calculate the outcome for each insurance question.

Figure 4: Example Figure for Index Insurance Literacy Question

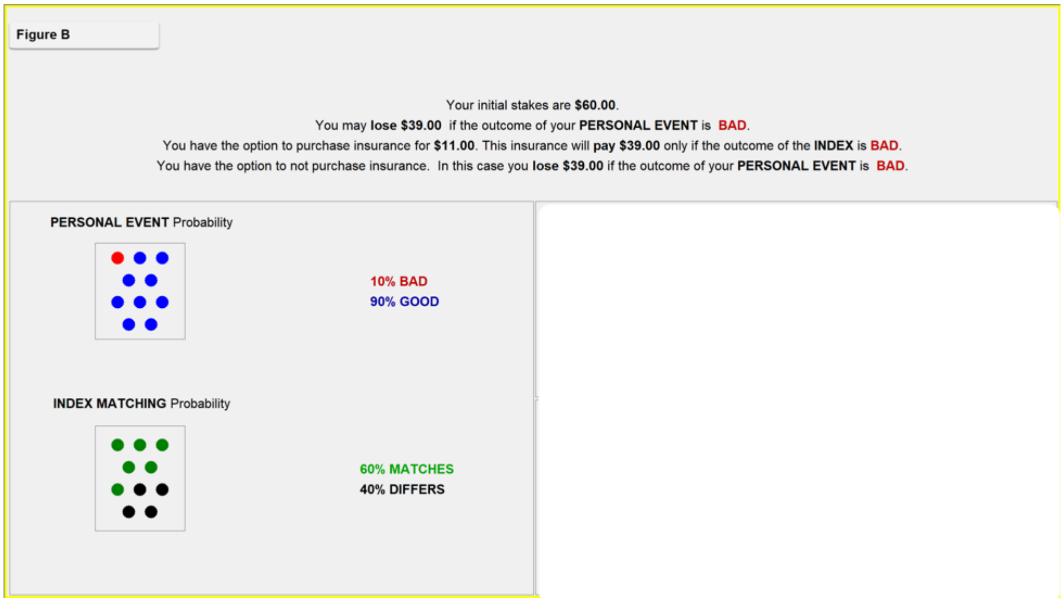


Figure 5: Example Decision Screen for Index Insurance Literacy Question

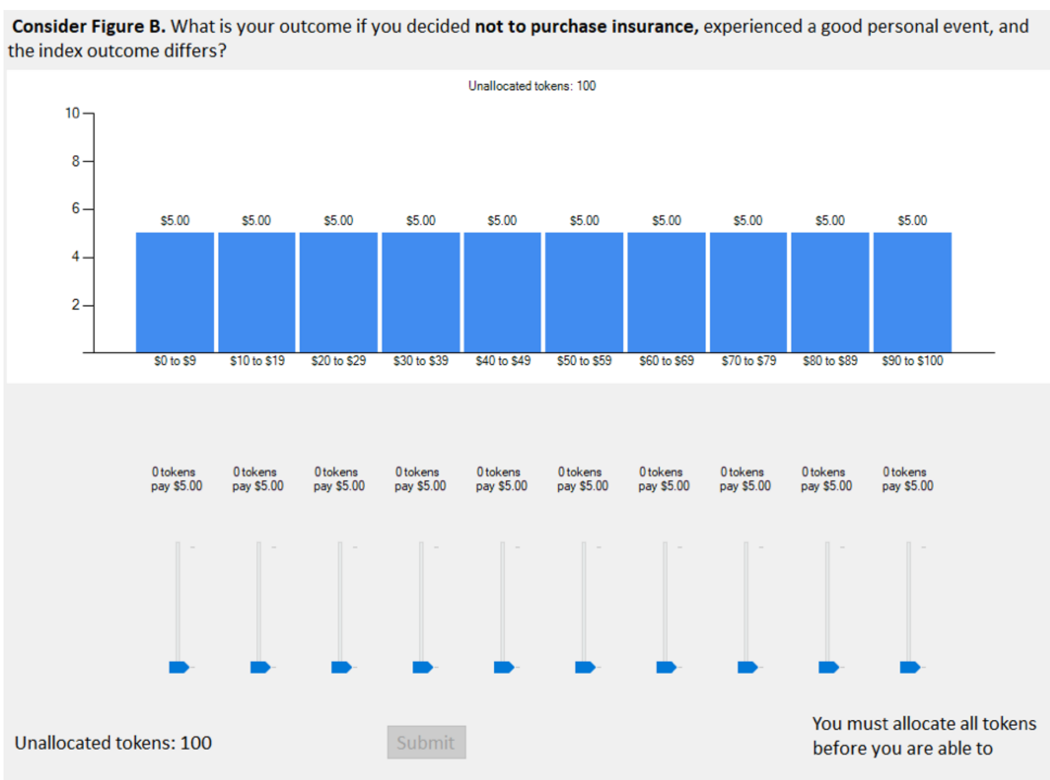


Figure 5 is an example of an insurance literacy question asked to the subject. It reads: *"Consider Figure B. What is your outcome if you decided not to purchase insurance, experienced a good personal event, and the index outcome differs?"* Here the subject refers to Figure B in the reference materials in order to calculate the payoff, and then place some bets on their beliefs about the answers to the question being asked.

Working through the earlier example, 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 Figure 5. Subjects could earn up to $10 in this task.

For both the Financial Literacy Index Score and the Index Insurance Literacy Score 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 50 tokens were allocated to the correct bin, then L = 0.50 (= 50/100). 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

We first describe the data and then draw inferences using appropriate statistical methods.

4.1 Descriptives

Figure 6 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 questions 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.[27] 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). On it's face this is suggestive of a familiar hypothetical bias in experimental economics, where incentives lower the noise in responses.

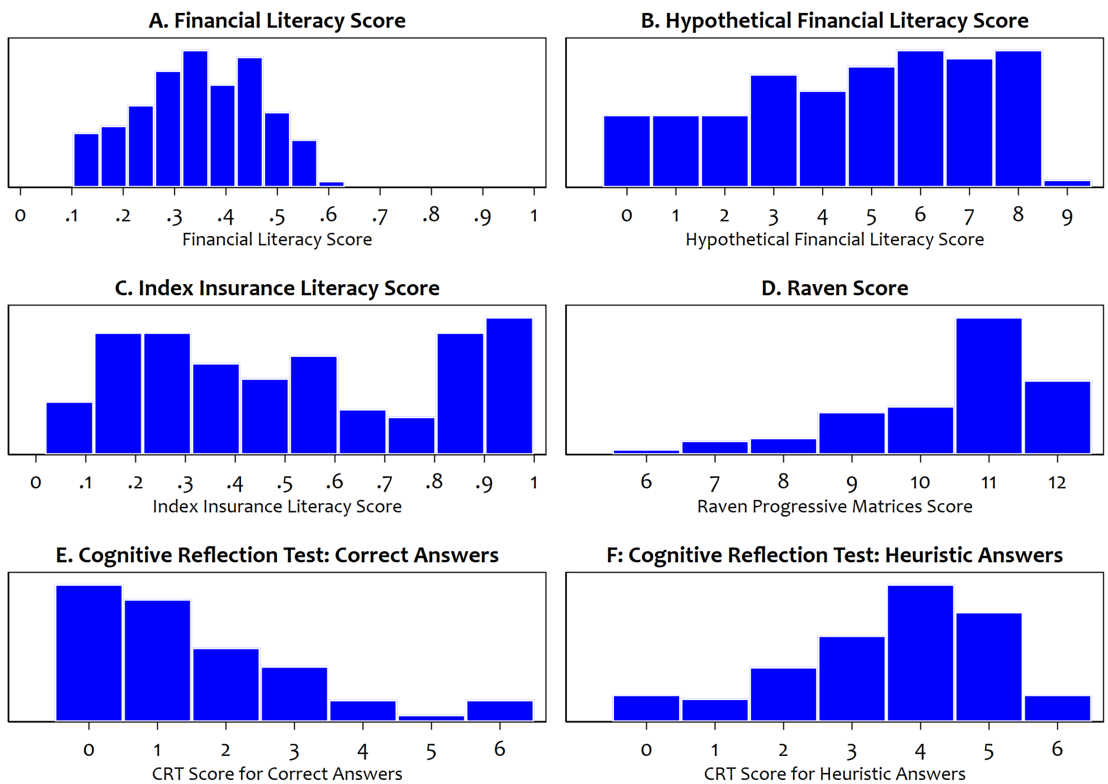
The Index Insurance Literacy Score in panel C displays a much wider distribution than the general Financial Literacy Score in panel A. The latter is intended as a general financial literacy measure, and the former as a tightly domain-specific literacy measure. These differences tell us three things. 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 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. And Set I of the RAPM was in large part designed to quickly determine if a subject should then be given the harder Set II. Hence we might not see much inferential power from these scores, with most subjects bunched towards the top possible scores.

The CRT scores display a familiar pattern from previous research, where the "false allure" 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

[27]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. For example, in separate research we compare inventivized and hypothetical responses to Set II of the RAPM, and find a 17 percentage point increase in the score with incentives that is statistically significant. That is of great importance for determining the level of fluid intelligence a person has, but does not neceesarilly imply that the hypothetical responses are determining, say, the average effect of gender on fluid intelligence.

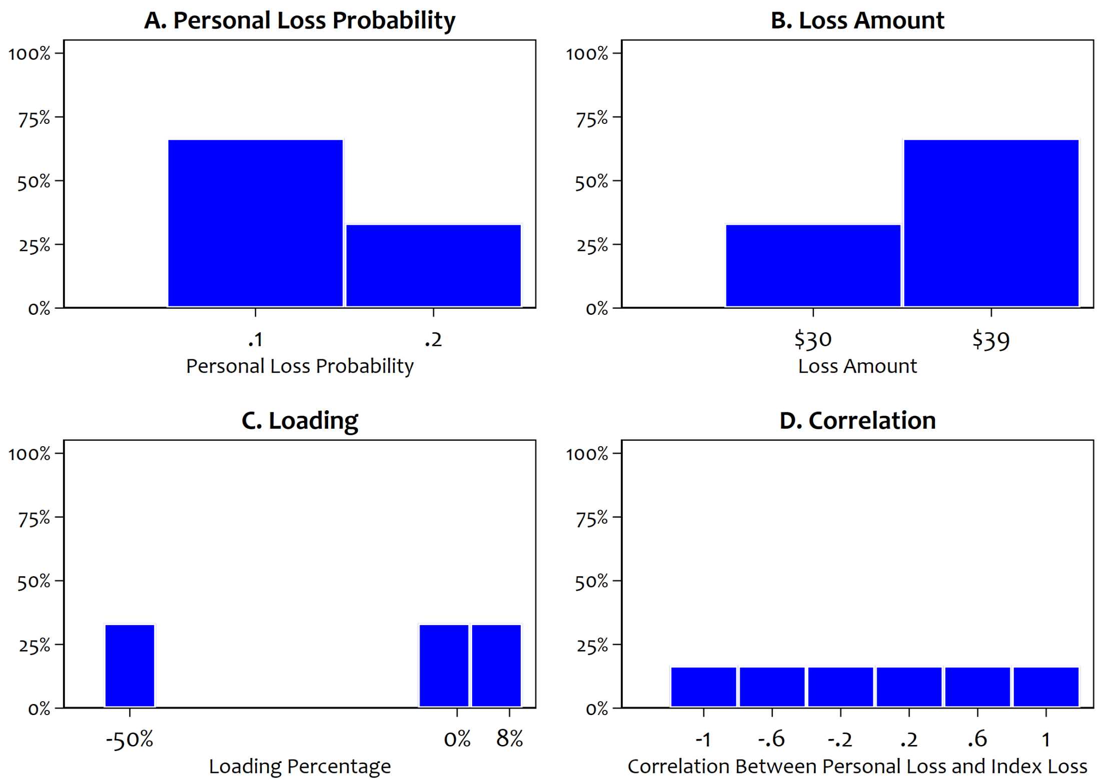
Figure 6: Distributions of the Literacy and Cognitive Reflection Variables



Note: For each distribution the number of subjects is 119. "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.

