Studies in Risk Perception and Financial Literacy: Applications Using Subjective Belief Elicitation

Ph.D. Dissertation

Mark Schneider
School of Economics
University of Cape Town
Rondebosch 7701
Cape Town
South Africa

June 2019

Supervisor:

- Don Ross – University of Cape Town and University College of Cork

Supervisory Committee Members:

- Glenn W. Harrison – Georgia State University and University of Cape Town
- E. Elisabet Rutström – Örebro University, Stockholm School of Economics, and University of Cape Town

Declaration:

I declare that this dissertation presented for the degree of Doctor of Philosophy in Economics, was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Signed by candidate

Mark Patrick Schneider

15 June 2019

Chapter 1 – Risk Perception and Financial Literacy

1.1 Motivation & Introduction

The financial environment that consumers face today has become more complex, and has expanded dramatically, in just a generation. Over the past 20 years consumers have been given access to exotic mortgages and financial derivative product offerings, as well as increased access to credit and borrowing options. The baby boomers[1] in the U.S. have had a front row seat to this changing financial environment. At the same time there has been a five-fold increase in personal bankruptcy filings in the US in the last 30 years (White 2009). Boomers have also experienced a dramatic change in the options available to build savings for retirement compared to mechanisms enjoyed by their parents. The defined benefit pensions of the boomers' parents are being replaced by voluntary defined contribution retirement systems, such as 401(k) and 403 (b) offerings. This transition simplified the balance sheets of employers, but shifted the cognitive burden to employees to decide how much to save, where to invest, and how to make lump sum payouts or annuity streams last throughout retirement (McKenzie and Liersch 2011; Yoong 2011).

With the shift in retirement planning, how are individuals managing these costly and risky processes themselves? In many households, not well; one study reports that about one-third of adults in their 50s have done little or no planning for retirement (Lusardi 2003). Further studies argue that poor planning for retirement, low levels of wealth accumulation (Lusardi and Mitchell 2007), and a lack of participation in stock markets (Yoong 2011) may be a direct result of financial illiteracy.

Many academics, experts, and politicians, observing the changing landscape of financial planning and decision making coupled with low rates of financial literacy, have prescribed the same remedy: increased financial literacy and financial education. In fact, in 2010 President Obama declared the month

---

[1] The Oxford English Dictionary (Third Edition) defines a baby boomer as "a person born during the baby boom following the Second World War," and is commonly viewed as those children born between 1946 and 1964. The baby boomer generation makes up nearly 20% of the American population.

of April as *National Financial Literacy Month*, dedicated "to teaching ourselves and our children about the basics of financial education" (Johnson 2010).

There is a clear need to study financial literacy. However, there remain fundamental methodological issues about how to define, classify, and measure financial literacy that the current literature has not coherently addressed. This thesis fills that niche, presents a clear definition of financial literacy, and demonstrates how to measure it using a rigorous framework. That framework is then deployed in a lab and an artefactual field setting to measure financial literacy at various levels of agency: an individual decision-maker, a group of two individual decision-makers, and decision-makers in a naturally formed household. Further, we study the effect that having access to the Internet has on the financial literacy of individual decision-makers.

The conceptual framework for the elicitation of measures of literacy is a set of procedures to elicit subjective belief distributions from individuals. These procedures have a theoretical basis, and have been applied in recent literature. The conceptual framework for the treatments, generating hypothesized changes in literacy, is the broad literature on "scaffolding." That literature spans philosophy, developmental learning, psychology, artificial intelligence, and cognitive science.

The remainder of this chapter is organized as follows: Section 1.2 provides an introduction to the concept of financial literacy, differentiates financial literacy from financial capability, provides a brief survey of the literature, and introduces an applied experimental method of eliciting individuals' risk attitudes and subjective beliefs in order to measure financial literacy; Section 1.3 reviews the methodology behind subjective belief elicitation and operationalizes the framework used in this research; Section 1.4 introduces some hypotheses and proposed applications using the elicitation methodology in both laboratory and field settings; and Section 1.5 concludes.

1.2 Literature Review

1.2.1 What is Literacy?

The *Oxford English Dictionary* defines the adjective "literate" as someone who is "acquainted with letters or literature; erudite, learned, lettered." Here this *literal* literacy is the ability to read written text. However, the idea of literacy has grown to include domain specific topics and skill sets: health literacy, computer literacy, statistical literacy, and financial literacy (Lusardi and Mitchell 2007, 2008; Di Girolamo et al. 2015; Huston 2010) as well as other domains. These extensions can be collectively viewed as *metaphorical* literacy, each referring to a particular cognitive competence in a particular domain.

Agencies of Literacy

Literacy can be considered at different levels of agency, which begin with an individual decision-maker. When one starts to consider the combined literacy of $n \geq 2$ individual decision-makers in a group or household, which we will call "effective literacy," there are two important new dimensions of literacy that need to be defined.

The first dimension of literacy associated with thinking about the literacy of groups or households is what economists refer to as "common knowledge." In game theory this refers to the idea that each member of the group, or each player in the game, shares some knowledge of the game with the other player and, further, that this knowledge survives an infinite regress. In other words, I know that you know that I know that you know…the facts about the game. Clearly, knowledge is a logically demanding concept, but it is also suggestive of a weaker concept which addresses the extent to which individuals in a group share the same understanding of some fact. This topic has been studied in many disciplines, apart from economics and game theory. For example, it is a major issue in linguistics, where the notion of "semantics" refers to the meaning of a word, phrase, sentence, or text, and the notion of "pragmatics" refers to dealing with language in use and the contexts in which it is used. For example, Putnam (1975, pg. 143) cites that he cannot personally differentiate an elm from a beech tree, and further claims that his "concept" of a beech is the same as his "concept" of an elm. However, he goes on to say that his word "elm" applies only to elms and his word "beech" applies only to beeches. If he points at an elm and asserts "That's a beech", he speaks incorrectly. Does he therefore not know what elm and beech mean? Putnam (1975, pg. 144) uses that example to motivate "that there is a division of linguistic labor" and points to the fact that there are experts who can tell these trees apart when the need arises, and that users of the words know (roughly) how to find these experts. Thus Putnam's effective literacy is wider than his "private" literacy; it is scaffolded by the existence in his environment of the experts.

We can see that the moment that one considers the literacy of several individuals engaged in some activity, the extent to which they share knowledge about a web of concepts is a dimension of the concept of literacy. Literacy, is then a kind of social knowledge. One way to measure this shared literacy is in terms of a "concordance correlation coefficient," which is a measure from the biostatistics literature to measure the extent to which two distributions are the same as each other. This statistical concept provides an immediate formal characterization of what is meant by shared literacy in the context of studying the literacy of groups and households.

The second dimension of literacy associated with thinking about the literacy of groups and households is the concept of domain specific literacy. This refers to the idea that one might be literate, however measured, in one domain of activity and understanding, and yet be illiterate in another domain of activity and understanding. Much of the literature on literacy has considered different domains in isolation (e.g. health literacy, risk literacy, computer literacy, financial literacy, etc.). The point is that when one considers groups or households, one needs to consider the complementarity or substitutability of domain specific literacy at the level of individual, group or household. In other words, the literacy of the household might depend on the distribution of domain specific literacies within the household. If there is a household that is composed of two individuals that are financially literate, but health illiterate, then their effective literacy in one domain might be expected to be high, but their effective literacy in the other domain might be expected to be low. Now consider the possibility that one member of the household is literate in one domain and illiterate in the other domain, and the other member of the household is the perfect complement: illiterate in the domain where the partner is literate, and literate in the domain where the partner is illiterate. Again, one would expect that the potential literacy of the household is high in both domains, but it is far from obvious that the effective literacy of the household will be high in both domains. The difference between potential and effective literacy has to do with the extent to which the members of the household share their knowledge in these two domains, which in turn has to do with the extent to which the members of the household make joint decisions. In the "economics of the household" literature this is referred to as the bargaining problem within the household; an extensive survey of this literature can be found in *Economics of the Family* by Browning, Chiappori, and Weiss (2014). It is not always assumed that members of the household agree on every joint decision, so in the example considered here it could be that the members of the household do not acknowledge their distribution of comparative advantage. If the members of the household do not exhibit this "meta-awareness," either because they do not know one another's varying literacies or because they are acting strategically with respect to this knowledge, then it is possible that the effective literacy that is produced by the household in a joint decision could fall well short of the potential literacy of that household.

A particular channel for failure of shared knowledge and consequent coordination failure of the literacy of groups and households is the possibility that this bargaining problem fails for reasons related to stereotyped expectations, or bias based on some characteristic such as gender. The result could easily be a situation in which the effective literacy of the household is significantly lower than the potential literacy of the household. Of course, this failure of efficiency within the household can lead to differences between potential and effective literacy even within one domain. But it might be exacerbated if the members of the household have significantly different levels of literacy across the domains that households make joint decisions about.

Effective Literacy

Building on agencies of literacy, the idea and measure of *effective literacy* was introduced by Basu and Foster (1998) as a new approach to evaluate the aggregate literacy level in a country or region. They note that literacy is typically measured by taking the number of adults who are literate[2] as a percentage of the total number of adults in a country. However, they argue that *effective literacy* should account for the intrahousehold externality arising from the presence of a literate member. As long as a household has at least one literate member, then it is assumed to raise the *effective literacy* of the household. In these households an illiterate member of a household has immediate access to some portion of the functioning typically associated with literacy if the other household member is more literate. The assumption is that the production of literacy within a household is a simple matter: it is equal to the highest individual literacy over the members of the household. Thus, literate members generate a positive externality, or local public good, for illiterate members of a household.

Basu and Foster (1998) distinguish between two types of illiterate persons when assessing the effective literacy rate, as shown in Figure 1.1: 1) a *proximate illiterate*, an illiterate person who lives in a household with at least one literate member; and 2) an *isolated illiterate*, an illiterate person whose household has no literate member. Although this characterization measures literacy in a crude binary fashion, the logic is general. They show that the effective literacy measure can fluctuate widely based on the actual dispersion of literate individuals within a society. Indeed, if all literate individuals lived only with other literate individuals then the effective literacy measure collapses and is equal to the traditional literacy level. However, if literate individuals are dispersed in a society and households are comprised of both literate and illiterate individuals then the *effective literacy* measure will be higher than the traditional literacy measure.

[2] This study is about literal literacy, as determined by some measure of reading ability. The measure of literacy is not fundamental to the notion of effective literacy.

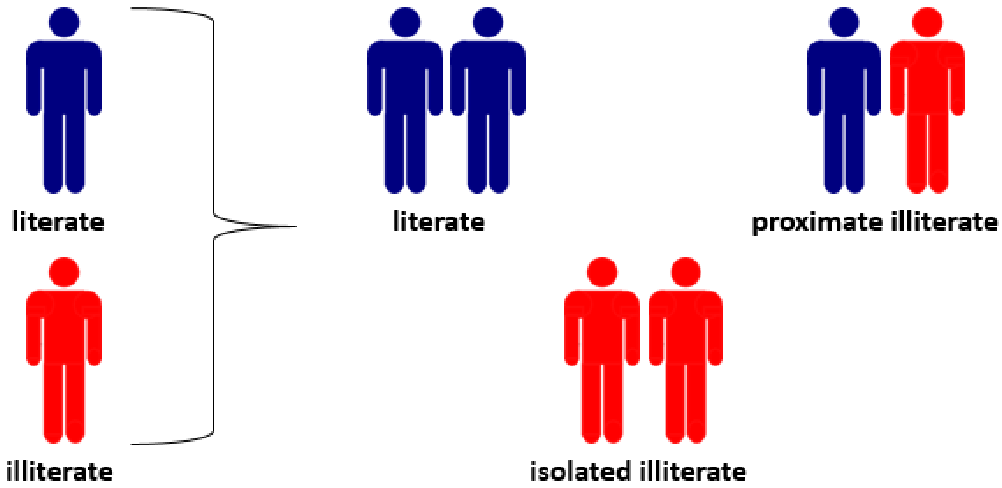


Figure 1.1 - Effective Literacy Composition of a Household with Literate and Illiterate Members

The Effects of Scaffolds on Literacy

Scaffolding has been defined generally by Clark (1997, pg. 45) as "exploitation of external structure" by an agent. The idea is that the mind of an agent can be extended by using external structure: indeed, "the extended mind" thesis proposed by Chalmers and Clark (1998) and central in Clark (2011) argues that the mind isn't reducible to or "inside" the brain. Your mind, according to them, is the integrated set of all your gear for forming, testing, and adapting beliefs, some of which is biologically inherited but many that is engineered. What is meant by external structure is not just limited to physical tools either, although they play a role (e.g., computers, paper and pen). It includes language and concept formation, as stressed by Vygotskij (1962) and Bruner (1968). Vygotskij (1962) focused on the manner in which concept formation and language helped experts communicate with and teach novices, and Bruner (1968) focused on the acquisition of oral language by infants. Each viewed scaffolding as aiding an agent in learning things. As Vygotskij (1987, pg. 210) put it, "what the child is able to do in collaboration today he will be able to do independently tomorrow." However, scaffolding can include the use of external structure that is ongoing, as in the use of language itself in social interaction.

The first clear statement of what is now thought of as scaffolding comes from the critiques by Dreyfus (1965, 1972) of the so-called "classical" approach to Artificial Intelligence (AI).[3] The attack on AI that Dreyfus launched was aimed the idea that we could design AI systems that contained all of the information they needed, and just had to "look up" this information when making a decision. Intelligent systems, Dreyfus argued, must be embedded in, and learn on the basis of dynamic interaction with, external environments.