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 7: Distributions of the Actuarial Variables



Note: Each of the 119 subjects made 54 decisions. The Panels A, B, C and D present the characteristics of these 54 decisions in terms of personal loss probability, loss amount, loading percentage, and correlation.

Figure 7 displays the experimental design in terms of the actuarial variables defining the insurance contracts offered. Each subject was asked to consider purchasing an insurance contract 54 times, and each panel shows the distribution of these variables across these 54 choices. As explained earlier, we had two personal loss probabilities, two loss amounts, three loading percentages, and six correlations between the personal loss and the index loss. 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.

Figure 8: Unconditional Correlations of Purchase and Literacy Variables

	Probability of Purchase	Financial Literacy Score	Hypothetical Financial Literacy Score	Insurance Literacy Score	Raven Progressive Matrices Score	CRT Score for Correct Answers
Financial Literacy Score	0.05					
Hypothetical Financial Literacy Score	0.12	0.39				
Insurance Literacy Score	0.05	0.42	0.23			
Raven Progressive Matrices Score	-0.01	0.09	0.11	0.25		
CRT Score for Correct Answers	0.16	0.49	0.52	0.40	0.26	
CRT Score for Heuristic Answers	-0.19	-0.35	-0.39	-0.28	-0.07	-0.85

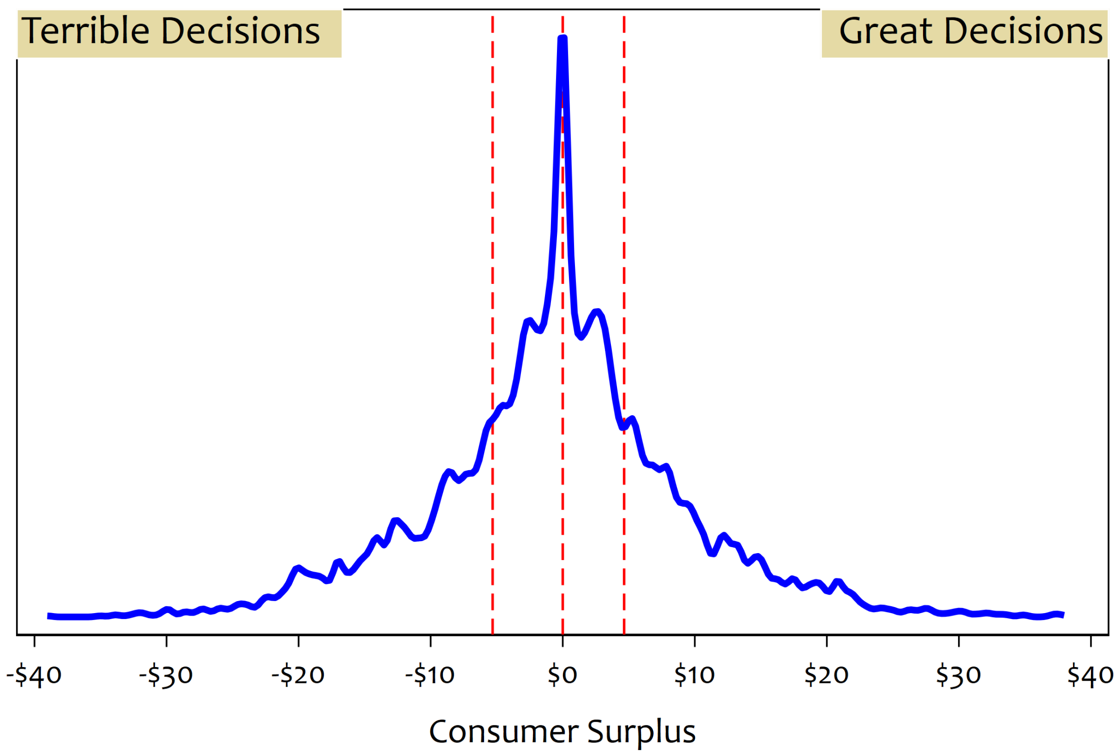
Note: Unconditional correlations at the level of a subject. For each financial literacy variable there are 119 observations, corresponding to one observation per subject. The "Probability of Purchase" represents the fraction of the number of purchase decisions per subject where the subject decided to purchase insurance divided by the total number of possible purchase decisions per subject. Each subject was asked to make 54 decisions to purchase insurance or not.

Figure 8 displays unconditional correlations[28] of the insurance purchases observed and the literacy variables, where "literacy" is a broad term to cover all of our measures. The Probability of Purchase is just the empirical frequency of a purchase decision over all 54

[28] Unless otherwise noted, any correlation referred to in the text is statistically significantly different from zero at the 5% level.

decisions by a given subject. In later analyses this will be the predicted probability from a statistical model, but here it is just observed purchases.

Figure 9: The Quality of Insurance Purchase Decisions



Note: Kernel density of the realized CS for each of the 119 subjects making 54 choices. Hence the sample is 6426. Dashed red vertical lines show the 25^{th}, 50^{th}, and 75^{th} centiles. We classify terrible decisions as those that led to a realized CS below the 25^{th} centile (N=938 decisions) and great decisions as those that led to a realized CS above the 75^{th} centile (N=893 decisions).

In an important sense, Figure 8 is a straw man, which we present just to tear down. It tells someone what factors are correlated with observed take-up of the insurance product. It is the sort of information that someone interested in marketing the product, and consumer welfare be damned along the way, would live and die by. Normatively, we argue that such marketing practices for insurance should indeed die. In any event, Figure 8 tells us that general financial literacy and domain-specific insurance literacy are not closely correlated on average with take-up. Similarly, the Raven measure of fluid intelligence has close to zero correlation on average with take-up. Of course, there are segments of the population where such factors might be more correlated with take-up *per se*, and we illustrate that point shortly when we look at welfare-improving purchases and welfare-reducing purchases.

Figure 9 displays the realized Consumer Surplus from all of the 54 decisions that all of the 119 subjects made. The measure for CS is dollars, and we also display the 25^{th}, 50^{th} and 75^{th} centiles in dashed, red lines. The median CS is virtually zero, so we see roughly 50% of the purchase decisions reducing welfare and 50% of the purchase decisions improving welfare. To focus on essentials, we separate these decisions into those that we deem Great Decisions, generating a CS (gain) in excess of the 75^{th} centile, and those that we deem Terrible Decisions, generating a CS (loss) below the 25^{th} centile. This separation helps us avoid comparing lots of decisions that generated *de minimus* CS gains and losses around zero.

We stress, again and for good cause, that Figure 9 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 will see.

Figure 10: Unconditional Correlations of Purchase and Literacy Variables, for Purchase Decisions that are Terrible or Great for Welfare

	Probability of Purchase	Financial Literacy Score	Hypothetical Financial Literacy Score	Insurance Literacy Score	Raven Progressive Matrices Score	CRT Score for Correct Answers	CRT Score for Heuristic Answers
Probability of Purchase							
Financial Literacy Score	0.21						
Hypothetical Financial Literacy Score	0.58	-0.22					
Insurance Literacy Score	0.27	0.56	0.39				
Raven Progressive Matrices Score	-0.40	-0.25	0.20	0.31			
CRT Score for Correct Answers	0.36	0.41	0.56	0.76	-0.03		
CRT Score for Heuristic Answers	-0.20	-0.07	-0.39	-0.62	-0.20	-0.79	

A. Terrible Decisions for Welfare

	Probability of Purchase	Financial Literacy Score	Hypothetical Financial Literacy Score	Insurance Literacy Score	Raven Progressive Matrices Score	CRT Score for Correct Answers	CRT Score for Heuristic Answers
Probability of Purchase							
Financial Literacy Score	-0.11						
Hypothetical Financial Literacy Score	0.06	-0.25					
Insurance Literacy Score	-0.13	0.23	-0.04				
Raven Progressive Matrices Score	-0.45	-0.24	0.33	0.15			
CRT Score for Correct Answers	-0.19	-0.51	0.17	0.04	0.32		
CRT Score for Heuristic Answers	-0.08	0.62	-0.17	0.20	-0.10	-0.89	

B. Great Decisions for Welfare

Note: Unconditional correlations at the level of a subject. For each financial literacy variable there are 119 observations, corresponding to one observation per subject. We classify terrible decisions as those that led to a realized CS below the 25^{th} centile (N=938 decisions) and great decisions as those that led to a realized CS above the 75^{th}) centile (N=893 decisions). The "Probability of Purchase" represents the fraction of the number of purchase decisions per subject where the subject decided to purchase insurance divided by the total number of possible purchase decisions per subject. Each subject was asked to make 54 decisions to purchase insurance or not.

Figure 10 provides the first, descriptive insights into addressing this question. It repeats the unconditional correlations from Figure 8, but after stratifying by the quality of the insurance decision. One must be careful here, since these are only correlations with purchases, and the "terrible decision" might have been to purchase or not purchase. Similarly, the "great decision" might have been to purchase or not purchase. All we know for now is that one refers to bad decisions and the other refers to good decisions.

However, we do have one very useful aggregate statistic that can help guide intuition here: the average purchase decisions. For all subjects the purchase decision was made 57% of the time. But for those making terrible decisions it was a purchase decision 83% of the time, and for those making great decisions it was a purchase decision only 26% of the time. So the stylized fact is that the terrible decisions tended to reflect excessive take-up, and the great decisions tended to reflect caution when deciding to purchase the product. Recalling panel C of Figure 7, 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.[29]

Looking at the terrible decisions, in panel A of Figure 10, we are now wary of anything that is leading to increased take-up, since that seems to be the behavioral pattern associated with such poor quality decisions. In the far left column we do observe that all of these factors[30] appear to be having a noticeable correlation with purchase. For the terrible decisions, we would then expect, and can verify, that these factors with positive correlations were all higher for subjects making terrible decisions, and those with negative correlations (especially Raven) were lower for subjects making terrible decisions.

The contrast with the great decisions, in panel B of Figure 10, is the point of real interest here. In this case we are now looking for clues as to why the behavioral pattern of *reduced* take-up arose. For the great decisions it is only the Raven score for fluid intelligence that is heavily correlated with purchase decisions, although there is a noticeable correlation

[29] A reminder 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).

[30] In the case of CRT scores, however, there is an offsetting effect from the CRT Score for Heuristic Answers that must be taken into account.

with Financial Literacy, Insurance Literacy, and the CRT Score for Correct Answers.[31] For the great decisions, we would then expect, and can verify, that these factors were all higher for subjects making great decisions.

Figure 11 displays the same distributions shown earlier in Figure 6, but stratifying according to whether the choices were terrible or great in terms of CS.[32] For Financial Literacy in panel A we observe that those that made great decisions had slightly lower scores, but were not significantly different. For Index Insurance Literacy in panel C those that made great decisions were, indeed, entirely in the upper modes, as conjectured above. In contrast, those that made terrible decisions tended to have better Raven scores of fluid intelligence. These comparisons point sharply to the role of domain-specific literacy as explaining why some decisions were great and other decisions were terrible.

4.2 Statistical Analysis

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

The core statistical model is a panel probit regression, recognizing that each subject contributed 54 purchase decisions. The unobserved heterogeneity of each subject is accommodated with a random effects specification, allowing us to examine the effects of characteristics that are associated with the individual (e.g., literacy measures, demographics). 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 average marginal effects of covariates of interest.[33]

[31] In the case of CRT scores, however, there is again an offsetting effect from the CRT Score for Heuristic Answers that must be taken into account.

[32] Since the literacy variables are defined for an individual, and these distributions are generated at the level of the decision, they reflected weighted distributions. Thus an individual with 25% of her decisions that were terrible would contribute 25% of her literacy score to the distribution for terrible decisions, and 75% of her literacy score to the distribution for great decisions.

[33] The demographics we control for are the respondents' age, 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

Figure 11: Distributions of the Literacy Variables,
for Purchase Decisions that are Terrible or Great for Welfare

Note: We classify terrible decisions as those that led to a realized CS below the 25^{th} centile (N=938 decisions) and great decisions as those that led to a realized CS above the 75^{th}) centile (N=893 decisions). "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.

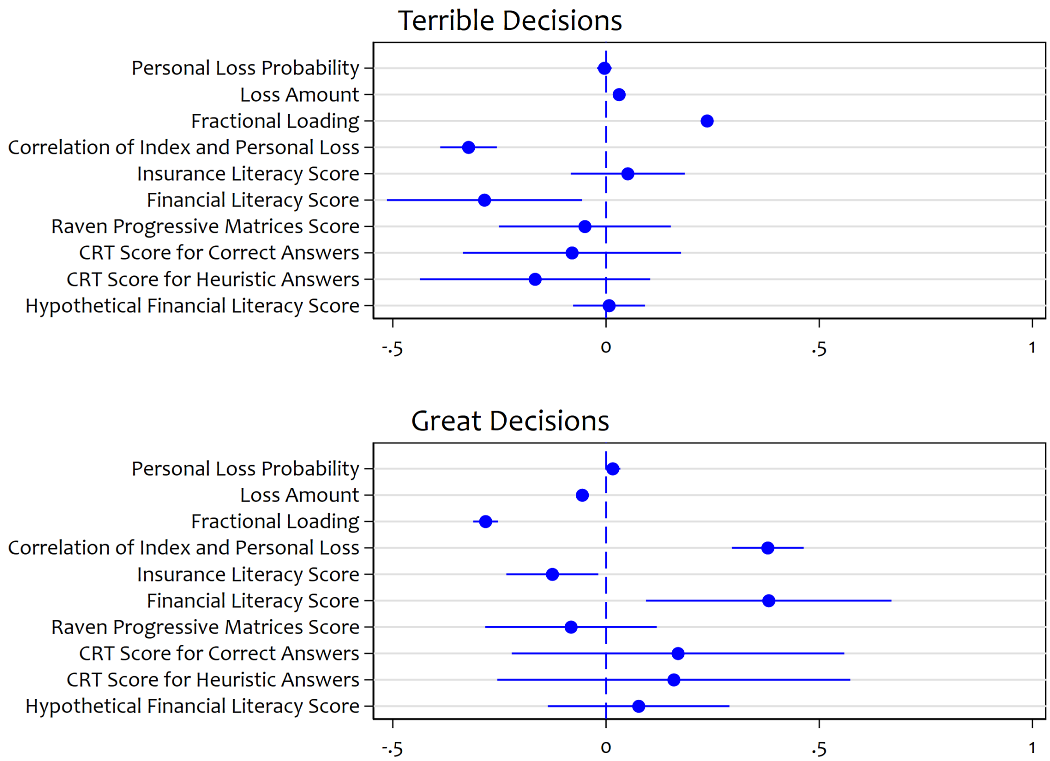
One regression model considers the purchase decisions that led to terrible welfare outcomes, and the other regression model considers the purchase decisions that led to great welfare outcomes. The point estimates of the average marginal effects of the actuarial and literacy covariates of interest, along with their 90% confidence interval, are displayed in Figure 12. For ease of comparison these variables have been normalized to fall in the interval between 0 and 1. The effect here, shown on the horizintal axis, is on the predicted probability of deciding to purchase the insurance product.

Focussing initially on the actuarial characteristics of the choice, we immediately observe some strange effects from those that made terrible decisions and some expected effects from those that made great decisions. In general, and assuming EUT, one would expect *a priori* that increases in the premium relative to the actuarially fair premium, would decrease purchases. In our specification this is captured by the fractional loading variable: greater values of the fractional loading mean higher premia, *ceteris paribus*. And higher loading factors are associated with statistically significant increases in the likelihood of purchasing insurance for those that made terrible purchase decisions, and exactly the opposite for those that made great purchase decisions. Similarly, one would expect *a priori* that improvements in the correlation between individual losses and the index loss would lead to increased purchases, since basis risk is being reduced in expectation. This is exactly what we observe for those making great purchase decisions, and exactly the opposite of what we observe for those making terrible purchase decisions. The personal loss probability and the loss amount have no significant effect on purchase probability, after controlling for the effects of loading and correlation.

Turning to the literacy measures, we find that improvements in general Financial Literacy leads to better, or less poor, decisions for both groups. For those that make terrible decisions, predominantly by excess take-up, greater Financial Literacy *reduces* take-up significantly. For those that make great decisions, predominantly by cautious take-up, greater Financial Literacy *increases* take-up significantly, and greater Insurance Literacy *decreases* take-up significantly. On the face of it, this last result might seem paradoxical, since

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 respondents' parents have a high or very high income (versus low or very low), and whether the respondent is Christian.

Figure 12: Effects on Purchase Probability of Actuarial and Literacy Variables, for Purchase Decisions that are Terrible or Great for Welfare



Note: Average marginal effects, and their 90% confidence intervals, of covariates of a panel probit regression of the purchase decision for a subsample of individuals who made terrible decisions in terms of CS, and great decisions in terms of CS. Each of the 119 subjects made 54 decisions. We classify terrible decisions as those that led to a realized CS below the 25^{th} centile (N=938 decisions) and great decisions as those that led to a realized CS above the 75^{th}) centile (N=893 decisions). We cluster standard errors at the session level, to correct for potential session-level heteroskedasticity. Demographic controls are the respondents' age, 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 respondents' parents have a high or very high income (versus low or very low), and whether the respondent is Christian.

"literacy" is pushing in different directions with respect to take-up. But "great decisions" 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.*[34] 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 general Financial Literacy helps the decision-maker decide on the former, and greater Insurance Literacy cautions on the latter. For the great decisions about index insurance, it is each *type* of literacy that matters, and for interestingly different ways when one examines the underlying actuarial characteristics of the choices.[35]

Although the effect is just not statistically significant, greater fluid intelligence appears to also caution those making great decisions to be cautious about purchasing.

In general, an important insight from these results is that the *incentivized* literacy measures have a much more important impact on the quality of insurance decision-making, both for welfare-improving and welfare-worsening decisions. The hypothetical survey responses, with the possible exception of the Raven score, provide no insight into why these are good quality decisions or bad quality decisions.

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. In principle, settlement can occur prior to any disaster, allowing mitigation of the personal loss and/or 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

[34] 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.

[35] This is consistent with the findings of Dercon et al. (2014) that the *content* of training sessions significantly affects take-up. They do not evaluate the quality of the extra take-up.

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 arena as well.

However there are challenges in 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 level of understanding of these products, what makes people make low-quality decisions in terms of their welfare, and eventually what role interventions might have. Our starting point was to understand how terrible decisions are made, and how great decisions are made, in a controlled setting in which we can measure all of the actuarial and behavioral "moving parts" behind those decisions. This allowed us to identify measures for general literacy, domain-specific literacy, cognitive reflection, and fluid intelligence. It also allowed us to control the key actuarial parameters, loss probabilities, personal loss amounts, and critically the correlation of personal and index losses.

We find that our incentivized literacy measures play an important role in identifying how great decisions were made, and that domain-specific literacy plays a particularly important role. This role interacted with responses to the usual actuarial parameters, helping understand perverse responses by poor decision-makers to loading factors and the correlation between personal and private loss. Hypothetical surveys of literacy provided no useful information. The next step is to see how these insights explain effects on the quality of decisions when we add policy interventions.

The final conclusion concerns a theme that lies at the heart of any rigorous evaluation of policy using the insights of behavioral welfare economics: how to judge if some policy is encouraging good decisions or bad decisions. One approach, which drives the popular "nudge" movement, is to assume that judgment away, and simply assert that some change in an observable must be good for all preferences and beliefs. Isn't it obvious that people should save more, eat less fatty foods, drink less wine, and take-up any insurance product on offer? Of course not: economics teaches us that the demand for these behaviors depends on preferences and beliefs. In fact, no sensibly realistic economics that recognizes heterogeneity of preferences and beliefs even tells us that this would be good behavior *on average*. And even if it is good behavior on average, are you willing to lash your policy advice to the social welfare function mast that this entails? Not clearly, and definitely not for us.

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.[36] And it repeatedly[37] comes to a major conclusion: that blindly encouraging take-up is an outright dangerous thing to do. Blind watchmakers[38] typically end up making a *lot* of terrible watches before they come across one good watch.

[36]See Harrison (2019, §6) for a full list, in the context of the behavioral welfare economics of insurance.
[37]See Harrison and Ng (2016, 2018, 2019).
[38]Dawkins (1986)

References

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

Attali, Y. and Bar-Hillel, M. (2020). The False Allure of Fast Lures. *Judgment and Decision Making*, 15(1):93–111.

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

Baron, J., Scott, S., Fincher, K., and Metz, S. E. (2015). Why Does the Cognitive Reflection Test (Sometimes) Predict Utilitarian Moral Judgment (and Other Things)? *Journal of Applied Research in Memory and Cognition*, 4(3):265–284.

Binswanger, H. P. (1980). Attitudes Toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, 62:395–407.

Camerer, C. F. and Hogarth, R. M. (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.

Campitelli, G. and Gerrans, P. (2014). Does the Cognitive Reflection Test Measure Cognitive Reflection? A Mathematical Modeling Approach. *Memory & Cognition*, 42(3):434–447.

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3):404.

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, `http://www.basis/wisc.edu`.

Carter, M. R. and Chiu, T. (2018). Quality Standards for Agricultural Index Insurance: An Agenda for Action. *The State of Microinsurance 2018*, `https://www.gbgfund.org/quality-standards-agricultural-index-insurance-agenda-action`.

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

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

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

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

Collins, D., Morduch, J., Rutherford, S., and Ruthven, O. (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.

de Haan, T. and Jona, L. (2018). "Good Nudge Lullaby": Choice Architecture and Default Bias Reinforcement. *The Economic Journal*, 128(610):1180–1206.

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

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

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

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

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

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

Harrison, G. W. (2007). Hypothetical Bias over Uncertain Outcomes. In List, J. A., editor, *Using Experimental Methods in Environmental and Resource Economics*. Northampton, MA: Elgar, 2007.

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

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

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

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

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

Harrison, G. W. and Ng, J. M. (2018). Welfare Effects of Insurance Contract Non-performance. *The Geneva Risk and Insurance Review*, 43(1):39–76.

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

Harrison, G. W. and Phillips, R. D. (2014). Subjective Beliefs and the Statistical Forecasts of Financial Risks: the Chief Risk Officer Project. In Andersen, T. J., editor, *Contemporary Challenges in Risk Management*. New York: Palgrave Macmillan.

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

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

Harrison, G. W. and Swarthout, J. T. (2016). Cumulative Prospect Theory in the Laboratory: A Reconsideration. *CEAR Working Paper 2016-05*. Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

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

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

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the Top of Mind: How Reminders Increase Saving. *Management Science*, 62(12):3393–3411.