Bruner (1968) picked up the early ideas of Dreyfus (1965) and applied them to developmental learning, as noted above. This extension was important for moving the notion of scaffolding to include social interactions, although not requiring that the external context be social. This extension is important for another reason: it allows us to see how scaffolds need not always be Pareto-improvements, even if they are costless to implement. For example, consider the use of the word "boat." What if the definition of a boat required that it be made of wood, drawing on historical precedent when the only boats were made of wood. Then when someone comes along to suggest making a boat of steel, one might expect a chorus of complaints that "that is not a boat!" Even more seriously, this concept of a boat, as scaffold, might inhibit one drawing inferences from the history of wooden boats to help inform the design of steel boats.

"Embodied cognition" means two different things, one of which is related to scaffolding. The unrelated item is a generalization of the idea of "muscle memory." The related item is: knowledge that we build into artifacts, so we can then "forget" it. For example: users once knew how to program their word processors, but not anymore because the knowledge is "embodied" in the interface. So that sense of embodied cognition is a form of scaffolding.

We should also be aware that even controlled experiments entail the use of some scaffolds, because of the use of language to convey questions and possible responses. A particularly striking example in time discounting experiments is when one tells the subject the interest rates implied by their

[3] This refers to AI based on pre-loaded axioms and factual knowledge, as opposed to more contemporary approaches based on deep learning of statistical patterns. See Graubard (1988).

choices between "smaller, sooner" amounts of money and "larger, later" amounts of money (e.g., Andersen, Harrison, Lau and Rutström (2014)). Individuals have varying field experience with what interest rates are, and hence treat that information as a scaffold to varying degrees. Indeed, the history of behavioral economics, particularly when focused on framing anomalies, is largely about manipulating, in a controlled manner, the nature of the scaffold provided to subjects. An open issue, which is a major theme of the comparison of lab and field experiments, is whether behavior in the context of artefactually provided scaffolds is the same as behavior in the context of natural scaffolds, including scaffolds that are endogenously sought out by agents (e.g., checking Google or Wikipedia). In a related vein, there is some evidence that humans process probabilities better when presented as "natural frequencies" rather than as probability statements, reflecting the intuition that humans have used frequencies more generally over time as a scaffold than they have used probabilities (e.g., Cosmides and Tooby (1996) and Gigerenzer and Hoffrage (1995)).

Scaffolding can both promote and detract from efficiency, depending on the context.

Wisdom of the Crowd

The "wisdom of the crowd" is a hypothesis that a diverse collection of independently deciding individuals is likely to make certain types of decisions and predictions better than a given individual or single expert. The wisdom of the crowd seeks to compare the quality of the collective opinion of a group of people with that of a single person. An extensive survey of this literature and its applications can be found in *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societiesces* by Scott Page (2007). The wisdom of the crowd results when social scaffolding produces a net gain in effective literacy.

The logic behind the wisdom of the crowd hypothesis can be traced back to a theorem that was first conveyed by the Marquis de Condorcet (1785) in his work *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (Essay on the Application of Analysis to the Probability of Majority Decisions). The Condorcet Jury Theorem, as it is known, in its most reduced form assumes that a group (or jury) is required to reach a decision by majority vote. Summarizing from Weisstein

(2018), each individual in the group has one vote that is counted and can either be *correct* with probability $0 \leq p \leq 1$ or *incorrect* with probability $(1-p)$. The theorem goes on to ask how many voters should be included in the group to maximize the chances of a *correct* majority decision. Here it can be shown that the result depends on whether the value of p is greater than or less than 0.5. The two outcomes can be:

- $p > 0.5$, so that each voter is more likely to vote correctly. In this case we can show that adding more voters to the group increases the probability that the majority vote is correct. In the limit, the probability that the majority votes correctly will approach 1 as the number of voters increases.
- $p < 0.5$, so that each voter is more likely to vote incorrectly. In this case, adding more voters to the group makes things worse, and the optimal jury consists of a single voter.

We can show mathematically, for the case of 3 voters, that the probability of a correct majority vote is higher than the probability of a correct decision by any given individual in the group for any $p > 0.5$. Consider the case of $p = 0.6$ and n=3. Here we see that, for each voting outcome that results in a correct decision by the group, (0.6*0.6*0.6) + (0.6*0.6*0.4) + (0.6*0.4*0.6) + (0.6*0.6*0.4) = 0.648 is the probability of a correct majority decision, which is indeed greater than the individual $p = 0.6$. Additionally, using *Mathematica* and code from Saito (2018) we can show this concept graphically for the case of $p = 0.6$ as n $\rightarrow$ 100 in Figure 1.2. Note in the bottom panel of the figure, the probability that the correct choice wins approaches 1 as the number of voters increases.

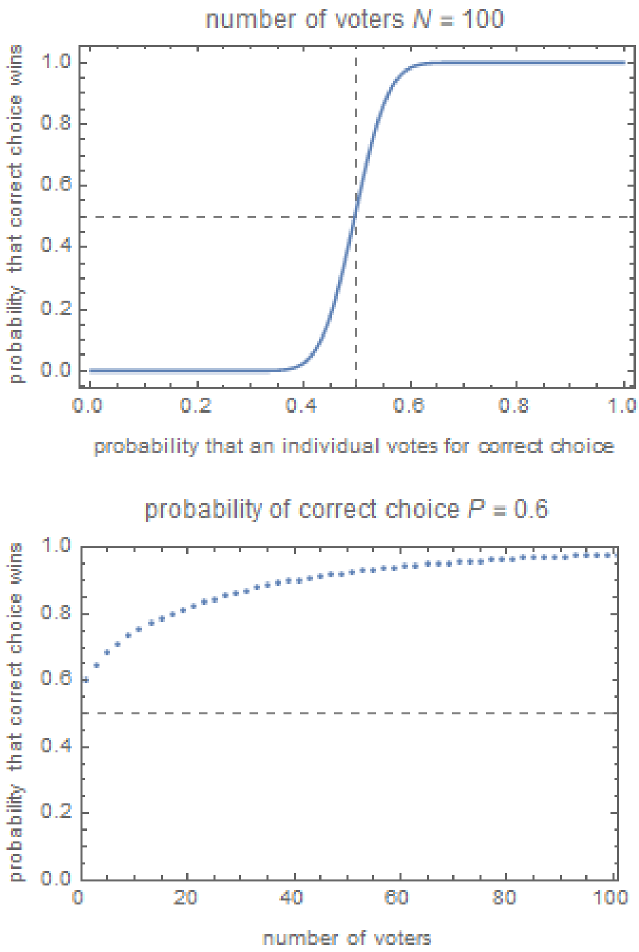


Figure 1.2 - Illustration of Condorcet's Jury Theorem

An often cited early example of the wisdom of crowds is an application by social scientist and statistician Sir Francis Galton to find a point estimate from a continuous distribution in his article "*Vox Populi"* in *Nature* (1907). Here he was interested in the results of a weight-judging competition that was carried out at the 1906 annual show of the *West of England Fat Stock and Poultry Exhibition* held in Plymouth, England. The event sold tickets for a sixpenny fee to competitors who were shown a preselected fat ox. Individuals would then provide their best estimates of what the ox would weigh after it had been slaughtered and dressed; they also included their personal information to be contacted later. Prizes were awarded to those with the most accurate predictions. There were 787 valid tickets sold during that show, which Francis Galton was granted access to, and he analyzed the written responses. He reported that the median guess was 1,207 pounds and that it was accurate within 0.8% of the true weight of 1,198 pounds of the slaughtered and dressed ox. He contrasted that with the probable error of a single observation picked at random, which would have been off by 3.1% of the true weight, or a 37-

pound difference. He noted that "[the median result] is…more creditable to the trustworthiness of a democratic judgment than might have been expected," meaning the point estimate of the group was better than that of a randomly selected individual.

Groupthink

Although the wisdom of the crowd approach involves aggregating subjective beliefs across numerous individuals that can often hold different and varied beliefs, the concept of groupthink, pioneered by Janis (1971, p. 43), posits "a mode of thinking that persons engage in when *concurrence-seeking* becomes so dominant in a cohesive ingroup that it tends to override realistic appraisal of alternative courses of actions." In other words, groupthink is a psychological phenomenon that occurs within a group in which the desire for harmony or conformity in the group results in irrational or dysfunctional decision-making outcome. Group members seek to reach a consensus decision with minimal conflict and without critical evaluation of alternative viewpoints by actively suppressing dissenting viewpoints, and by isolating themselves from outside influences. Groupthink arises when social scaffolding produces a net loss in effective literacy

Peck (1996, p. 32) summarizes Janis (1972, 1982) noting that common characteristics of groupthink "include: an illusion of invulnerability; an unquestioned belief in the group's inherent morality; collective efforts to discount warnings; stereotyped views of the enemy as evil; self-censorship of deviations from group beliefs; a shared illusion of unanimity; suppression of dissent; and the emergence of self-appointed mind-guards who screen the group from dissidents."

Sunstein and Hastie (2015) examine several ways in which groups fall victim to groupthink and offer strategies to improve decisions made in group settings. They first review challenges they find contribute to poor group decisions. Four pertinent factors they cover are collective myopia, following the lead, closing ranks, and ignoring outliers. Collective myopia describes a lack of foresight on the part of the group and can occur for many reasons; for example, a group's tendency to be overoptimistic about future forecasts, a group's bias for the most familiar outcome, or even the unrecognized impacts of common mental shortcuts. Following the lead refers to the influence of what is discussed first and who speaks first. Sunstein and Hastie (2015, p.63) refer to these as cascade effects, which can be informational and reputational. Closing ranks in a group occurs when members become more polarized and skewed in a particular direction based on prior discussions, thereby making it more difficult to offer a dissenting view. Lastly, ignoring outliers can be described as a group giving little attention to what may be known only by a few.

Measurements

Measurements of literacy can be sorted into two main classifications: qualitative and quantitative. The work presented in this thesis will focus on quantitative results. However, it is worth mentioning two qualitative measures that are found in the literature. One measure involves focus groups which comprise either a group or an individual being interviewed. The interview process in a focus group provides structure and a narrative in order to understand issues that are identified for a cross section of the population being interviewed. Two examples of focus groups recruited to study financial planning are Kennickell, Starr-McCluer, and Sundén (1996), using high net-worth individuals, and Turnham (2010), using low-income populations. The second qualitative measure to mention is the “teach back” method that is widely employed in the health literacy literature (see Denny and Grady 2007; Kripalani et al. 2008; Nigolian and Miller 2011; Seley and Weinger 2007). The technique behind the teach back method, in a healthcare setting, is for a healthcare provider to check the understanding or comprehension of the patient, or their care-giver, by asking them to demonstrate or repeat in their own words what they believe are the take away points of the session. This process should not be done in quizzing or shaming way, and offers the provider an opportunity to check if the patient is health literate in order to be able to adhere to the protocol of care prescribed. Both qualitative methods described above have the capability of generating further insights and potentially improve validity of outcomes and results. They are often viewed as complimentary to quantitative data.

Quantitative research is based on quantities elicited using a specified measurement process. There is no one standard measure that is used across studies and the types of measurements vary across disciplines. For our purposes we will talk about measurements that are found in literacy literature.

Literacy at its most fundamental refers to a person's ability to read and write. Here there is often a binary measure if an individual can read and write or not, respectively coded as 1 or 0. Other literacy measures follow the same classification scheme in the sense that if some individual answers a fill-in-the-blank or multiple-choice question correctly then they 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 test downstream behavior. We will explore the literature regarding quantitative measurements with respect to financial literacy in subsection 1.2.2.

These literacy measures elicit a *point estimate* rather than recovering the *full distribution* of a subject's beliefs about an answer to a question. Merkle and Weber (2011, p.264) correctly illustrate this concept, which is reproduced as Figure 1.3. The top panel of the figure represents a belief distribution about a skill or ability over different quantiles (e.g. deciles). Shown in that panel is a person that assigns positive weights to deciles 4 through 10 as believed to contain the true value of their skill or ability. They assign less weight to deciles 4,5, 9, and 10, and more weight to deciles 6, 7, and 8. The point to note here is that the subject believes that their skill or ability falls somewhere across those 7 deciles. The bottom panel of the figure shows what a person may report as their *point estimate*, here in decile 7, which is in fact distilled from their entire underlying belief distribution. Merkle and Weber (2011) use this design to test various causes of overconfidence. The research presented in this thesis uses an improved method to test financial literacy in participants where a true answer is known. The method also allows for a rich characterization of the belief distribution underlying an individual's elicited response, a test for any bias, and an evaluation of their confidence of response with respect to a known, true, answer.

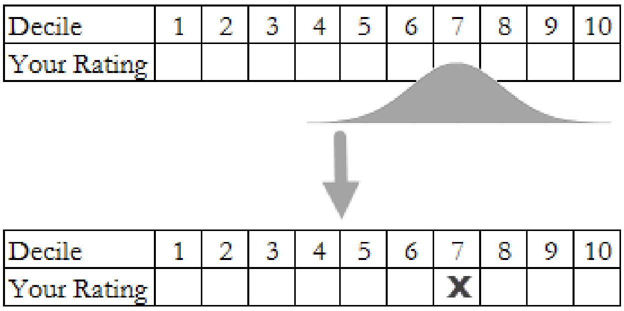


Figure 1.3 - Aggregation of beliefs. Full distribution (above). Point estimate (below).

Bias and Overconfidence

We can say that someone's subjective beliefs exhibit bias when the weighted average belief is different from the true outcome. Take the bottom panel of Figure 1.3 for instance, which shows an individual's point estimate belief in the 7^{th} decile. Now consider that the true answer is located in the 9^{th} decile: here we observe that the individual's belief was biased from the true answer by 2 deciles. This definition mimics the definition of bias in statistics, where the estimated expected value of the parameter being estimated differs from the true underlying quantitative parameter being estimated. Thus bias *per se* is a strict concept, and does not even "speak to" the statistical significance of any bias. Of course, the latter is what concerns economists, so we must always have a concern with the confidence with which a belief is held. A related concept in statistics is "consistency," which is the idea that bias disappears as samples get asymptotically large. Thus an estimator may be consistent, but still exhibit "small sample bias" if it is applied with a small sample. The concept of bias presumes, of course, that a true outcome exists and is defined. Our experimental design ensures this is the case and will be covered in Section 1.3.

Figure 1.4 illustrates this idea of bias. The two panels on the left exhibit zero bias, and the two panels on the right exhibit bias. In each case the true outcome on the horizontal axis is marked by a (green) diamond shape equal to 100 on the vertical axis.

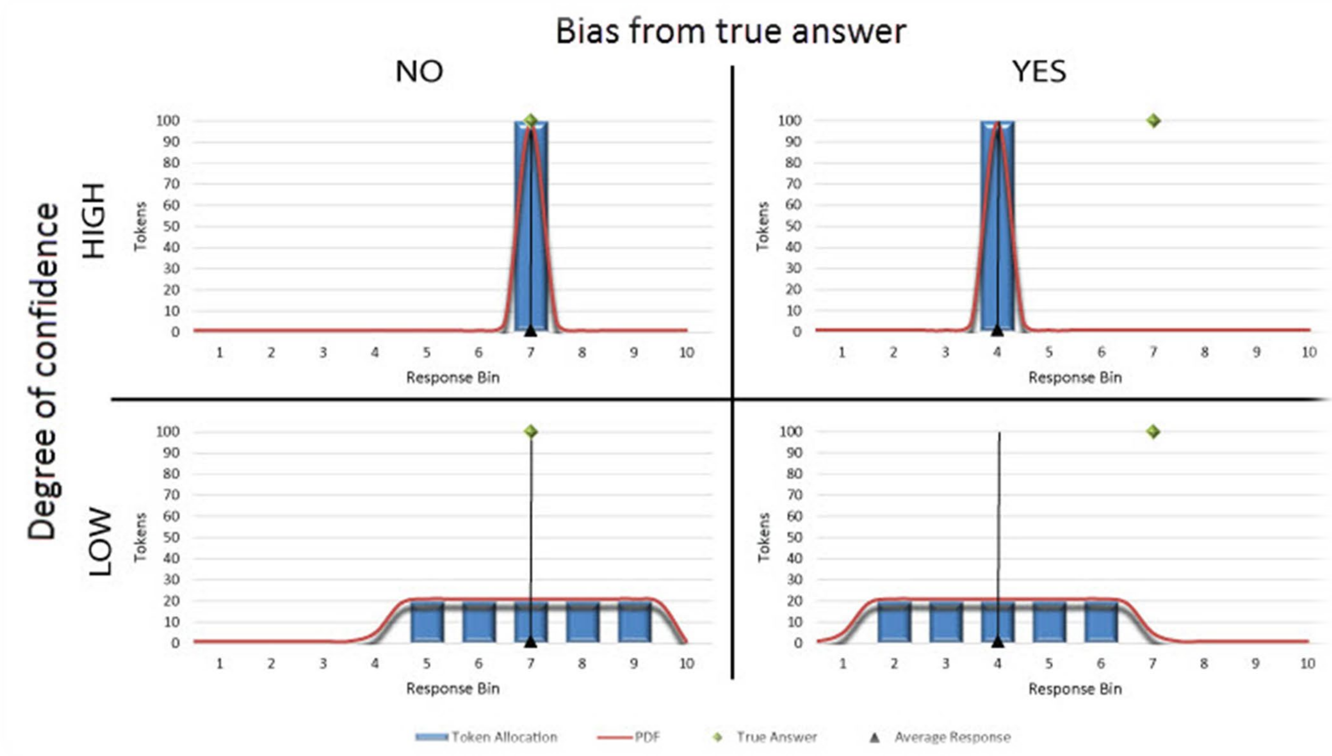


Figure 1.4 - Characterizing Responses Using Bias and Confidence

What does it mean when we say that someone displays overconfidence? The term overconfidence is meant to describe a bias whereby a person systematically overrates their subjective confidence in their judgements compared to the objective accuracy of those judgements. Overconfidence is one possible characteristic of a miscalibration of subjective probabilities. Summarizing Moore and Healy (2008), overconfidence has been defined in three distinct ways: (1) overestimation of one's actual performance; (2) overplacement, or the "better than average effect", of one's performance relative to others; and (3) overprecision by expressing unwarranted certainty in the accuracy of one's beliefs.

Moore and Healy (2008, p.508), Merkle and Weber (2011, p.264) and Benoît and Dubra (2011, p.1605) explain why elicited subjective belief distributions are needed to evaluate hypotheses about overconfidence in the third sense defined above. Moore and Healy (2008) and Merkle and Weber (2011) explicitly recognize the need to elicit subjective belief distributions using proper scoring rules. They both apply variants of the Quadratic Scoring Rule (QSR), but their procedures are limited in several respects by comparison to the ones employed here. First, their interface is not as intuitive as the one we use.

Second, and most critically, they do not correct for the fact that preferences concerning risk will lead subjects to report beliefs that reflect those risk attitudes (Harrison and Ulm 2016).

Figure 1.4 illustrates this idea of confidence. The two panels on the top exhibit high confidence, and the two panels on the bottom exhibit low confidence. Note that we carefully do not label these as overconfidence and underconfidence, respectively. To do that one has to ascertain the appropriate level of confidence, and that depends on the priors that an individual has as well as the precision of the data that the individual has seen. For instance, just as we expect that an econometric estimator should have a large standard error when the sample is small, an individual should often have a large variance in their beliefs when they do not know much about an event contingency. In Figure 1.4 we implicitly associate confidence with variance, but that is just for illustration.

1.2.2 What is Financial Literacy?

Overview

What is it then to say that someone is financially literate? Table 1.1 illustrates a sample of the varied definitions that occur in the literature. The first 8 rows of Table 1.1 are taken from Huston's (2010) meta-analysis covering 71 published studies that included financial literacy/financial knowledge measures over the years 1996 to 2008, while the last 4 rows are further supplements from my review of the literature.

Definitions of Financial Literacy Found in the Literature	
1	Financial literacy is the ability to make informed judgments and to take effective decisions regarding the use and management of money (Noctor, Stoney, and Stradling (1992), definition used by Beal and Delpachitra (2003) and ANZ (2008)).
2	Personal financial literacy is the ability to read, analyze, manage and communicate about the personal financial conditions that affect material well-being. It includes the ability to discern financial choices, discuss money and competently to life events that affect everyday financial decisions, including events in the general financial issues without (or despite) discomfort, plan for the future and respond economy (Vitt et al. (2000); also cited by Cude et al. (2006)).
3	Financial literacy is a basic knowledge that people need in order to survive in a modern society (Kim 2001).
4	Financial literacy refers to a person's ability to understand and make use of financial concepts (Servon and Kaestner 2008).
5	Financial literacy is the ability to use knowledge and skills to manage financial resources effectively for lifetime financial security (Jump$tart Coalition 2007).
6	Financial literacy is the ability to use knowledge and skills to manage financial resources effectively for a lifetime of financial well-being (U.S. Financial Literacy and Education Commission 2006).
7	Financial knowledge is defined as understanding key financial terms and concepts needed to function daily in American society (Bowen 2002).
8	Consumer literacy, defined as self-assessed financial knowledge or objective knowledge (Courchane and Zorn 2005).
9	Financial literacy determines how well people make and execute financial decisions, including saving, investing, borrowing from one's retirement account, and planning for retirement (Lusardi, Mitchell, and Curto 2010).
10	Financial literacy as having the knowledge, skills and confidence to make responsible financial decisions (Task Force on Financial Literacy, Office of the Canadian Minister of Finance 2009).
11	By the most basic definition, financial literacy relates to a person's competency for managing money (Remund 2010).
12	Characterize [financial] literacy in terms of the subjective beliefs that someone has over possible responses to some question (Di Girolamo et al. 2015).

Table 1.1 - Definitions of Financial Literacy Found in the Literature

After reading through Table 1.1, one can see that the literature has not yet arrived at a unified voice as to how to define financial literacy. Partly because of this lack of unification and precision, the literature often treats the terms *financial literacy* and *financial capability* as synonymous, when they are not.

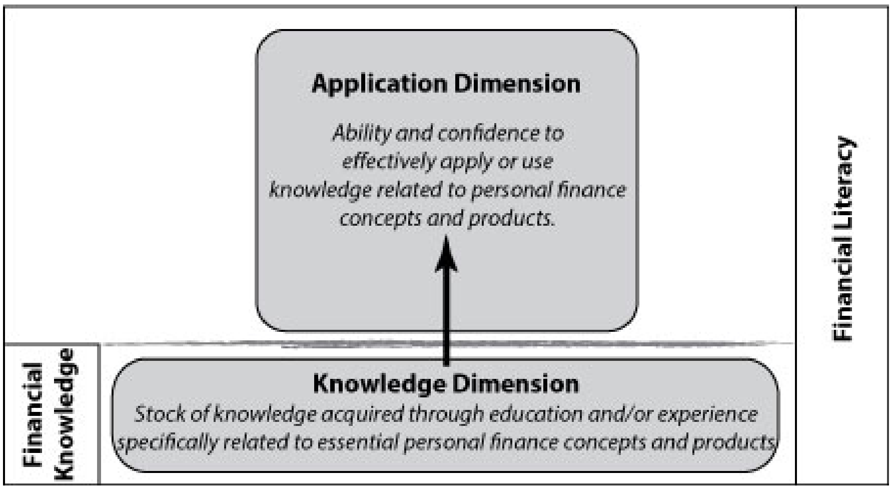
Financial capability can mean more than just understanding concepts, and refers to the capability of the individual to make financial decisions coherently, and to manage their financial well-being (Holzmann 2010). One of the first national financial capability surveys to be administered was by the United Kingdom's Financial Saving Authority with their "Financial Capability in the UK: Establishing a Baseline Survey" in 2005. Instead of only looking at financial knowledge in a narrow sense, the UK survey focused additionally on 5 content domain areas: making ends meet, keeping track of your finances, planning ahead, choosing financial products, and staying informed about financial matters (Atkinson et al. 2006).

While a consensus definition of financial literacy and financial capability is unlikely due to the evolving nature of the literature, Remund (2010) makes a contribution by categorizing financial concepts into five areas. They are: (1) knowledge of financial concepts, (2) ability to communicate about financial concepts, (3) aptitude in managing personal finances, (4) skill in making appropriate financial decisions, and (5) confidence in planning effectively for future financial needs. When these 5 categories are assessed together they account for financial capability.

Figure 1.5, from Huston (2010), offers a graphical representation of her conceptual framework and definition of financial literacy. Here we see that category (1) from Remund (2010) maps cleanly into Huston's (2010) "Knowledge Dimension", while categories (2) through (5) map broadly across what she refers to as the "Application Dimension." She incorrectly defines the two dimensions taken together as financial literacy. I suggest that the area labeled "Financial Knowledge" is actually financial literacy, and the "Application Dimension" should be referred to as financial capability.

The difference between literacy and capability is that literacy is rote knowledge and testable in a rigorous setting, whereas capability is the application of that knowledge in some field context. That is the key distinction between literacy and capability, but it is often blurred. Our research adopts the definition that financial literacy is rote financial knowledge that is testable and can be evaluated against a known, true answer.

Figure 1.5 - Concept of Financial Literacy



Financial Literacy Measurement

Moving from a conceptual definition of financial literacy to its measurement, the literature falls into two sets. The first category is *manipulated financial literacy*, studies involving experimental and quasi-experimental methods to analyze the effects of financial education interventions. The second area of the literature is *measured financial literacy*, studies including correlation analysis and econometric studies that measure financial literacy by the percent of correct answers on tests of financial knowledge and/or predicted downstream financial behaviors (Fernandes, Lynch, and Netemeyer 2014).

Manipulated financial literacy is the smaller of the two literatures. One example is Clark, Morrill, and Allen (2011), which examined data on over 1,000 retirement-eligible workers, focusing on their hypothetical preferences for either a lump sum or an annuity stream of retirement income before and after their participation in a retirement planning seminar. The participants came into the study belonging to either a group that owned a defined contribution (DC) plan (401(k), 403(b), or 457) or a defined benefit (DB) plan (employer pension). The authors note that the unique characteristics of the default options of the different plans were akin to an annuity stream for the DB and a lump sum for the DC. However, there are provisions in both plans according to which a worker, upon retirement, can opt for the other payment method; e.g. buying an annuity stream from the lump sum of their DC.