Loomes, G. and Sugden, R. (1998). Testing Different Stochastic Specifications of Risky Choice. *Economica*, 65(260):581–598.

Lusardi, A. and Mitchell, O. S. (2007). Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth. *Journal of Monetary Economics*, 54(1):205–224.

Lusardi, A. and Mitchell, O. S. (2008). Planning and Financial Literacy: How do Women Fare? *American Economic Review, Papers & Proceedings*, 98(2):413–17.

Lusardi, A. and Mitchell, O. S. (2011). Financial Literacy and Planning: Implications for Retirement Well-Being. In Lusardi, A. and Mitchell, O. S., editors, *Financial Literacy: Implications for Retirement Security and the Financial Market Place* . New York: Oxford University Press.

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

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

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

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

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

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

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., and Hamilton, J. (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.

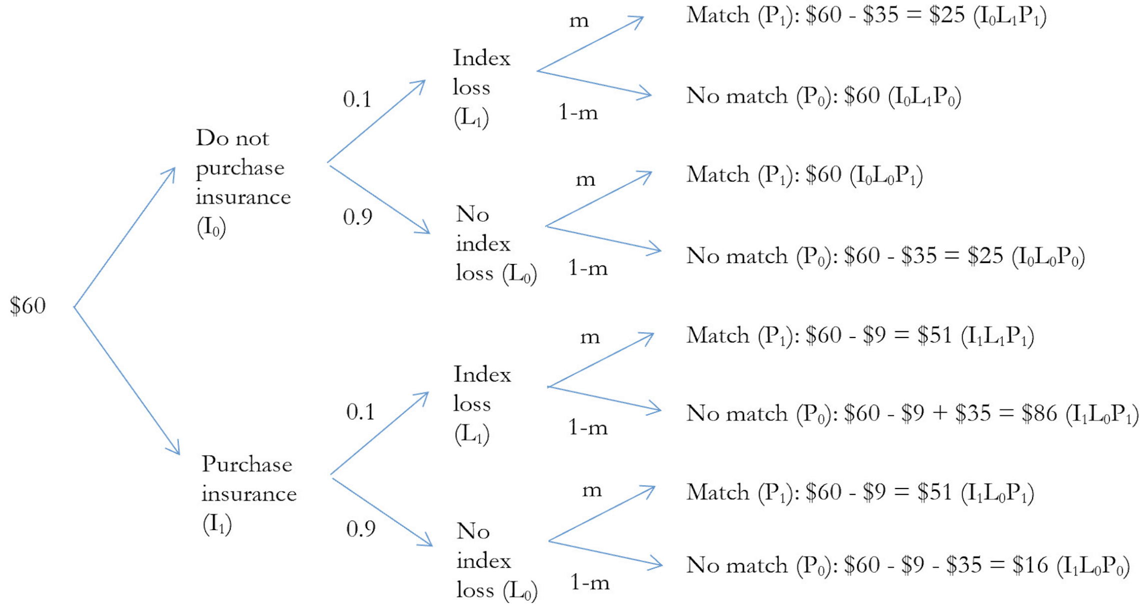
Raven, J., Raven, J., and Court, J. (1998). *Manual for Raven's progressive Matrices and Vocabulary Scales: Section 4, Advanced Progressive Matrices, Sets I & II*. San Antonio, TX: Pearson.

Schlesinger, H. and Schulenburg, J. G. v. (1987). Risk Aversion and the Purchase of Risky Insurance. *Journal of Economics (Zeitschrift für Nationalönomie)*, 47(3):309–314.

Stango, V. and Zinman, J. (2009). What do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs. *American Economic Review, Papers & Proceedings*, 99(2):424–429.

A Appendix A: Additional Figures

Figure A.1: Decision Tree for Index Insurance Product



B Appendix B: Instructions

B.1 Instructions for the Index Insurance Decisions

Choices Over Insurance Prospects

In this task you will make choices about whether to insure against possible monetary loss. In each choice you will start out with an initial amount of money and, in the event of a loss, the loss amount will be taken from this initial stake. In each choice you will have the option to buy insurance to protect you against the possible loss, although you are not required to buy the insurance.

You will make 54 choices in this task. You will actually get the chance to play one of the choices you make, and you will be paid in cash according to the outcome of that choice. So you should think carefully about how much each insurance choice is worth to you.

Each choice has two random events: a **Personal Event** and an **Index Event**. Each event has two possible outcomes: Good or Bad. If the Personal Event outcome is Bad, then you will suffer a loss. Before you know the outcome of the Personal Event, you must decide whether to purchase insurance against this possible loss. However, the insurance only pays a claim if the Index Event outcome is Bad.

If you do not purchase insurance, then only the outcome of the Personal Event will decide your earnings:

Personal Event	Your Earnings
Bad	Initial stake - Loss
Good	Initial stake

If you do purchase insurance, it is important for you to understand that an insurance claim is not paid according to whether you actually suffer a loss. Instead, an insurance claim is paid only according to the Index Event. Both events will decide your earnings:

So there are four possible outcomes if you purchase insurance. You might suffer a

Personal Event	Index Event	Your Earnings
Bad	*Bad*	Initial stake - Insurance cost – Loss + Insurance coverage
Bad	*Good*	Initial stake - Insurance cost – Loss
Good	*Good*	Initial stake - Insurance cost
Good	*Bad*	Initial stake - Insurance cost + Insurance coverage

loss and receive an insurance claim payment. Or you might suffer a loss but not receive an insurance claim payment. You might not suffer a loss and also receive no insurance claim payment. Finally, you might receive an insurance claim payment even when you do not suffer a loss.

Each event is determined by randomly drawing a colored chip from a bag. In general, each draw will involve two colors, and each decision you make will involve different amounts and mixtures of two colors. When making each decision, you will know the exact amounts and mixtures of colored chips associated with the decision. After you have decided whether or not to purchase insurance, the two events will be determined as follows.

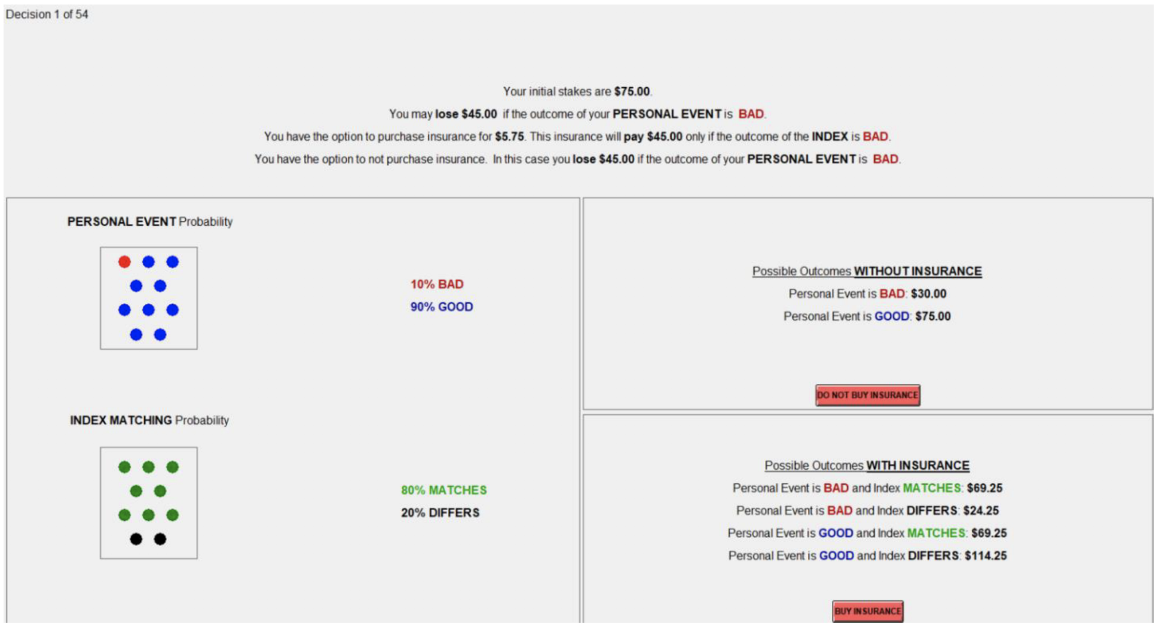
First, the Personal Event will be determined with blue and red chips.

- If you draw a **blue** chip, then the Personal Event outcome is Good and you do not suffer a loss.
- If you draw a **red** chip, then the Personal Event outcome is Bad and you suffer a loss.

Next, if you purchased insurance, the Index Event will be determined with green and black chips.

- If you draw a **green** chip, then the Index Event outcome Matches the Personal Event outcome
- If you draw a **black** chip, then the Index event outcome Differs from the Personal Event outcome

Here is an example of what your decision would look like on the computer screen. The display on your screen will be bigger and easier to read.



In this example you start out with an initial stake of $75. If the outcome of the Personal Event is Bad you will lose $45, and if the outcome of the Personal Event is Good you will not lose any money. If you faced the choice in this example and chose to purchase insurance, you would pay $5.75 from your initial stake. You would pay this $5.75 before you drew any chips, so you would pay it regardless of the outcomes of your draws.

You will be drawing colored chips from bags to determine the outcomes of both events. First, you will draw a chip to determine the Personal Event outcome. The image on the left shows that there is a 10% chance that the Personal Event outcome is Bad, and a 90% chance that the Personal Event outcome is Good. This means there will be 9 blue chips and 1 red chip in a bag, and the color of the chip you randomly draw from the bag represents the outcome of the Personal Event. If a blue chip is drawn, the Personal Event outcome is Good, and if a red chip is drawn the Personal Event outcome is Bad.

Next, you will draw a chip to determine the Index Event outcome. There is an 80% chance that the Index Event outcome Matches the Personal Event outcome and a 20% chance that the Index Event outcome Differs from the Personal Event outcome. This means there will be 8 green chips and 2 black chips in a bag. If a green chip is drawn the Index

Event outcome Matches the Personal Event outcome, and if a black chip is drawn the Index Event outcome Differs from the Personal Event outcome.

You will indicate your choice to purchase, or not purchase, the insurance by clicking on your preferred option on the computer screen.

There are 54 decisions like this one to be made, each shown on a separate screen on the computer. Each decision might have different chances for the Personal Event outcome, the Index Event outcome, the initial stake, or the cost of insurance, so pay attention to each screen. After you have worked through all of the insurance decisions, please wait in your seat and an experimenter will come to you. You will then roll two 10-sided dice to determine which insurance decision will be played out. Since there are only 54 decisions, you will keep rolling the dice until a number between 1 and 54 comes up. There is an equal chance that any of your 54 choices will be selected, so you should approach each decision as if it is the one that you will actually play out to determine your payoff. Once the decision to play out is selected, you will draw chips from the Index bag and the Personal bag to determine the outcome.

In summary:

- You will decide whether or not to purchase insurance in each of the 54 scenarios.
- One of your decisions will be randomly selected to be played for cash.
- You will suffer the specified monetary loss only if the Personal Event outcome is Bad.
- If you purchase insurance, it will pay a claim payment only if the Index Event outcome is Bad. This can happen in two ways:
    1. Your Index draw Matches a bad Personal Event outcome;
    2. Your Index draw Differs from a good Personal Event outcome.

Whether or not you prefer to buy the insurance is a matter of personal taste. You may choose to buy insurance on some or all of your 54 choices, or none of the choices. The people next to you may be presented with different choices, insurance prices, and may have different preferences, so their responses should not matter to you. Please work silently, and make your choices by thinking carefully about each prospect.