Before the seminar the individuals were asked what type of payment option they would hypothetically prefer: lump sum or annuity. The same individuals then participated in an employer-provided pre-retirement seminar, which ranged from a half-day to two days in duration. After the seminar the individuals were asked again what their hypothetical preference was. Results were shown for both groups (DB and DC); however, I will only focus on those with DB plans. Prior to the workshop, 70.5% said they would stay with the annuity stream and 29.5% preferred the lump sum. After the seminar, 72.4% preferred the annuity stream and 27.6% the lump sum. This shows that the preferences of *some* individuals reversed, presumably from the information presented in the seminar. However, this could just be statistical noise. Strictly speaking, they all could have reversed, but in opposite directions.

Measured financial literacy is the more pervasive of the two types of literature, as it has gained the attention and funding of governments, non-profit organizations, non-governmental organizations, and special interest groups on a large scale. A well-known example of measured financial literacy is the Personal Financial Survey, which is administered in the United States by the *Jump$tart Coalition for Personal Financial Literacy*. *Jump$tart* is often credited with the first large scale financial literacy survey of American youth. Their nationwide survey is conducted with 12th grade students to determine the financial ability of young people. The survey is presented in a multiple choice format and covers topics surrounding income, money management, saving and investing, and spending and credit. A student is deemed literate for a specific question if they answer the question correctly. Mandell (2008), the study author, claims that students are financially literate if they score 75% or more; however, how he arrives at that specific value is not discussed.

The first *Jump$tart* survey was conducted in 1997-1998 and was then repeated in years 2000, 2002, 2004, 2006 and 2008. In 2008 it was expanded to include a survey of college students as well. The average grades for all those sampled are listed in Table 1.2 as follows:

Year	Number of High School students polled	Average Grade (defined as amount of questions answered correctly out of 31 total questions)
1997	1,532	57.3%
2000	723	51.9%
2002	4,024	50.2%
2004	4,074	52.3%
2006	5,775	52.4%
2008	6,856	48.3%
2008 (full-time college)	1,030	62.2%

Table 1.2 - Financial Literacy of 12th Grade Students - Results from Jump$tart

This is only one example of measured financial literacy, but most follow the same classification scheme in the sense that if an individual answers a multiple-choice question correctly then they are deemed financially literate with respect to the topic of that question. Indices of how literate the individual is are then often built and these indices are tested on downstream behaviors such as whether or not an individual saves for retirement, enters into a stock market, takes out payday loans, or any number of other behaviors.

Despite the widespread inclusion of financial literacy questions in various surveys and research, there have been only few attempts at developing a rigorous methodology for its measurement and to standardize the approach.

1.2.3 Risk Perception

In the same way that it is necessary to define financial literacy and how it is measured, it is also necessary to define risk and how it can be measured. Risk can be categorized in two classes: *objective risk* and *subjective risk*. Objective risk is associated with random physical systems such as roulette wheels, rolling dice, or flipping coins. An example of objective risk is flipping a fair two-sided coin 100 times and computing that it lands on heads with probability 0.5. Subjective risk, however, refers to an individual's personal judgment or perception about how likely an event is to occur.[4] Subjective risk will differ from person to person due to many factors, such as past experience, risk attitudes, expert knowledge, or personal bias. Continuing from the example above, since subjective risk varies by individual, one person

[4] Even though we *define* subjective risk at the level of the individual, one can also enquire about the subjective risks of groups of agents, such as households. Indeed, the "effective literacy" concept, discussed in Section 1.2.1 and also in later chapters, does precisely that.

could state that their subjective belief that flipping a fair two-sided coin and observing it landing on heads on their next flip is a 25% chance, whereas another person may state a belief in an 80% chance.

1.2.4 Subjective Belief Elicitation

The von Neumann-Morgenstern axioms of completeness, transitivity, independence, and continuity, which define a rational decision maker under Expected Utility Theory (EUT), were originally formulated to be used with objective probabilities. However, Savage (1954) adopted the theory to characterize the fact that different people make different decisions because they have may have different utility functions and/or different beliefs about the probabilities of different outcomes for event contingencies, or different beliefs about some verifiable answer. Savage (1954) developed the standard theory of Subjective Expected Utility (SEU).

SEU builds on EUT by adding additional structure which states that a rational decision maker can believe that an uncertain event has possible outcomes with differing probabilities and each of these possible outcomes has a utility associated with it. Choices can then be characterized by a function in which an individual believes that there is a subjective probability over each outcome and the subjective expected utility is simply the subjectively evaluated expected value of the utility. In effect, SEU assumes that people behave as if they take the average of their subjective belief distribution when placing bets on binary events. Which decision an individual prefers depends on which subjective expected utility is higher.

Savage (1971) showed that under the conditions of SEU, and using a proper scoring rule, it is possible to elicit a person's subjective probabilities and other expectations over binary events. Matheson and Winkler (1976) develop families of scoring rules for the elicitation of probability distributions over continuous events. Building on these studies, Andersen et al. (2014) operationalize a method to recover an individual's subjective probabilities using the incentive-compatible Quadratic Scoring Rule (QSR) over binary events, adjusting for distortions generated by risk aversion. Harrison et al. (2017) operationalize the QSR over continuous events, and characterize the QSR theoretically and empirically.

1.2.5 Behavioral Welfare Economics Approach to Financial Literacy

Bernheim and Rangel (2009) and Bernheim (2009) present an approach to behavioral welfare economics that recognizes the methodological challenge of evaluating welfare when one does not accept that one can rely on (naive) revealed preference. Their approach is a generalized method that defines welfare directly in terms of observed choice. Ambuehl et al. (2014, 2017, and 2018) utilize the Bernheim and Rangel (2009) and Bernheim (2009) framework to develop two frames with which to ask a question bearing on financial choices, where two conditions are met, and are couched here in terms of a financial literacy application:

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

Note that both conditions rely on *a priori* judgments. There is nothing wrong with this, but of course the "proof is in the pudding" when one gets to specific applications, and different readers might have different priors on the validity of these two conditions.[5] The application of these ideas in Ambuehl, Bernheim, and Lusardi (2014)(2017) and Ambuehl, Bernheim, Ersoy and Harris (2018) provide just such an instance, focused squarely on financial literacy.

[5] This is the approach adopted in Ambuehl, Bernheim, and Lusardi (2014), to view one of the frames as revealing true, latent valuations. In Ambuehl, Bernheim, and Lusardi (2017) this position was qualified, allowing that there might be some normative metric that does not lead one to accept that either frame represents the true, latent valuation. The example provided is when subjects exhibit Quasi-Hyperbolic discounting in response to both questions, with Exponential discounting *a priori* deemed to be normatively attractive and Quasi-Hyperbolic discounting deemed *a priori* to be normatively unattractive. In this case, they claim, both responses might be "contaminated" by the "passion for the present" one expects from Quasi-Hyperbolic responses. They then present a formal mathematical result that essentially says that if the responses to statements A and B are equally contaminated, then as one takes the *limit of the difference between the responses as that difference goes to zero*, then a first-order approximation to a valid welfare measure can be obtained. But that says nothing about whether the difference between the responses that are non-zero, or not close to zero, have any valid interpretation, unless one wants to invoke stringent path-independence assumptions from welfare economics (see Broadway and Bruce (1984, pg. 199) or Harrison, Rutherford, and Wooton (1993)). The bulk of responses of interest are decidedly non-zero, and not close to zero, as illustrated in Ambuehl et al. (2018, Figure 1, pg. 16). The general methodological issue of concern here is discussed directly by Harrison and Ng (2016, pg. 115ff) and Harrison and Ross (2017). Bernheim (2016) provides a general statement of his approach to behavioral welfare economics, including discussion of the application to financial literacy.

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

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

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

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

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

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

Another, subtler interpretation issue concerns the information about a 3% interest rate. A subject might reasonably presume that this is taken to be the market (borrowing and lending) interest rate for this question. Then we know from the Fisher Separation Theorem that we cannot recover estimates of the subjective discount rate due to censoring: see Coller and Williams (1999), Harrison, Lau, and Williams (2002), Harrison, Harstad, and Rutström (2004) for extended discussions. All that we would recover is their knowledge of the interest rate, which is again included in statement B, hence we would again expect a spike of responses at $22.

Extending this point, the mere mention of interest rates provides a scaffold that might affect responses differently for statement B compared to statement A. In effect, statement B offers a scaffold, mentioned earlier in Section 1.2.1, that could be expected to change the response compared to statement A, where there is no such explicit scaffold mentioned. Thus what is claimed to be the welfare effect of literacy might just be the welfare effect of having access to a scaffold, and that is ambiguous as a theoretical matter.

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

The general concern here is that to apply the method of Bernheim and Rangel (2009) one must find frames that convince readers that they meet the two conditions noted earlier, and this is not likely to be an easy task across domains.

1.3 Methodology

The research presented here follows in the footsteps of Merkle and Weber (2011), Andersen et al. (2012), Harrison and Phillips (2014), Di Girolamo et al. (2015), Harrison and Ulm (2016), and Harrison et al. (2017) utilizing *Subjective Belief Elicitation* to characterize and measure the financial literacy of an individual by eliciting their entire subjective belief distribution. This method ascertains how precise an individual's knowledge is in response to some question, controlling for their attitude to risk. This thesis uses these methods to measure financial literacy in a rigorous manner that compares participants' responses to a known, true answer. This approach allows for richer and more rigorous characterization of financial literacy compared to multiple-choice questions or "fill-in-the-blank" responses.

Harrison and Phillips (2014) and Harrison et al. (2017) document in detail the subjective belief elicitation procedure that will be followed to test financial literacy in the lab and field. The approach procedurally as follows: a study participant in our research will be exposed to a task in which they will be paid according to how accurate their beliefs are about certain financial literacy questions. They will be asked to place bets based on their beliefs about the answers to each question. Figure 1.6 is an example of the display of a response screen.

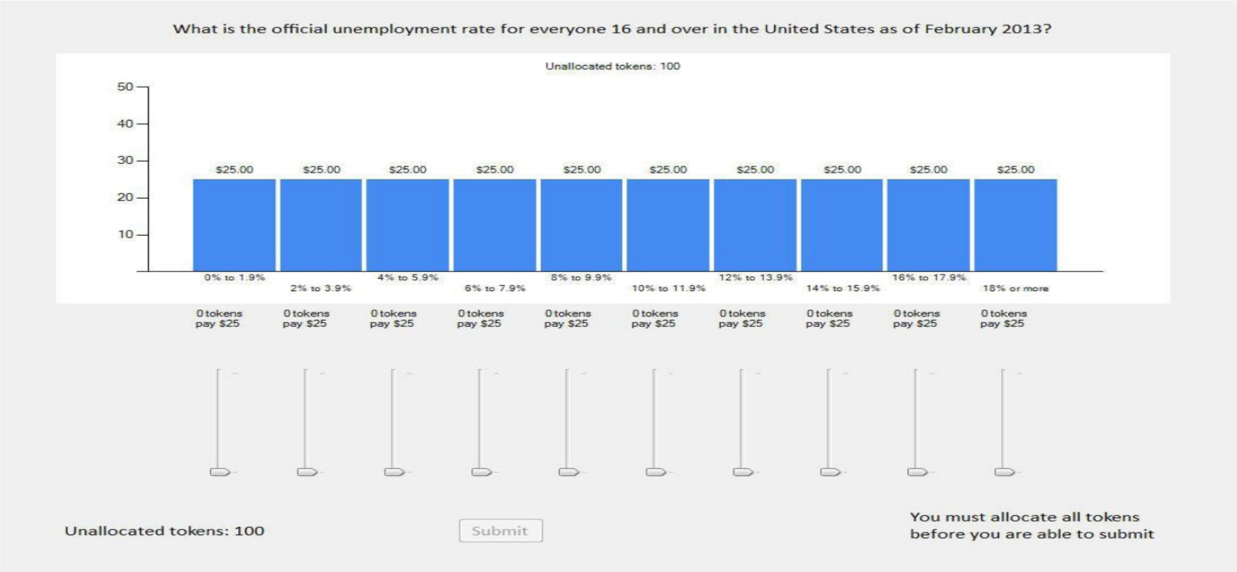


Figure 1.6 - Belief Elicitation Interface – Tokens Evenly Distributed

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 allows them to allocate tokens that reflect their belief about the answer to the question. 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 will change as shown in Figure 1.7.

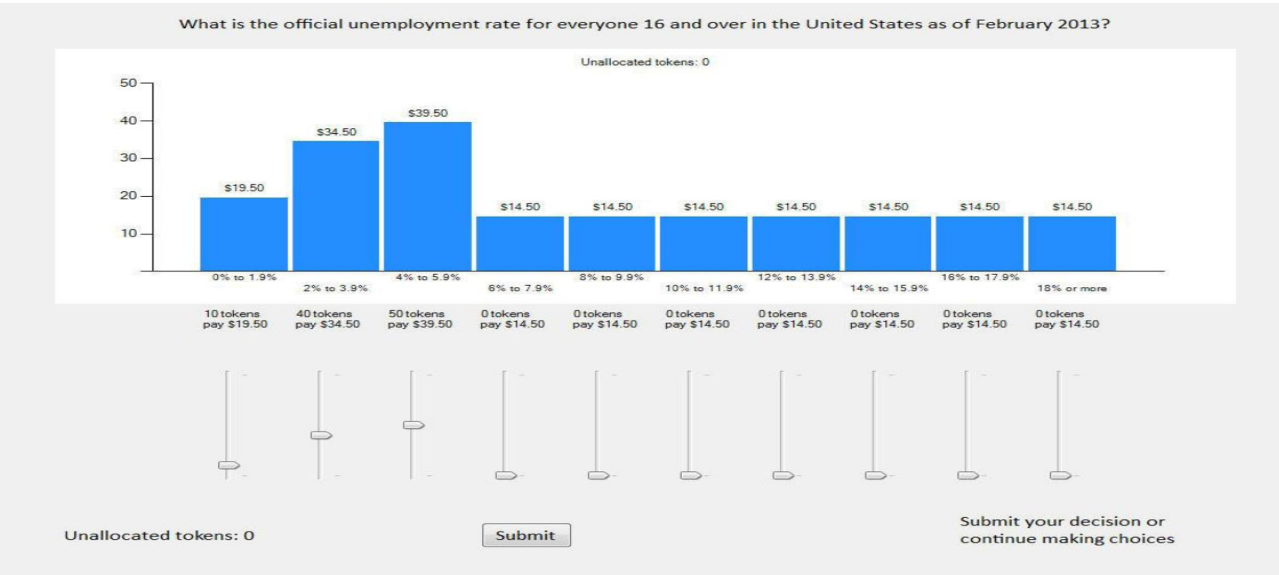


Figure 1.7 - Tokens Distributed According to Individual's Belief

Participants' earnings are based on the payoffs, which is generated by the QSR applied to their token allocation and displayed in real time as they re-allocate all 100 tokens. A participant is paid the displayed amount above an interval if and only if that interval contains the true answer.

It is therefore up to the participants to balance the strength of their personal beliefs with the possibility of them being wrong. Their 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 participant is also incentivized to give statements about their subjective belief distribution. Their choices might also depend on their willingness to take risks. Harrison et al. (2017) prove several theorems characterizing the optimal responses of an SEU agent, including results that suggest that risk averse SEU agents will effectively reveal their true beliefs for plausible levels of risk aversion. Procedures for identifying non-EUT, and hence presumptively non-SEU, individuals will be employed to check on the empirical robustness of this assumption for each individual.[6]

1.4 Applications – Lab & Field

This section introduces proposed applications using subjective belief elicitation in both laboratory and field settings. Here it is important to note the complementarity of experiments conducted in the laboratory and field. The laboratory provides greater control over the administration of an experiment, is faster to implement, and is usually less costly than going directly into the field. It is in the lab that we can test with more precision the experimental design and methodology, and the effect of word choice or framing on the results of an experiment. The lab is the easiest place to ensure internal validity of tests of hypotheses. Going through this process allows for an efficient feedback mechanism to inform and understand experimental results. It is after an experiment and results are understood in a lab setting that we can then venture out with greater confidence to administer the experiment in the field, where there is often less control and mistakes can be costlier in terms of both time and money. Of course, the field experiment has greater potential to establish external validity than the lab experiment.[7]

This dissertation uses the latest state-of-the-art advances in subjective belief elicitation to evaluate the financial literacy of individuals in several settings. Chapter 2 opens with an introduction to

[6] Harrison and Phillips (2014) employ a "binary lottery procedure" to offset, in theory, the effects of risk aversion on the incentive to report truthfully. This procedure is evaluated by Harrison et al. (2015).
[7] Internal validity reflects the extent to which a study minimizes systematic error relative to some theory or causal hypothesis. External validity is the extent to which the results of a study can be generalized to other situations and to other people.

the techniques and tools used throughout this work and evaluates the financial literacy of individuals in the lab. Chapter 3 goes further and introduces a new measure of *extended financial literacy* when individuals in the lab have accesses to some external cognitive "scaffold"; in this example, access to the internet is the external scaffold. Chapter 4 builds on the concepts in the previous chapter and introduces the idea and measure of *effective literacy,* as introduced by Basu and Foster (1998), and uses their approach to test the financial literacy of individuals when they are part of a group. Chapter 5 transitions the techniques developed in the lab to the field and evaluates the financial literacy of naturally occurring households in Denmark.

1.5 Summary

In the beginning literacy measures were a simple metric of whether an individual could read and write, which is often now referred to as "reading literacy." Here we could simply count up the total number of reading literate in a country and divide by the total population to compute the percentage of literate individuals in that country. However, this simple approach ignores the situation in which an illiterate person has access to a literate person that could read to them. This was the premise of Basu and Foster (1998), introducing the new measure of *effective literacy* that accounts for the potential positive externalities that could arise from access to a literate individual. The research in this thesis builds on effective literacy as a measure and introduces a new measure of literacy called *extended literacy*, which gives a decision-maker access to an external scaffold during the decision-making process.

The next distinction is that literacy does not equal capability, despite the fact that the management communication literature often confounds the two as being the same. This would be like saying that because someone can read a word they are capable of understanding it and using it correctly in a sentence, or being able to numerically calculate a return on an investment and being capable of investing their money reliably using that calculation. The distinction between literacy and capability is that literacy is rote knowledge and testable in a rigorous setting, whereas capability is the application of that knowledge in some field context.

The research presented in this thesis measures financial literacy at different levels of decision-making: individuals, individuals given access to scaffolds, and groups or households. The techniques involved in the measurement of literacy reflect state-of-the-art advances in subjective belief elicitation that allow for the recovery of a decision-maker's entire underlying subjective distribution. This technique generates a rich characterization of beliefs and allows us to talk about bias and confidence with respect to a known, true answer. We have the tools and use them to answer interesting questions.

Chapter 2 – Toolbox: How to Measure Financial Literacy

2.1 Introduction

This chapter will focus on deriving a "toolbox" of experimental and econometric methods that will be used throughout this thesis. It will introduce the experimental design and financial literacy questions being examined throughout Chapters 2, 3, and 4, and provide results for a group of control subjects. It develops a theoretical framework of how and why risk attitudes and the elicitation methods for subjective belief distributions are used, along with a discussion of their properties. There is a detailed discussion on how to interpret the econometric estimations and draw inferences from those experimental tasks.

2.2 Literature Review

2.2.1 Financial Literacy

After conducting a review of current literature, an initial 16 questions were identified that test financial and statistical knowledge that were both topically relevant and well fit to be answered using subjective belief elicitation.[8] Questions were drawn from various sources throughout the literature and Table 2.1 provides a listing of those questions, the correct answer at the time of the initial ask, and the two different bin labeling schemes used in this research; see Appendix A for more information on labels. An asterisk is placed beside the correct answer in both the initial and new bin labels. All questions were asked to university students in the United States and answers are US-specific.

Table 2.1 - Financial Literacy Questions, Answers, and Bin Labels

- **fin1**: 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?
  Actual: $110.41
  Initial bin labels: [$92, $94, $96, $98, $100, $102, $104, $106, $108, $110*]
  Newer bin labels: [$102, $104, $106, $108, $110*, $112, $114, $116, $118, $120]

[8] "Fit" in this context means that each question has an objectively correct response and could be incorporated into the software display. Further, the financial literacy questions that were selected for this research was guided by the types of questions used in the literacy study by Di Girolamo et al. (2015), which was circulating as a CEAR working paper in 2013. They introduced the idea of asking questions that evaluated knowledge of facts, as one component of a broader notion of literacy.

- **fin2**: What is the earliest age at which you can start Social Security benefits?
  Actual: 62 (US-specific)
  Initial bin labels: [54, 56, 58, 60, 62*, 64, 66, 68, 70, 72]
  Newer bin labels: [62*, 64, 66, 68, 70, 72, 74, 76, 78, 80]

- **fin3**: What is the age that you can receive a full or unreduced Social Security benefit?
  Actual: 66 (US-specific)
  Initial bin labels: [54, 56, 58, 60, 62, 64, 66*, 68, 70, 72]
  Newer bin labels: Not applicable, this question was dropped.

- **fin4**: If you start Social Security benefits at the earliest possible age, you will receive a benefit that is X percent of the benefit that you would have received at the normal retirement age. What is X?
  Actual: 75% (US-specific)
  Initial bin labels: [55%, 60%, 65%, 70%, 75%*, 80%, 85%, 90%, 95%, 100%]
  Newer bin labels: Not applicable, this question was dropped.

- **fin5**: What is the earliest age at which you will be eligible for Medicare?
  Actual: 65 (US-specific)
  Initial bin labels: [55, 57, 59, 61, 63, 65*, 67, 69, 71, 73]
  Newer bin labels: [49, 51, 53, 55, 57, 59, 61, 63, 65*, 67]

- **fin6**: Suppose you had $100 in a savings account and the interest rate was 20 percent per year, and you never withdraw money or interest payments. After 5 years, how much would you have in this account in total?
  Actual: $248.60
  Initial bin labels: [$90, $100, $125, $148, $173, $196, $207, $233, $249*, $271]
  Newer bin labels: Not applicable, this question was dropped.

- **fin7**: Imagine that you have $100 in a savings account and the annual interest rate on your savings account was 1 percent per year, and annual inflation was 2 percent per year. After one year, how much purchasing power would you have on the initial $100?
  Actual: $98.98
  Initial bin labels: [$95, $96, $97, $98, $99*, $100, $101, $102, $103, $104]
  Newer bin labels: [$98, $99*, $100, $101, $102, $103, $104, $105, $106, $107]

- **fin8**: Assume that you have $200 in a savings account, and the interest rate that you earn on these savings is 10 percent a year. How much would you have in the account after two years?
  Actual: $242.00
  Initial bin labels: [$200, $220, $231, $240, $242*, $266, $293, $322, $330, $341]
  Newer bin labels: Not applicable, this question was dropped.

- **fin9**: David just found a job with a take-home pay of $2,000 per month. He must pay $900 for rent and $150 for groceries each month. He also spends $250 per month on transportation. If he budgets $100 each month for clothing, $200 for restaurants, and $250 for everything else, how long will it take him to accumulate savings of $600?
  Actual: 4 months
  Initial bin labels: [1 month, 2 months, 3 months, 4 months*, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months]
  Newer bin labels: [4 months*, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 13 months]

- **fin10**: If your credit card is stolen and the thief runs up a total debt of $1,000, but you notify the issuer of the card as soon as you discover it is missing, what is the maximum amount that you can be forced to pay according to Federal law?
  Actual: $50 (US-specific)
  Initial bin labels: [$0, $25, $50*, $100, $250, $500, $1000, $2500, $5000, $10000]
  Newer bin labels: [$0, $10, $20, $30, $40, $50*, $100, $250, $500, $1000]

- **fin11**: If your ATM or debit card is stolen and the thief runs up a total debt of $1,000, and you notify the bank 5 days later, what is the maximum amount that you can be forced to pay according to Federal law?
  Actual: $500 (US-specific)
  Initial bin labels: [$0, $25, $50, $100, $250, $500*, $1000, $2500, $5000, $10000]
  Newer bin labels: [$0, $50, $500*, $750, $1000, $2500, $5000, $7500, $1000, $2500]

- **fin12**: On average, if you went to college and earned a four-year degree, how much more money could you expect to earn than if you only had a high school diploma? (For example, 1X would be no increase, and 2X would be twice as much)
  Actual: 7X (US-specific)
  Initial bin labels: [The Same, 1.1X, 1.3X, 1.5X, 1.7X, 1.9X, 2X, 4X, 7X*, 10X]
  Newer bin labels: Not applicable, this question was dropped.

- **fin13**: You lend $100 to a friend one evening and he gives you $105 back the next day. How much interest has he paid on the loan?
  Actual: 5%
  Initial bin labels: [0%, 1%, 2%, 3%, 4%, 5%*, 6%, 7%, 8%, 9%]
  Newer bin labels: [-4%, -3%, -2%, -1%, 0%, 1%, 2%, 3%, 4%, 5%*]

- **fin14**: Suppose you put $100 into a savings account with a guaranteed interest rate of 2% per year. You don't make any further payments into this account and you don't withdraw any money. How much would be in the account at the end of the first year, once the interest payment is made?
  Actual: $102
  Initial bin labels: [$98, $99, $100, $101, $102*, $103, $104, $105, $106, $107]
  Newer bin labels: [$95, $96, $97, $98, $99, $100, $101, $102*, $103, $104]

- **fin15**: Based on 2006 statistics, if a man lived to be 20 in the United States, how many more years would he expect to live? Note that this is not asking for the age he would die at, but rather how many more years he would expect to live?
  Actual: 56.1 years (US-specific)
  Initial bin labels: [0, 10, 20, 30, 40, 50, 60*, 70, 80, 90]
  Newer bin labels: [15, 20, 25, 30, 35, 40, 45, 50, 55*, 60]

- **fin16**: Based on 2006 statistics, if a woman lived to be 20 in the United States, how many more years would she expect to live? Note that this is not asking for the age she would die at, but rather how many more years she would expect to live?
  Actual: 61 years (US-specific)
  Initial bin labels: [0, 10, 20, 30, 40, 50, 60*, 70, 80, 90]
  Newer bin labels: [15, 20, 25, 30, 35, 40, 45, 50, 55, 60*]

The order of presentation of questions was held constant across all subjects. Because several of the questions related to each other, and we were concerned that altering their order for different subjects would cause a potential confound when comparing results over subjects. The sections below provide background on each question and is ordered by category. Every question was not asked in every session due to the time allotted for different session, some questions were dropped from analysis for reasons explained below.

Interest and Inflation

Several of the questions in our battery tested subject's knowledge about the concepts of interest and inflation. These questions covered calculations for the nominal rates of simple interest (fin13, fin14) and compound interest (fin1, fin6, fin8), and the real rate of interest adjusting for inflation (fin7).