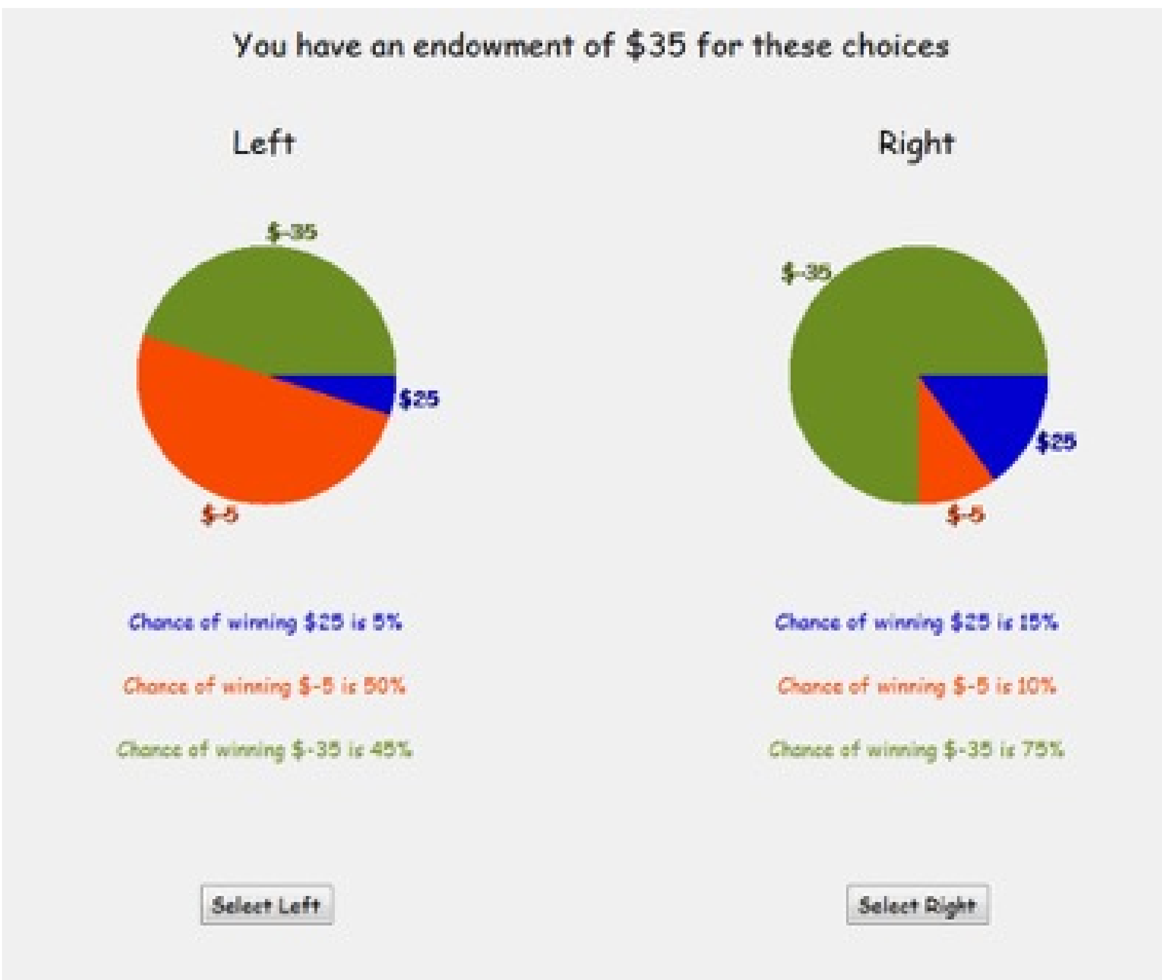
Your payoff from this task is in cash and is in addition to the show-up payment that you receive just for being here, as well as any other earnings in other tasks. If you have a question, raise your hand and someone will come over and answer it.

B.2 Instructions for the risk preferences elicitation

Choices Over Risky Prospects

This is a task where you will choose between prospects with varying prizes and chances of winning each prize. You will be presented with a series of pairs of prospects where you will choose one of them. For each pair of prospects, you should choose the prospect you prefer. You will actually get the chance to play one of these prospects for earnings, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer on each decision screen.

Here is an example of what the computer display of such a pair of prospects will look like.



The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

You might be told your cash endowment for each decision at the top of the screen. In this example it is $35, so any earnings would be added to or subtracted from this endowment. The endowment may change from choice to choice, so be sure to pay attention to it. The endowment you are shown only applies for that choice.

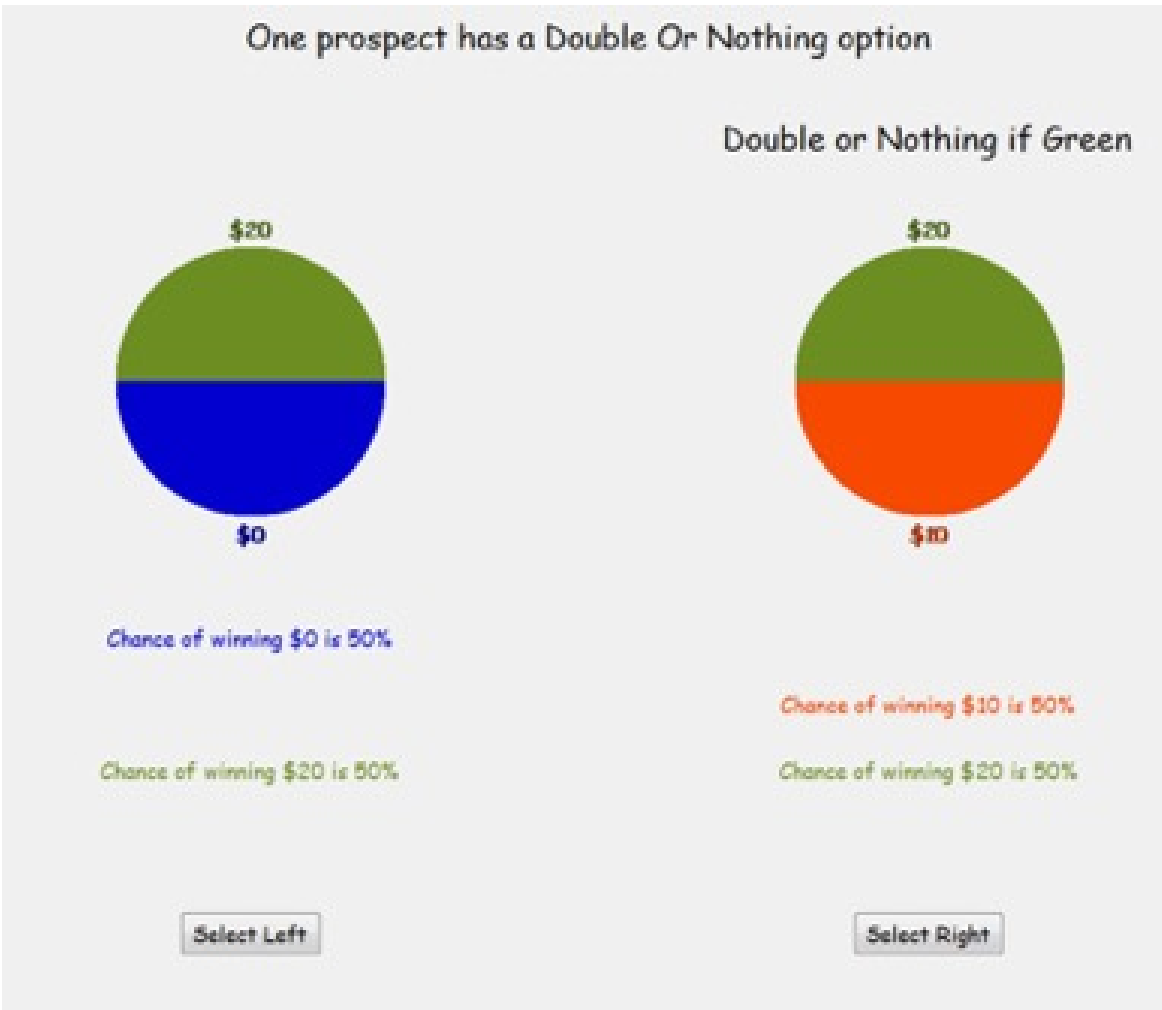
In this example the left prospect pays twenty-five dollars ($25) if the number drawn is between 1 and 5, pays negative five dollars ($-5) if the number is between 6 and 55, and pays negative thirty-five dollars ($-35) if the number is between 56 and 100. The blue color in the pie chart corresponds to 5% of the area and illustrates the chances that the number drawn will be between 1 and 5 and your prize will be $25. The orange area in the pie chart corresponds to 50% of the area and illustrates the chances that the number drawn will be between 6 and 55 and your prize will be $-5. The green area in the pie chart corresponds to 45% of the area and illustrates the chances that the number drawn will be between 56 and 100. When you select the decision screen to be played out the computer will confirm the die rolls that correspond to the different prizes.

Now look at the pie on the right. It pays twenty-five dollars ($25) if the number drawn is between 1 and 15, negative five dollars ($-5) if the number is between 16 and 25, and negative thirty-five dollars ($-35) if the number is between 26 and 100. As with the prospect on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $25 pie slice is 15% of the total pie.

Even though the screen says that you might win a negative amount, this is actually a loss to be deducted from your endowment. So if you win $-5, your earnings would be $\$30 = \$35 - \$5$.

Each pair of prospects is shown on a separate screen on the computer. On each screen, you should indicate which prospect you prefer to play by clicking on one of the buttons beneath the prospects.

Some decision screens could also have a pair of prospects in which one of the prospects will give you the chance for “Double or Nothing”. For instance, the right prospect in this screen image pays "Double or Nothing" if the Green area is selected, which happens if the number drawn is between 51 and 100. The right pie chart indicates that if the number is between 1 and 50 you get $10. However, if the number is between 51 and 100 we will flip a coin with you to determine if you get either double the amount or $0. In this example, if it comes up Heads you get $40, otherwise you get nothing. The prizes listed underneath each pie refer to the amounts before any "Double or Nothing" coin toss.



After you have worked through all of the pairs of prospects, please wait quietly until further instructions. When it is time to play this task out for earnings, you will then roll two 10-sided dice until a number comes up to determine which pair of prospects will be played out. If there are 40 pairs we will roll the dice until a number between 1 and 40 comes up, if there are 80 pairs we will roll until a number between 1 and 80 comes up, and so on. Since there is a chance that any of your choices could be played out for real, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will roll the two ten-sided dice to determine the outcome of the prospect you chose, and if necessary we will then toss a coin to determine if you get "Double or Nothing".

Here is an example: suppose your first roll was 81. We would then pull up the 81st decision that you made and look at which prospect you chose – either the left one or the right one. Let's say that the 81st lottery was the same as the last example, and you chose the left prospect. If the random number from your second roll was 37, you would win $0; if it was 93, you would get $20.

If you picked the prospect on the right and drew the number 37, you would get $10; if it was 93, we would have to toss a coin to determine if you get "Double or Nothing". If the coin comes up Heads then you would get $40. However, if it comes up Tails you would get nothing from your chosen prospect.

It is also possible that you will be given a prospect in which there is a "Double or Nothing" option no matter what the outcome of the random number. This screen image illustrates this possibility.

In summary, your payoff is determined by five things:

- by your endowment, if there is one, shown at the top of the screen;
- by which prospect you selected, the left or the right, for each of these pairs;
- by which prospect pair is chosen to be played out in the series of pairs using the two 10-sided dice;
- by the outcome of that prospect when you roll the two 10-sided dice; and
- by the outcome of a coin toss if the chosen prospect outcome is of the "Double or Nothing" type.

Which prospects you prefer is a matter of personal choice. The people next to you may be presented with different prospects, and may have different preferences, so their responses should not matter to you or influence your decisions. Please work silently, and make your choices by thinking carefully about each prospect.

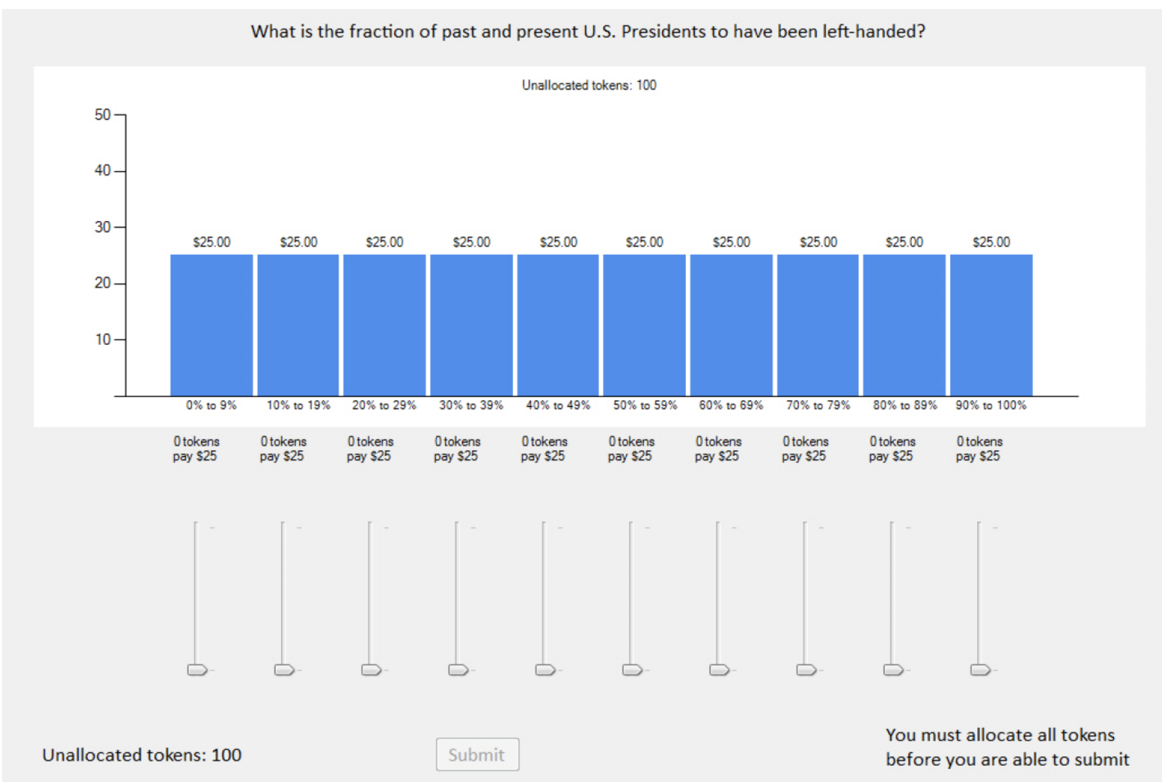
All payoffs are in cash, and are in addition to the $5 show-up fee that you receive just for being here, as well as any other earnings in other tasks from the session today.

B.3 Instructions for the Financial Literacy Beliefs

Instructions

This is a task where you will be paid according to how accurate your beliefs are about certain things. You will be presented with some questions and asked to place some bets on your beliefs about the answers to each question. You will be rewarded for your answer to one of these questions, so you should think carefully about your answer to each question. The question that is chosen for payment will be determined after you have made all decisions, and that process is explained below.

Here is an example of what the computer display of a question might look like. We pick a question that is not going to be asked of you, just for illustration.

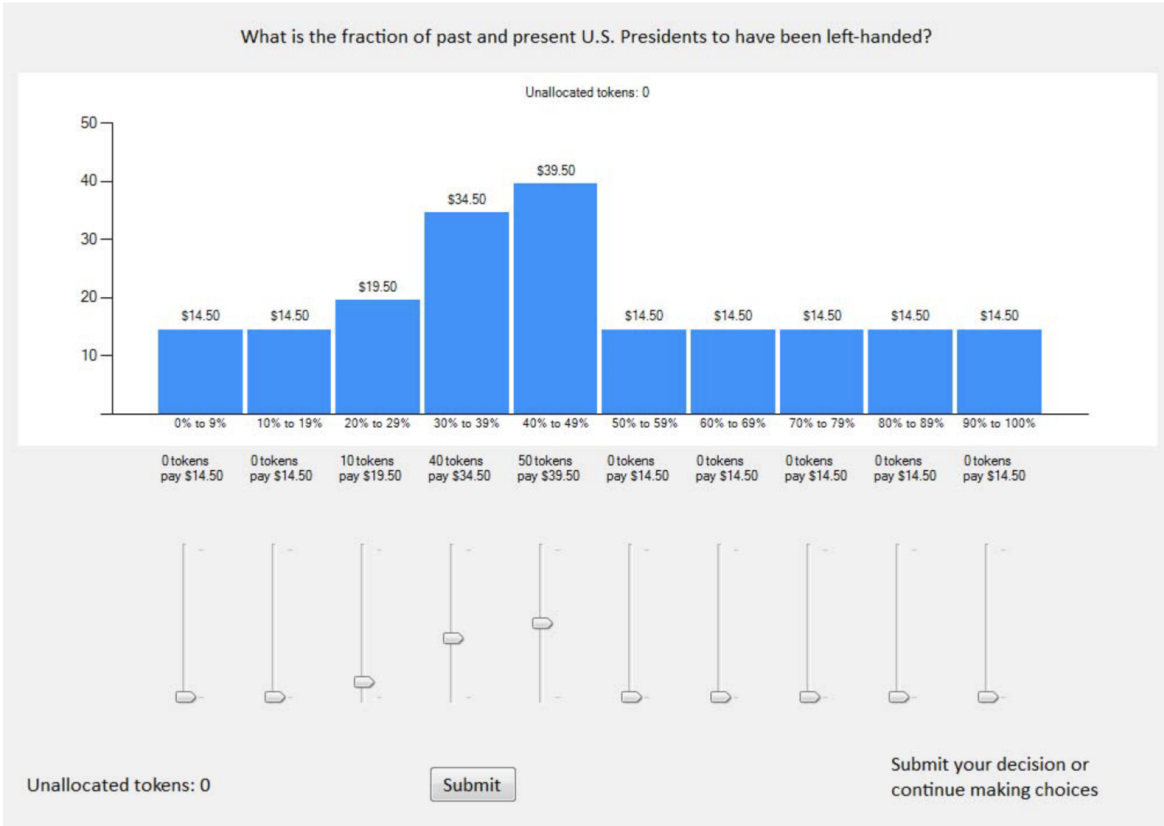


The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate across the sliders. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens, and in this example we start with 0 tokens allocated to each slider. As you allocate tokens, by adjusting sliders, the payoffs displayed on the screen will change. Your earnings are based on the payoffs that are displayed after you have allocated all 100 tokens.

You can earn up to $50 in this task.

Where you position each slider depends on your beliefs about the correct answer to the question. Note that the bars above each slider correspond to that particular slider. In our example, the tokens you allocate to each bar will naturally reflect your beliefs about the proportion of left-handed Presidents. The first bar corresponds to your belief that the proportion is between 0% and 9%. The second bar corresponds to your belief that the proportion is between 10% and 19%, and so on. Each bar shows the amount of money you could earn if the true proportion is in the interval shown under the bar.

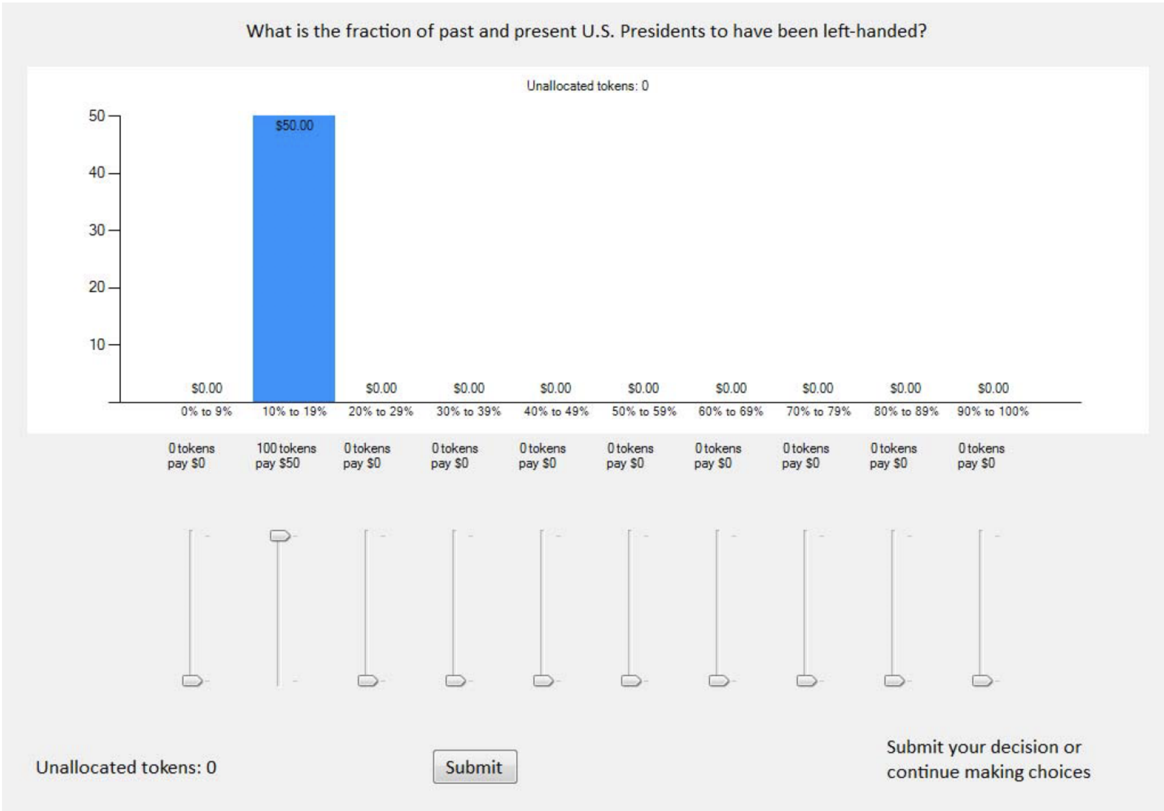
To illustrate how you use these sliders, suppose you think there is a fair chance the true answer is just under 50%. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 40% to 49%, 40 tokens to the interval 30% to 39%, and 10 tokens to the interval 20% to 29%. So you can see in this picture that if indeed the proportion of left-handed Presidents is between 40% and 49% you would earn $39.50. You would earn less than $39.50 for any other outcome. You would earn $34.50 if the proportion of left-handed Presidents is between 30% and 39%, $19.50 if it is between 20% and 29%, and for any other proportion you would earn $14.50.



You can adjust the allocation as much as you want to best reflect your personal beliefs about the proportion of left-handed Presidents.

Your earnings depend on your reported beliefs and, of course, the true answer. For instance, suppose you allocated your tokens as just shown. The true answer is that there are 8 left-handed Presidents out of 44, so the true proportion is 18.2%, and we would round that to 18%. So if you had reported these beliefs, you would have earned $14.50.