Questions fin13 and fin14 are adapted from an OECD International Network on Financial Education pilot study undertaken in 14 countries. That data were collected in 2010 and 2011 and administered to a nationally representative sample of adults over 18 in each country. The study was undertaken to identify needs and gaps in financial education and develop national policies to raise literacy levels. Our fin13 question adapts the following question of theirs "You lend X to a friend one evening and he gives you X back the next day. How much interest has he paid on the loan?" The value of X was tailored to the local currency and price levels. Responses to their question were open ended, and if the participant answered correctly the data were coded as a 1, otherwise a 0. Atkinson and Messy (2012) report that 95% of the subjects from Hungary answer the "interest paid on a loan" question correctly, while only 60% from the British Virgin Islands give the correct answer. Our fin14 question is asked with exactly the same wording as their question. Responses to their question were open ended and recorded in the same manner as described above. For this "calculation of interest plus principal" question, Atkinson and Messy (2012) report Ireland as the country with the highest percentage of respondents giving the correct answer at 76%, while Albania and Peru tied for the lowest with only 40% of the respondents able to give the correct answer.

The questions fin1 and fin7 are natural extensions of questions asked by Lusardi and Mitchell

(2007, 2008) and used in the *Health & Retirement Survey* (HRS) of 2004 in the United States.[9] This survey is naturally representative of Americans over the age of 50. Our fin1 question adapts the following question of theirs: "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: more than $102, exactly $102, less than $102?" Our fin7 question adapts this question of theirs: "Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?" The main difference for both questions is that we ask for beliefs about the true answer over a wide range and allow for reports over that range. Lusardi and Mitchell (2011; Table 2.1) report that only 67.1% and 75.2% of their sample gave the correct response to each question, respectively. These fractions drop significantly (their Figures 2.1a and 2.1b) as one considers Black and Hispanic respondents. When the same questions were posed to a nationally representative sample of young Americans, aged between 22 and 28 in Wave 11 of the *National Longitudinal Survey of Youth* conducted in 2007-2008, 79.3% and 54.0% gave the correct responses to the interest rate and inflation questions, respectively (Lusardi, Mitchell, and Curto (2010, Table 1, p. 365)).[10]

Questions fin6 and fin8 were dropped after the first phase of data collection in our study, since both questions require the same compound interest calculation needed to answer fin1. These three questions, fin1, fin6, and fin8, only differed by the initial dollar amount, the annual interest rate, and the time horizon. Because these questions require the same calculation, and time in the experimental lab was at a premium, we decided to ask only fin1 throughout the remainder of the project. By way of background, fin6 is the natural adaptation of a question that was first used in the Module for Financial Literacy from the *RAND: American Life Panel MS5: Retirement Decisions and Saving for Retirement* survey.

[9] A third question they asked was: *Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund."* This question was posed in order to understand if individuals know about diversification and risk. The 3 questions from the HRS are often referred to as "The Big 3" in the financial literacy literature. In a later Dutch national survey van Rooij, Lusardi, and Alessie (2011) increased the set of questions posed to individuals. Apart from 5 questions aimed at characterizing "basic" financial literacy (p. 452), they added 11 questions to characterize "advanced" financial literacy (p. 454). Similar extensions were undertaken by Bateman et al. (2012) in surveys in Australia.

[10] Bateman et al. (2012) ask these questions of adult retirement savers in Australia, and find that 78.4% correctly answer the inflation question and 71.8% correctly answer the interest rate question.

Question fin8 was the adaptation of the question used by the *Chilean Social Protection Survey*, as reported in (Mitchell, Hastings, and Chyn 2011).

Rules, Regulations, and Procedural

Our study took place in the United States, so several questions in our battery tested subject's knowledge about rules, regulations, and procedures unique to that country. Study participants were asked about their knowledge of Social Security benefits (fin2, fin3, fin4), Medicare (fin5), credit and debit cards (fin10, fin11), and expected returns to education (fin12).

The United States Social Security Administration (SSA) is an independent agency of the U.S. Federal government that administers Social Security, a social insurance program consisting of retirement, disability, and survivors' benefits. To qualify for these benefits in the typical manner, most workers pay Federal Insurance Contributions Act (FICA) taxes on their earnings; the claimant's benefits are based on the wage earner's contributions. According to a 2017 report[11] by the SSA, social security benefits account for 33% of aggregate total income of couples and nonmarried persons aged 65 or older in 2015. The other sources of aggregate income included 34% for current earnings, 9% for asset income, 8% for government employee pensions, 12% from private pensions, and 4% other. What is noteworthy is that, on average and for those taking social security, one-third of retirement income is funded from those payments.

The age at which full (normal) social security can be claimed has varied from 65 to 67, and is currently 67 years old for those born in 1960 and later. However, the *earliest* age at which social security can be claimed has remained at 62 years old. Additionally, the percentage reduction in Social Security benefits if taken at the earliest possible age (62) has varied from a 20 percent to a 30 percent reduction, and is currently at a 30 percent reduction for those born in 1960 and later.[12]

[11] Available at www.ssa.gov under SSA Publication No. 13-11785.
[12] Benefits by year of birth can be found at https://www.ssa.gov/planners/retire/agereduction.html.

Our questions fin2, fin3, and fin4 explore subject's knowledge of social security. However, only question fin2 was kept throughout the study and included in the analysis. Questions fin3 and fin4 had a potential confound by not explicitly stating for what age range the question was being posed. A simple qualifier at the beginning of each of these question, such as: "For persons born in 1960 and later, …", would have eliminated this confound, but as some participants expressed confusion it was decided to exclude questions fin3 and fin4 from analysis.

Medicare is a national health insurance program administered by the U.S. Federal government that is available to U.S. citizens over the age of 65, regardless of income or medical history. It is funded through multiple sources including a federal income tax, general revenue, and premiums. According to the *2017 Annual Report to the Medicare Board of Trustees*, Medicare covered 47.8 million people aged 65 and older. Our fin5 question asks subjects if they know at which age they can normally enroll in Medicare.

Our fin10 and fin11 questions ask if subjects know how much exposure and personal liability is should their personal credit or debit card is stolen, respectively. Our fin10 question is adapted from Mandell (2008) as reported in the *Jump$tart* survey: "If your credit card is stolen and the thief runs up a total of $1,000, but you notify the issuer of the card as soon as you discover it is missing, what is the maximum amount that you can be forced to pay according to Federal law?" Answers were in multiple choice format and given as a) $500, b) $1000, c) Nothing, and d) $50. The correct response is $50, and on average only 13% of students in the 2008 study answered this correctly. Hilgert, Hogarth, and Beverly (2003) asked a true or false version of this question in a quiz administered as part of the *Survey of Consumers*: "If your credit card is stolen and someone uses it before you report it missing, you are only responsible for $50, no matter how much they charge on it." The percentage of respondents answering correctly was 50%. Our fin11 question was constructed similarly, but takes advantage that U.S. Federal Law is slightly different with respect to the maximum amount for which one can be held personally liable when using an ATM or debit card.

Our fin12 question ask if subjects understand the average return for obtaining a 4-year college degree compared to obtaining only a high school diploma. It is also adapted from Mandell (2008) as reported in the *Jump$tart* survey: "If you went to college and earned a four-year degree, how much more money could you expect to earn than if you only had a high school diploma?" However, the fin12 question was dropped from the study after piloting it in a few sessions, primarily for timing purposes.

Budgeting

Our question fin9 presents participants with information about a basic budget and then asks them to calculate how long it will take to accumulate a targeted amount of savings. The question was also adapted from Mandell (2008) as reported in the *Jump$tart* survey: "David just found a job with a take-home pay of $2,000 per month. He must pay $900 for rent and $150 for groceries each month. He also spends $250 per month on transportation. If he budgets $100 each month for clothing, $200 for restaurants and $250 for everything else, how long will it take him to accumulate savings of $600." Answers were in multiple choice format and given as a) 3 months, b) 4 months, c) 1 month, and d) 2 months. The correct response is 4 months. For the 2008 survey, 60.2% of high school students and 77.8% of college students answered correctly.

Longevity Risk

The final two questions, fin15 and fin16, ask about a basic informational input to retirement planning: expected remaining lifetime, conditional on reaching the age of 20. Smith, Taylor, and Sloan (2001, p. 1126) call this "the most important subjective risk assessment a person can make," although they were referring to own-mortality. We separate out the question for men and women, to ascertain if the differential expected mortality between the two is recognized by individuals. These questions do not condition on the health, income, or any other relevant characteristics of the individual that would affect expected mortality. One could extend these questions to elicit more precise beliefs about someone more closely resembling the subject by conditioning on their age, gender, ethnicity, etc.

The most widely used subjective beliefs about longevity come from the *Health and Retirement Survey*, which has asked a simple question for respondents under the age of 65 since 1992: "With 0 representing absolutely no chance, and 100 absolute certainty, what is the chance that you will live to be 75 years of age or older?". A comparable question asks the chance that they would live to be 85, and for respondents *over* 65 a variant asked the chances of them living 11-15 years more. In the 2006 wave of the *Health and Retirement Survey* a sub-sample was asked questions that elicited their beliefs about the population life tables: "Out of a group of (men/women) your age, how many do you think will survive to the age of X?" The value of X was 75 for those under 65, and 11-15 years older for those over 65. These questions are closer to those we asked, although we only conditioned on the single age 20.

Of course, the questions in the *Health and Retirement Survey* are not incentivized, and do not elicit information on the confidence of the subjective belief. Smith, Taylor, and Sloan (2001) show that responses to this question are reasonably good predictors of future, actual mortality, even if they do not perfectly reflect new health information when updated. Perozek (2008) makes an even stronger case for the predictive value of these subjective belief questions, arguing that responses to these questions actually outperform population life tables. In contrast, Elder (2013) stresses that only with the 2006 wave can one evaluate the actual predictions, as early respondents reach the target ages of 75 or 85. And in that respect he presents a sharply contrary view, arguing that the evidence supports a "flatness bias," a "tendency for individuals to understate the likelihood of living to relatively young ages while overstating the likelihood of living to ages beyond 80." He attributes this bias to a failure to recognize that mortality risk increases with age.

Recap

To recap, we began with 16 financial literacy questions spanning topics of interest and inflation, rules, regulations, and procedures, budgeting, and longevity risk. These questions were identified from reviewing the current literature and adapted to the method of subjective belief elicitation used here. While we began with 16 questions, only 11 questions were used in the final analysis. Of the 5 questions that were dropped, questions fin3, fin4, and fin12 were dropped due to time restrictions in the lab and questions fin6 and fin8 were dropped due to possible ambiguity in the wording of the question. The remaining 11 questions are used in the analysis throughout this chapter and in chapters 3 and 4.

2.2.2 Theoretical Framework

The theoretical framework used throughout this research was first introduced in Section 1.3. This section will further develop that framework and show the complementary nature of why it is needed to jointly elicit risk preferences as well as the subjective beliefs of subjects. Following Harrison and Ulm (2016) each individual subject's risk preferences is used to recover latent beliefs form the scoring rules employed.

Atemporal Risk Attitudes

Consider the evaluation of a risky prospect consisting of a poor outcome x that will occur with known probability p, and a good outcome y that will occur with probability (1-p). Assume that this risky outcome will be resolved at a point in time, typically the present. Choices over risky prospects of this kind can be viewed as arising from the maximization of expected utility (EU) defined by:

$$EU = p \times U(x) + (1-p) \times U(y), \quad (1)$$

where U is a utility function.

The psychological process underlying EUT theory is that the decision-maker has attitudes towards the variability of outcomes. The notion of "variability" includes variance, skewness and kurtosis. Someone might be averse to mean-preserving variance ($U''<0$), skewness ($U'''>0$), and even kurtosis ($U''''<0$). However, even if someone is averse to variance, it is possible that they are attracted to skewness.[13]

If one relaxes the assumption that individuals and households employ Expected Utility Theory (EUT) to evaluate atemporal risk, then there are several ways to model their choices. One popular alternative is Rank Dependent Utility (RDU), due to Quiggin (1982), which posits that individuals weight the decumulative probabilities of outcomes using some Probability Weighting Function (PWF) reflecting

[13] The simplest mathematical forms for utility functions rule out such complexities, but they remain a part of EUT.

"optimism" or "pessimism" towards better outcomes. With non-satiation, U(y) > U(x), so we can instead of (1) consider the decision to engage in a risky act as guided by the evaluation

$$RDU = (1-w(1-p)) \times U(x) + w(1-p) \times U(y), (2)$$

where w(p) is some probability weighting function.[14] For instance, if $w(p) = p^{\gamma}$ and $\gamma>1$, then it can be shown that the decision-maker has a positive risk premium even if the utility function is linear, since the decision-maker is "pessimistic" about the better outcome occurring.[15] It is immediate from (2) that optimism or pessimism about the probability can drive behavior, in addition to any effect from convexity or concavity of the utility function used to evaluate final outcomes. In short, to know if individuals or households are risk averse, we need to know both their utility function and their PWF.

Just from these two alternatives we can see that the identification of the risk preferences involves several psychological processes: aversion to variability of outcomes, probability or pessimism towards probabilities, and desires to beat the odds.