Suppose you had put all of your eggs in one basket, and allocated all 100 tokens to the interval corresponding to a proportion between 10% and 19%. Then you would have faced the earnings outcomes shown here:



Note the "good news" and "bad news" here. Since the proportion of left-handed Presidents is between 10% and 19%, you earn the maximum payoff, shown here as $50. But if the true proportion had been 20%, you would have earned nothing in this task.

It is up to you to balance the strength of your personal beliefs with the possibility of them being wrong. There are three important points for you to keep in mind when making your decisions:

- First, your belief about the correct answer to each question is a personal judgment that depends on the information you have about the topic of the question.
- Second, depending on your choices and the correct answer you can earn up to $50.
- Third, your choices might also depend on your willingness to take risks or to gamble.

The decisions you make are a matter of personal choice. Please work silently, and make your choices by thinking carefully about the questions you are presented with.

When you are satisfied with your decisions, you should click on the Submit button and confirm your choices. When you are finished we will roll dice to determine which question will be played out. The experimenter will record your earnings according to the correct answer and the choices you made.

All payoffs are in cash, and are in addition to the show-up fee that you receive just for being here, as well as any other earnings in the session today.

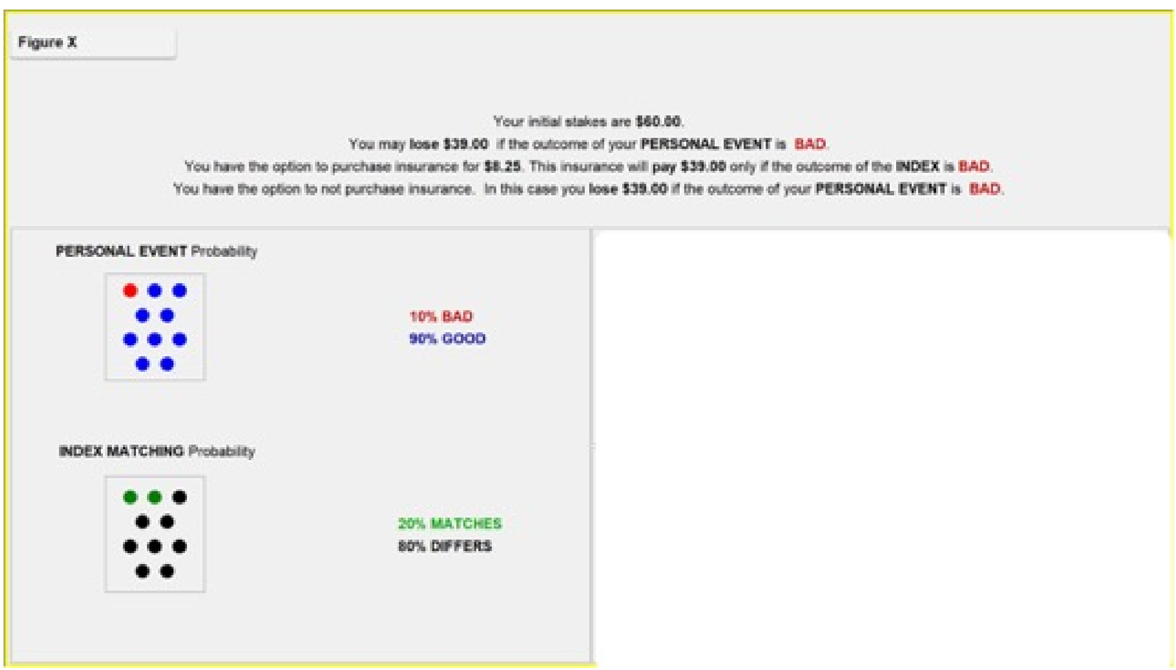
We will now have a video demonstration of how you make decisions in this task, using the same hypothetical example. You can then raise your hand if you need more explanation, or replay these instructions. The actual questions we will ask you to state your beliefs about will be presented after we explain the next task. Here is the video demonstration...

B.4 Instructions for the Index Insurance Literacy Beliefs

Beliefs over Potential Outcomes, Monetary Endowments, and Insurance Purchases

In this task you will be presented with possible outcomes of monetary endowments that are exposed to random events that can be protected through insurance or not. In this task you will be paid according to how accurate your beliefs are about the outcomes of these endowments, shocks, and insurance purchases. In this task you will not be asked to make any insurance purchases yourself. You will be presented with some questions and asked to place some bets on your beliefs about the answers to each question. You will be rewarded for your answer to one of these questions, so you should think carefully about your answer to each question. The question that is chosen for payment will be determined after you have made all decisions, and that process is explained below.

Here is an example of the outcomes that will be presented to you on paper. The paper display will be bigger and easier to read. From the top left corner, note that this display refers to Figure X.

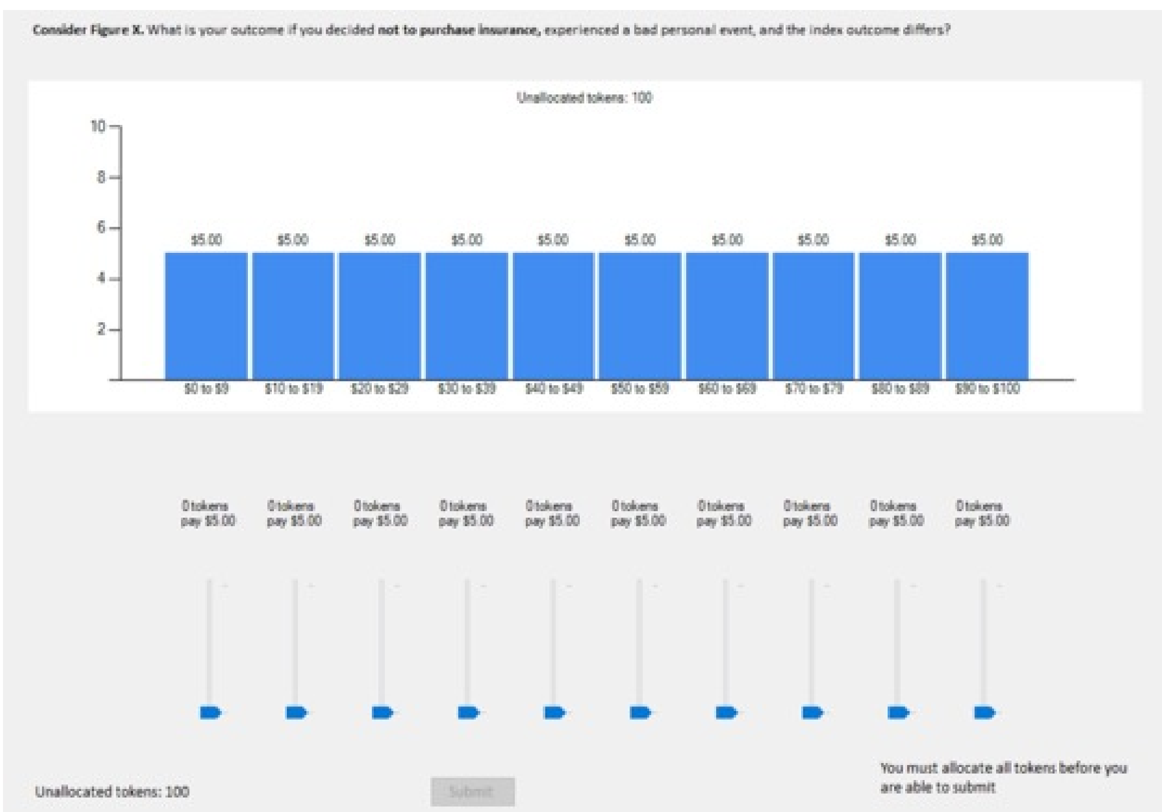


We will ask you 10 questions about potential outcomes of initial stakes, the personal event and the index event, and the effect of purchasing insurance or not.

Examples of the questions we might ask you are:

- **Consider Figure X. What is your outcome if you decided not to purchase insurance, experienced a good personal event, and the index outcome matches?** There will be 10 categories of possible answers: $0-$9; $10-$19; $20-$29; $30-$39; $40-$49; $50-$59; $60-$69; $70-$79; $80-$89; $90-$100.
- **Consider Figure X. What is your outcome if you decided to purchase insurance, experienced a bad personal event, and the index outcome differs?** There will be 10 categories of possible answers: $0-$9; $10-$19; $20-$29; $30-$39; $40-$49; $50-$59; $60-$69; $70-$79; $80-$89; $90-$100.

We will then ask you to place some bets on your beliefs about the answers to each question.

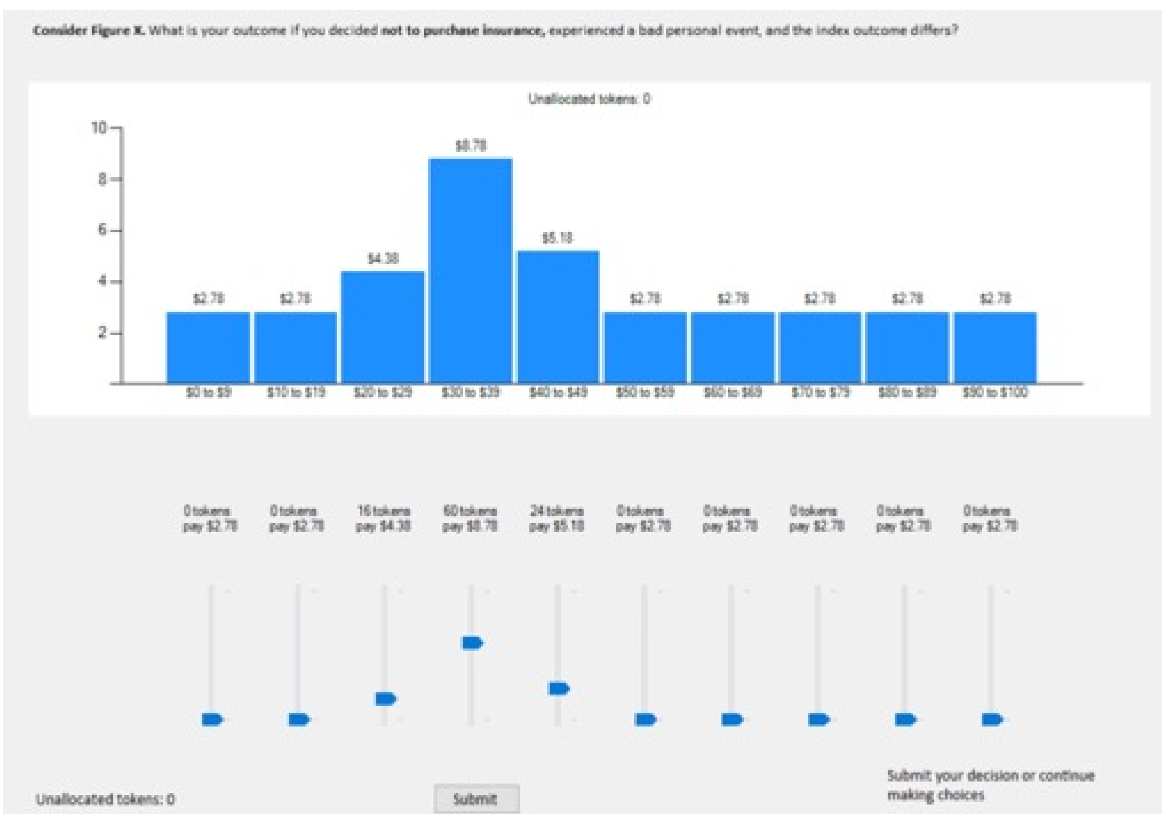


Here is an example from the first question shown above. You will probably recall this task from a previous session. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate across the sliders. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens, and in this example we start with 0 tokens allocated to each slider. As you allocate tokens, by adjusting sliders, the payoffs displayed on the screen will change. Your earnings are based on the payoffs that are displayed after you have allocated all 100 tokens.