Subjective Beliefs

Our interest in eliciting subjective beliefs is so that we can characterize how literate individuals are in certain domains. Characterizing and measuring the literacy of an individual requires then that we have some way of assessing *how* knowledgeable the person is about certain topics. There are some topics about which one can have "crisp" knowledge, in the sense of Boolean truth values. However, there are many domains of knowledge that one naturally expects varying levels of precision. We characterize literacy in terms of the subjective beliefs that someone has over possible responses to some question. By eliciting the subjective belief *distribution*, rather than just the answers to true/false or multiple choice

[14] We have written (2) in the slightly awkward form in which w(·) is directly applied to 1-p to generate the decision weight w(1-p) from the probability of the higher-ranked outcome y, and then the lower-ranked outcome is the residual decision weight. When there are only two outcomes we could have written (2) as w(p) U(x) + w(1-p) U(y) since the decision weights sum to 1. The approach in (2) generalizes to cases in which there are 3 or more outcomes: the decision weight for the highest-ranked outcome is just the weighted probability of that outcome, the decision weight for the second-highest outcome is the weighted probability of at least that outcome (i.e., it is the weighted cumulative probability of at least the second-highest outcome), and so on. The decision weight for the lowest-ranked outcome is the residual decision weight.
[15] This is true when ranking the options from best to worst.

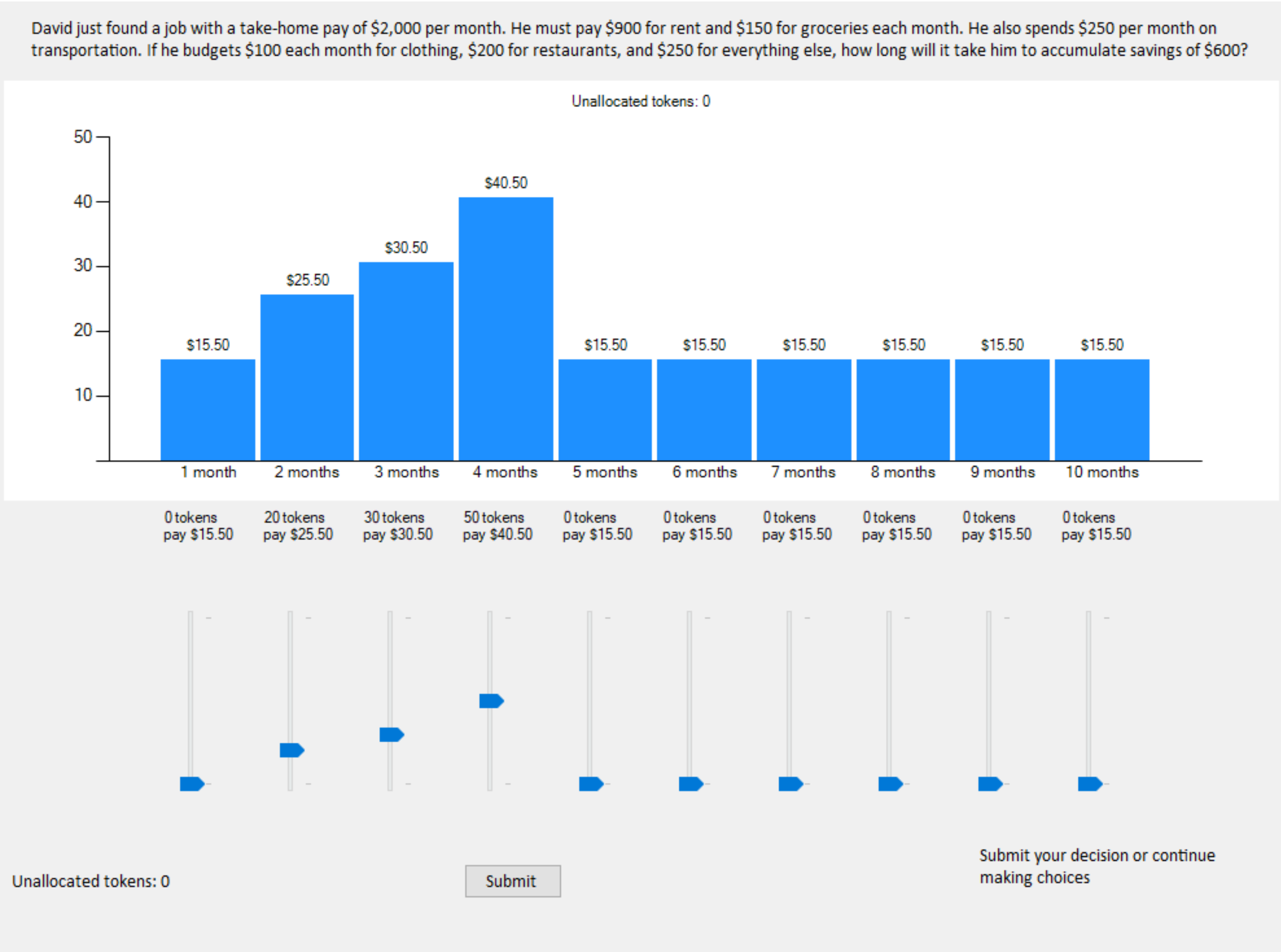
questions (e.g., Lusardi and Mitchell 2011 and 2014), we can directly measure the confidence that an individual has about their knowledge of some fact.

Following Savage (1971, 1972), we *define* subjective beliefs by the choices that individuals make when facing bets whose outcomes depend on those beliefs. The measurement of the literacy that someone has in a specific domain entails the elicitation of their subjective beliefs. For that task we conduct an experiment using proper scoring rules, which are simply structured bets offered to the individual by an observer (the experimenter). All of the elicited beliefs were incentivized and incentive-compatible, so that the subjects were making real choices with real economic consequences.[16]

We elicit subjective beliefs about the answers to 11 specific questions asked of each subject. In each case there is a correct answer, and responses were elicited over a continuous range of possible answers presented in terms of 10 intervals or "bins." A computer interface was used to present the belief elicitation tasks to subjects and record their choices, allowing them to allocate tokens in accordance with their subjective beliefs. Figure 2.1 presents the interface. The interface implements the Quadratic Scoring Rule (QSR) discussed below. Subjects could move the sliders at the bottom of the screen to re-allocate the 100 tokens as they wished, ending up with some distribution.

[16] Harrison (2014a) presents evidence that one cannot rely on hypothetically elicited subjective belief distributions to generally provide the same responses as appropriately incentivized methods.

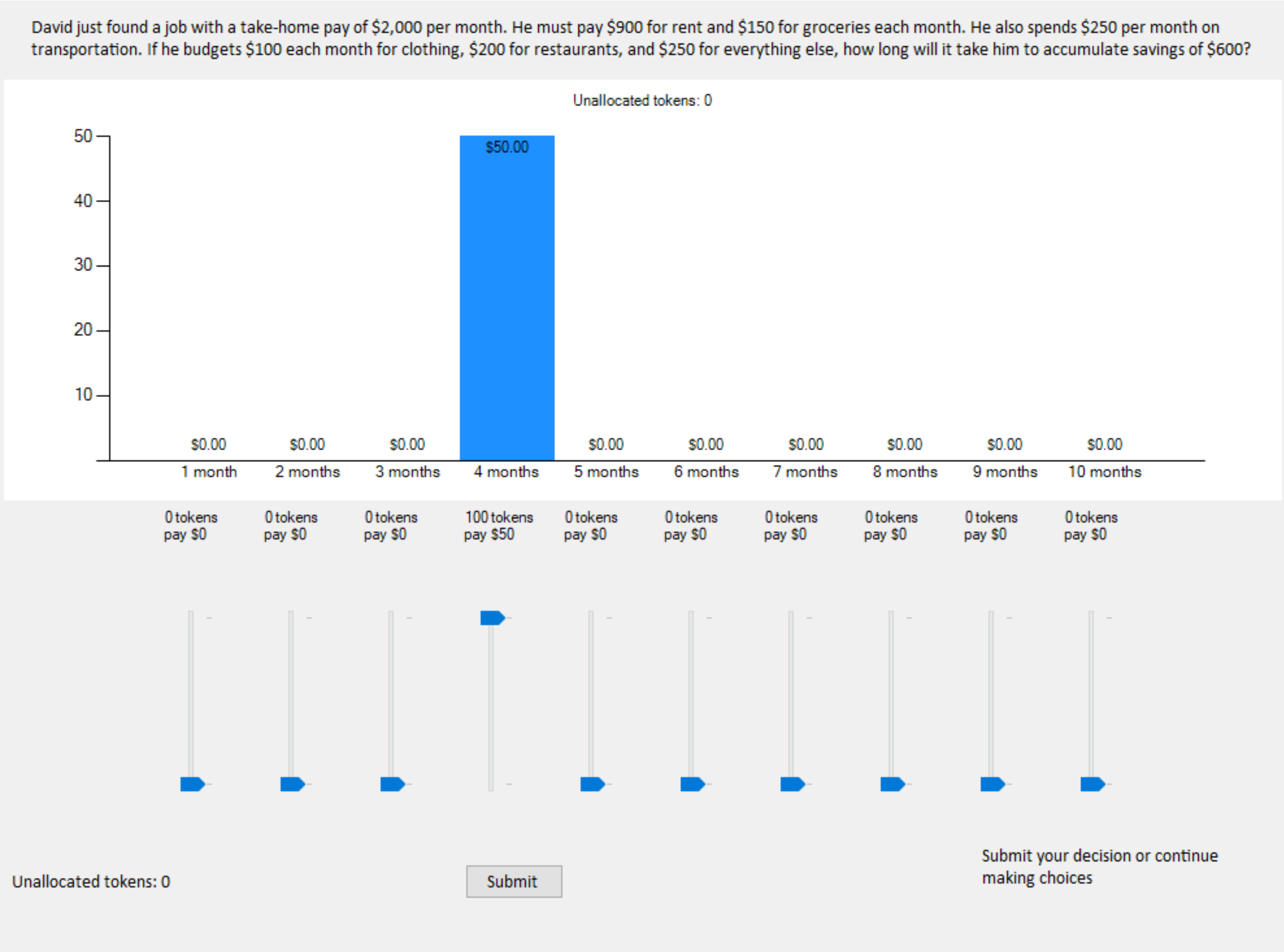
Figure 2.1 - Illustration of Subjective Belief Interface, Potential Token Allocation for *fin9*



The instructions explained that they could earn up to $50 dollars, as shown in Figure 2.2, but only by allocating all 100 tokens to one interval *and* that interval contains the true answer: if the true answer was just outside the selected interval, they would in that case receive the payout indicated on the top of the slider bar for the bin containing the true answer.[17]

[17] The software is also capable of an alternative display that shows points instead of cash payouts. These displays would be used if applying the "binary lottery procedure" to induce risk neutrality, following Harrison, Martínez-Correa, and Swarthout (2014). However, our experiments also elicited a measure of risk for each individual, and corrected for those risk attitudes as explained later.

Figure 2.2 - Illustration of Subjective Belief Interface, All Tokens Allocated to Correct Interval for *fin9*



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 was revealed, and a participant's earnings recorded. For example, if the respondent had reported the beliefs in Figure 2.1, they would have been paid $25.50 if the correct answer was 2 months. As it happens, the correct answer here is 4 months, so the subject would have actually received $40.50.

The decision maker in our experiment reports their subjective beliefs with a discrete version of a QSR for continuous distributions, developed by Matheson and Winkler (1976).[18] Partition the domain into K intervals, and denote as r_k the report of the density in interval $k = 1, \ldots, K$. Assume for the

[18] This formal exposition closely follows Di Girolamo et al. (2015)

moment that the decision maker is risk neutral, and that the full report consists of a series of reports for each interval, $\{ r_1, r_2, \ldots, r_k, \ldots, r_K \}$ such that $r_k \geq 0 \ \forall k$ and $\sum_{i=1 \ldots K} (r_i) = 1$.

If k is the interval in which the true value lies, then the payoff score is from Matheson and Winkler (1976, p. 1088, equation (6)):

$$S = (2 \times r_k) - \sum_{i=1 \ldots K} (r_i)^2$$

The reward in the score is a doubling of the report allocated to the true interval, and a penalty that depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval the punishment is severe. The punishment includes all possible reports.

Consider some examples, assuming $K = 4$. What if the subject has very tight subjective beliefs and puts all of the tokens in the correct interval? Then the score is

$$S = (2 \times 1) - (1^2 + 0^2 + 0^2 + 0^2) = 2 - 1 = 1,$$

and this is positive. But if the subject has a tight subjective belief that is wrong, the score is

$$S = (2 \times 0) - (1^2 + 0^2 + 0^2 + 0^2) = 0 - 1 = -1,$$

and the score is negative. So we see that this score would have to include some additional "endowment" to ensure that the earnings are positive.[19] Assuming that the subject has a very diffuse subjective belief and allocates 25% of the tokens to each interval, the score is less than 1:

$$S = (2 \times \tfrac{1}{4}) - (\tfrac{1}{4}^2 + \tfrac{1}{4}^2 + \tfrac{1}{4}^2 + \tfrac{1}{4}^2) = \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4} < 1.$$

The tradeoff from the last case is that one can always ensure a score of ¼, but there is an incentive to provide less diffuse reports, and that incentive is the possibility of a score of 1.

[19] This is a point of practical behavioral significance, but is not important for the immediate theoretical point.

To ensure complete generality, and avoid any decision maker facing losses, allow some endowment, α, and scaling of the score, β. We then get the generalized scoring rule

$$\alpha + \beta \, [\, (2 \times r_k) - \textstyle\sum_{i=1\ldots K} (r_i)^2 \,]$$

where we initially assumed α=0 and β=1. We can assume different values of α and β to transform the payoffs to any alternative range of levels we may want.