You can earn up to $10 in this task.

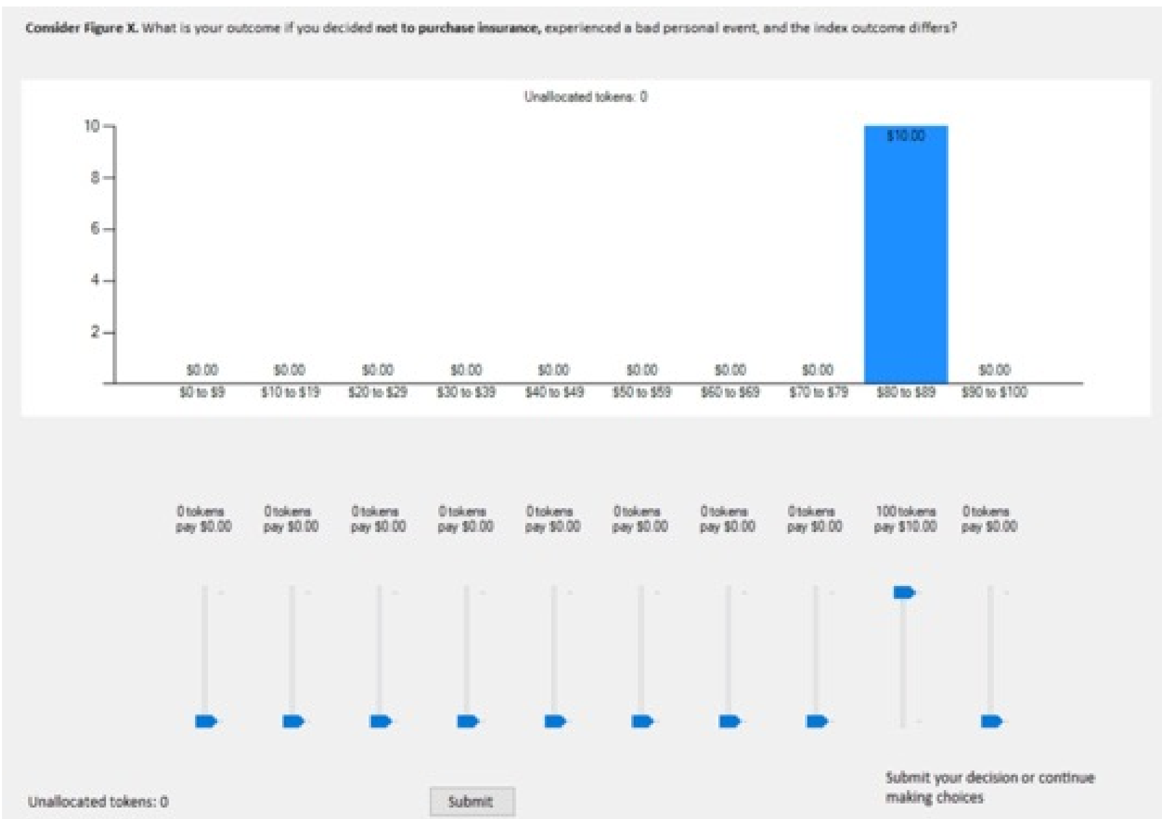
Where you position each slider depends on your beliefs about the correct answer to the question. Note that the bars above each slider correspond to that particular slider. Each bar shows the amount of money you could earn if the true outcome is in the interval shown under the bar.

To illustrate how you use these sliders, you might make these choices:



Your earnings depend on your reported beliefs and, of course, the true answer. For instance, suppose you allocated you tokens as just shown. If the true answer had been $13, you would have earned $2.78. If the true answer had been $33, you would have earned $8.78, and so on. If the true answer contains cents, we will round it to the nearest whole dollar. So if the true answer had been $13.26, we would have rounded it to $13.

Suppose you had put all of your eggs in one basket, and allocated all 100 tokens to the interval corresponding to an outcome between $80 and $89. Then you would have faced the earnings outcomes shown here:



Note the "good news" and "bad news" here. If the true answer had been $85, you would have earned $10. But if the true answer had been $15, or even $79, you would have earned nothing in this task.

It is up to you to balance the strength of your personal beliefs with the possibility of them being wrong. There are three important points for you to keep in mind when making your decisions:

- First, your belief about the correct answer to each question is a personal judgment that depends on the information you have about the topic of the question.
- Second, depending on your choices and the correct answer you can earn up to $10.
- Third, your choices might also depend on your willingness to take risks or to gamble.

The decisions you make are a matter of personal choice. Please work silently, and make your choices by thinking carefully about the questions you are presented with.

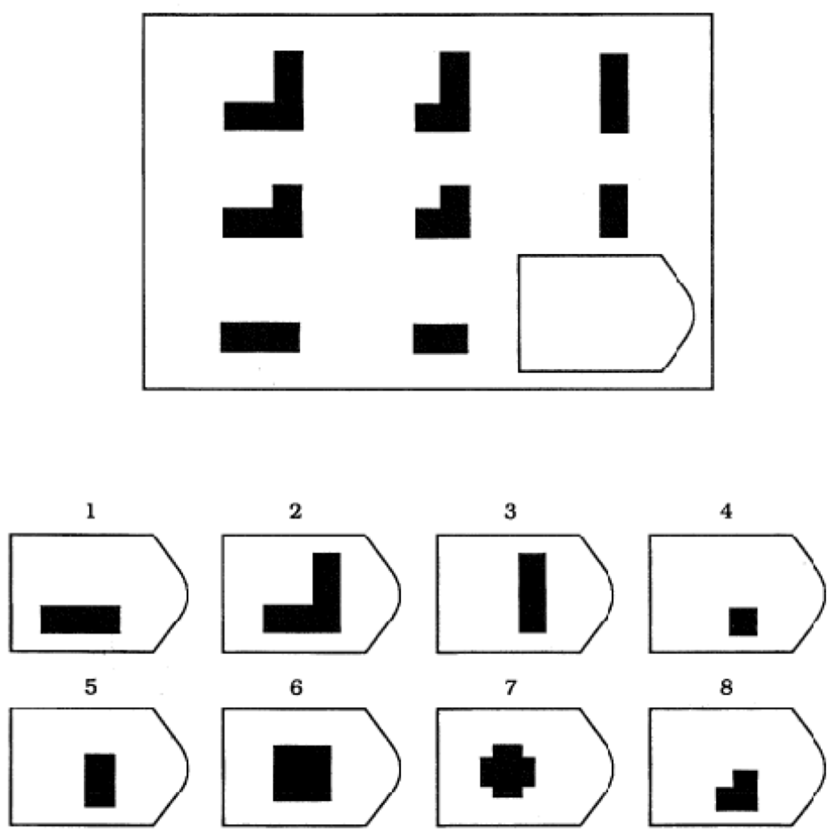
When you are satisfied with your decisions, you should click on the Submit button and confirm your choices. When you are finished we will roll dice to determine which question will be played out. The experimenter will record your earnings according to the correct answer and the choices you made. All payoffs are in cash, and are in addition to the show-up fee that you receive just for being here, as well as any other earnings in the session today.

B.5 Instructions for the Raven Advanced Progressive Matrices Test

Matrices Task

The first task today will involve completing a series of 12 matrices. For each item, you will be asked to identify the missing element that completes a pattern. Each matrix consists of 8 visual geometric designs with a missing 9th piece in the bottom right corner. You are asked to identify the missing element that completes the pattern from eight choices provided.

The figure below is representative of a matrix that you will encounter. You will notice in the bottom-right portion of the top figure that there is a section cut out of it. Look at the pattern in that top figure and think what piece is needed to complete the pattern correctly. Then find the right piece of out of the eight choices shown below.



Only one of these pieces is correct. In this case, number 8 completes the diagonal, but not going downwards or across. We see that each geometric shape is getting smaller, whether going down, across, or diagonally. Can you see the choice that completes the matrix? If you said Number 4, you are correct.

You will be presented with 12 of these matrices and asked to find the missing piece in each. You may find the problems easy at first and then becoming more difficult. Please work quietly and independently until all 12 matrices are completed.

B.6 Questions and Correct and Heuristic Answers Cognitive Reflection Test

Table B.1: Cognitive Reflection Test

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? *(correct answer: 5 cents; heuristic answer: 10 cents)*	Frederick (2005)
If it takes 5minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets? *(correct answer: 5 minutes; heuristic answer: 100 minutes)*	Frederick (2005)
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? *(correct answer: 47 days; heuristic answer: 24 days)*	Frederick (2005)
If three elves can wrap three toys in hour, how many elves are needed to wrap six toys in 2 hours? *(correct answer: 3 elves; heuristic answer: 6 elves)*	Primi et al. (2016)
Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class? *(correct answer: 29 students; heuristic answer: 30 students)*	Primi et al. (2016)
In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes? *(correct answer: 15 medals; heuristic answer: 20 medals)*	Primi et al. (2016)

B.7 Questions and Correct Answers Hypothetical Financial Literacy Questions

Correct answers listed in **bold**

1. An investment with a high return is likely to be high risk.
    - **True**
    - False
    - Don't know
    - Refuse to answer
2. High inflation means that the cost of living is increasing rapidly.
    - **True**
    - False
    - Don't know
    - Refuse to answer
3. It is usually possible to reduce the risk of investing in the stock market by buying a wide range of stocks and shares.
    - **True**
    - False
    - Don't know
    - Refuse to answer
4. If interest rates rise, what will typically happen to bond prices?
    - They will rise
    - **They will fall**
    - They will stay the same
    - There is no relationship between bond prices and the interest rate
    - Don't know

- Refuse to answer

5. A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.
    - **True**
    - False
    - Don't know
    - Refuse to answer

6. Suppose you owe $1,000 on a loan and the interest rate you are charged is 20% per year compounded annually. If you didn't pay anything off, at this interest rate, how many years would it take for the amount you owe to double?
    - Less than 2 years
    - **2 to 4 years**
    - 5 to 9 years
    - 10 or more years
    - Don't know
    - Refuse to answer

7. Suppose that you receive $20 in interest on your savings account during a year. At the same time, prices increase so that something that cost $100 at the beginning of the year now costs $120 at the end of the year. What will the $20 in interest earnings allow you to do?
    - Buy more things at the end of the year than I did at the beginning of the year
    - **Buy the same things at the end of the year that I did at the beginning of the year**
    - Buy fewer things at the end of the year than I did at the beginning of the year
    - Don't know
    - Refuse to answer

8. Which of the following is an accurate statement about investment returns?

- Usually, investing \$5,000 in shares of a single company is safer than investing \$5,000 in a fund which invests in shares of many companies in different industries
- **Usually, investing \$5,000 in shares of a single company is less safe than investing \$5,000 in a fund which invests in shares of many companies in different industries**
- Usually, investing \$5,000 in shares of a single company is equally as safe as investing \$5,000 in a fund which invests in shares of many companies in different industries
- Don't know
- Refuse to answer

9. Suppose you are a member of a stock investment club. This year, the club has about \$200,000 to invest in stocks and the members prefer not to take a lot of risk. Which of the following strategies would you recommend to your fellow members?
    - Put all of the money in one stock
    - Put all of the money in two stocks
    - **Put all of the money in a stock index fund that tracks the behavior of 500 large firms in the USA**
    - Don't know
    - Refuse to answer