In our experiment $K = 10$, and we do not know whether the subject is risk neutral. Indeed, the weight of evidence from past experiments clearly suggests that subjects will be modestly risk averse over the prizes they face. It is well-known that risk aversion can significantly affect inferences from applications of the QSR when eliciting subjective *probabilities* over *binary* events (Winkler and Murphy 1970; Kadane and Winkler 1988), and there are various methods for addressing these concerns. Harrison et al. (2017) characterize the implications of the general case of a risk averse agent when facing the QSR and reporting subjective *distributions* over *continuous* events, and find, remarkably, that these concerns do not apply with anything like the same force. For empirically plausible levels of risk aversion, one can reliably elicit the most important features of the latent subjective belief distribution without undertaking calibration for risk attitudes.

Specifically, they draw the following conclusions:

1. An individual reports having a positive probability for an event only if he has positive subjective probability for the event. We can infer from Figure 2.1, for instance, that this subject truly attaches zero weight to the possibility of a savings horizon 5 months or longer, no matter what his risk attitudes.
2. If an individual has the same subjective probability for two events, then the reported probabilities for the two events will also be the same if the individual is risk averse or risk neutral.
3. The converse is true for risk averse subjects, as well as for risk lovers. That is, if we observe two events receiving the same reported probability, we know that the true probabilities are also equal, although not necessarily the same as the reported probabilities.

4. If the individual has a *symmetric* subjective distribution, then the reported mean will be *exactly* the same as the true subjective mean, whether or not the subjective distribution is unimodal. Hence if we simply assume symmetry of the true distribution, a relatively weak assumption in some settings, we can elicit the mean belief directly from the average of the reported distribution.
5. The more risk averse an agent is, the more the reported distribution will resemble a uniform distribution defined on the support of their true distribution. In effect, risk aversion causes the individual to report a "flattened" version of their true distribution, but never to report beliefs to which they assign zero subjective probability.
6. It is possible to derive the effect of increased risk aversion on the difference between the reported distribution and true distribution. Harrison et al. (2017) show numerically that *a priori* plausible levels of risk aversion in laboratory settings implies no significant deviation between reported and true subjective beliefs in this setting.

Provided that our subjects exhibit the modest levels of risk aversion that are typically found in lab settings with similar stakes, these results provide the basis for using the reported distributions as if they are the true, subjective belief distributions.

One of the maintained assumptions in our main data analysis, then, is that the responses of subjects to the belief elicitation questions can be taken at face value as revealing the true subjective beliefs of the individual (to a reasonable approximation). One alternative assumption is that subjects exhibit RDU preferences over risk, in which case we are unable to take these responses at face value in all respects. To evaluate that alternative assumption our experimental design included binary lottery choice questions for each subject, with lotteries defined over objective probabilities. So the alternative, but significantly weaker, assumption is that evidence for RDU preferences over objective probabilities is evidence for RDU (or non-SEU) preferences over subjective probabilities. If we accept this assumption for some individuals, we can filter them out and see if it changes our inferences based on the average elicited subjective beliefs (Harrison 2014a).

Our approach to eliciting subjective belief distributions allows a rigorous characterization of the concept[20] of "overconfidence," widely used in the behavioral economics literature, since it provides information on the confidence that the individual holds certain beliefs. Moore and Healy (2008) and Merkle and Weber (2011) review the vast literature, and explain why one needs information on the distribution of beliefs to measure overconfidence relative to the "rational Bayesian updating" explanation offered by Benoît and Dubra (2011).

Recovering a Participant's True Beliefs

Once a participant has reported their beliefs to the scoring rule, one faces the question of recovering the latent subjective belief distribution underlying those reports. There are four approaches that will be adopted here to recover the latent subjective beliefs and all are at the level of the individual. They differ in terms of the importance of identifying theoretical assumptions, from the most restrictive to the least restrictive.

The first approach is to simply assume that every subject behaves consistently with SEU theory. In this case, we know from the results in Harrison, Martínez Correa, Swarthout and Ulm (2017) that subjects that exhibit the typical risk attitudes found in the laboratory will generate reports that are extremely close to their true, latent subjective belief distribution. This is a remarkable result, particularly in contrast to the results obtained for the elicitation of a subjective probability over a binary event. In the elicitation of a subjective probability even modest amounts of risk aversion generate significant distortions in the elicited probability compared to the observed report. The intuition is that risk aversion causes the subject to report a belief that is closer to a 50% belief, because at a 50% report the subject is equalizing the payoff for the two states of nature that are possible. By reporting closer to 50% then the subject actually believes, the subject is reducing the variability of payouts from the scoring rule depending on which of the binary events actually occurs. In the extreme, the most risk-averse subject will literally report 50% and remove all variability in payouts. The net result of this distortion from risk

[20] To be precise, the notion of "overplacement," or the "better-than-average" effect in which more than 50% of a group of individuals think themselves to be better than the average of that group in something.

attitudes, in the case of elicitation of a subjective probability of a binary event, is to cause reports to be left biased or right biased relative to true beliefs. However, when eliciting subjective belief distributions over a continuous event, which is what is being done here, the effect of risk aversion under SEU is to cause the reports to be "flattened" relative to true beliefs. In other words, the subject is reducing the variability of payoffs across the states of nature that he thinks will occur, and if there are four or five discrete states of nature that he thinks will occur, he will reduce the variability over those four or five states by equalizing the number of tokens allocated to each state. So instead of the distortion causing a left biased or right biased adjustment relative to true beliefs, it causes a flattening adjustment relative to true beliefs. This results in reports being much closer to true beliefs, at least for the standard levels of risk aversion found in the laboratory.

This is a valuable theoretical result, but rests on the assumption that the individual is indeed behaving consistently with SEU. In effect, that strong theoretical assumption is "buying" a great deal of identification: if we can make that assumption, then effectively the observed reports can be assumed to be the true, latent subjective beliefs. What does one do if one relaxes that assumption that the subject behaves consistently with SEU? The remaining three approaches consider various ways of relaxing that assumption.

One way to relax that assumption is to undertake a separate experimental task in which one determines if the subject is behaving consistently with respect to EUT or some alternative model of decision-making under risk. We should distinguish between SEU and EUT because the typical task that is used for this purpose is one in which there is objective risk, generated by physical dice, rather than subjective risk. We then make the assumption that if the subject is behaving consistently with respect to EUT, then it is reasonable to assume that they behave consistently with respect to subjective beliefs and hence behave consistently with SEU. Assume that the alternative specification to EUT is the popular Rank Dependent Utility (RDU) specification. In this case our approach boils down to deciding if the individual subject is better characterized, from their choices over risky lotteries, as an EUT subject or as a RDU subject. The simplest way to make that determination is to estimate the structural model of risk preferences for the individual, again based on their responses to the choices over risky lotteries. We discuss the econometric specification of this step in section 2.5. Because EUT is nested within RDU, we would estimate the more general model for each individual and then test the hypothesis that the estimates are consistent with EUT. That hypothesis boils down to the assumption that the subject does not engage in any "probability weighting." That hypothesis can be formally tested once we have econometric estimates of the RDU model for the individual. We can then adopt some conventional critical level at which we determine if the null hypothesis that the subject behaves consistently with EUT is rejected. If we use a 5% critical level, then if the p-value for the hypothesis test provides a value that is 5% or below, then we classify the subject as being consistent with RDU. On the other hand, if the p-value for the hypothesis test provides a value that is greater than 5%, we cannot reject the null hypothesis, and we conclude that the subject behaves consistently with EUT. Once we have classified each subject in this manner, we can simply repeat the analysis from the first approach but deleting the subjects who are not deemed to be behaving consistently with SEU. In general terms, based on evidence for samples drawn from the same population in Harrison and Ng (2016) and Harrison and Ulm (2016), we expect that approximately 50% of the subjects will be characterized as being consistent with EUT in this manner.

The third approach is a variant on the second approach just presented, but where we ascertain a probability that the subject is an SEU subject. The second approach simplifies this by assuming that we have classified the subject as 100% consistent with SEU, or 0% consistent with SEU, based on the critical value of the hypothesis test (assumed to be 5% above). A more flexible approach is to recognize that the hypothesis test provides information on the strength of the assumption that the subject is an SEU subject, and that by transforming that information on the strength of the null hypothesis into an "all or nothing" classification, we are in fact throwing valuable information way. If there is some way from the statistical hypothesis test to ascertain the probability that a given subject is an SEU subject, then this approach would simply reweight the responses used in the first approach. It would reflect the probability that the subject is behaving consistently with those reports actually being a close approximation of the true latent subjective probability belief distribution. It is an extremely easy matter to reweight observations in any statistical analysis, and one sacrifices virtually nothing by using this approach in terms of the range of econometric specifications that can be examined and hypothesis tests that can be conducted. But one does gain the use of all observations in the sample, whereas the second approach would, under the assumptions noted about the fraction of subjects that we find in this population to be consistent with SEU, lead us to drop about 50% of the sample.

The final approach is to undertake more sophisticated corrections to the reported subjective belief distributions in order to "recover" the true latent subjective belief distribution, even when one assumes that the subject behaves consistently with RDU. In this instance we use the theoretical results, and numerical methods, of Harrison and Ulm (2016). They show that if the subject does behave consistently with RDU, then you will also experience a left bias or right bias in the elicited subjective probability belief distribution. One still obtains the result, for perfectly intuitive reasons, that the subject will never report a positive belief over an interval that the subject actually assigns zero subjective probability weight to, even if the subject is a probability weighter. What will happen is that the subject will "skew" the reports to the left or to the right, depending on the nature of the probability weighting that is employed. But the subject will never "shift" the reports to the left or to the right, to include intervals that the subject has assigned a zero probability to subjectively. If one knows the parameters of the RDU specification for the subject, it is possible to recover the exact latent subjective probability distribution from the observed reports.

It is also possible to recover a distribution over these recovered distributions, reflecting the fact that the initial estimate of the RDU model is stochastic. In other words, we may have estimates of the parameters of the RDU model for a given subject, but we must recognize that those parameters have standard errors, and that imprecision in the estimate of the parameters will generate some imprecision in the recovered subjective probability distributions. This additional uncertainty is conventionally handled by undertaking a bootstrapping analysis of the recovery of subjective probability distributions. This approach is, as one can imagine, somewhat more involved numerically than the first three approaches, but allows one to say that true, latent subjective probability distributions have been employed, rather than making strong assumptions explicit in the first three approaches.

We see these approaches as providing a natural sequence with which to analyze the observed data. The first approach allows one to directly treat the observed data as subjective belief distributions from which we can ascertain measures of literacy, but at the price of making a very strong theoretical assumption. The second and third approaches allow one to relax that assumption to varying degrees, but still means that some observations are either thrown out or given relatively low weights compared to other observations. The fourth approach allows one to avoid assuming that the subject is behaving consistently with SEU, but at the price of requiring more numerical analysis of the raw results in order to recover the true latent subjective probability distributions.

2.3 Experimental Tasks

The experimental results reported here were collected as part of several unrelated projects. One was a project conducted for the Federal Reserve Bank of Atlanta designed to understand differences between survey measures of future inflation and incentivized belief distributions of future inflation. Another was a project designed to evaluate expected welfare gains to individuals from purchases of insurance, documented in Harrison and Ng (2016).

Of relevance for the results reported here, each subject completed a battery of binary choices over risky lotteries, a battery of token allocations for beliefs, and a standard socio-demographic survey. The order in which these were administered varied from session to session. Section 2.2 addresses the theoretical framework of the risky lotteries and the subjective belief tasks, and using characteristics of individuals derived from the standard socio-demographic survey allows the modeling of some of the observed heterogeneity in the sample.

2.4 Experiments and Demographics

The experiments in the United States, reported in chapters 2, 3 and 4, were conducted at Georgia State University's (GSU) Experimental Economics Center (ExCEN)[21] over the period from November 2013 to March 2016 and were often run in combination with other research projects. The lab can accommodate up to 40 subjects in a session. In total 31 sessions using the financial literacy questions were administered to current GSU students that registered interest on the recruiter over that time frame; note session 24 isn't applicable. Once a student participated in any of the 31 sessions, they were excluded from participating in subsequent sessions for this project. A copy of the email invitation via an online recruiter is in Appendix B.

Table 2.2 displays the number of sessions by treatments and the bin labeling scheme used. Over the 31 sessions, 20 sessions used "Initial" interval labels and 11 used "New" interval labels. This is a naming convention that was adopted to test if there is an anchoring effect or mid-point biases around the initial bin labels compared to new bin labels when shifting the interval labels around the bins, thus shifting the bin location containing the true answer, Appendix A provides more detail on these robustness checks. The second column labeled "Individual" is defined as a participant only having access to their own cognition with which to respond to the financial literacy questions being asked. The participants in these control sessions were allowed to use a piece of blank paper and a pencil if they wanted to do calculations by hand, but no other outside aides were allowed (e.g. calculator, phone, or internet). There were 16 Individual sessions using the initial bin label and 3 using the new labels, thus 19 sessions total. The third column shows the number of sessions that were allowed timed access to the Internet just prior to answering the financial literacy questions. There are 5 Internet sessions in total; 2 sessions using the initial labels and 3 sessions using the new bin labels. The Internet sessions are discussed fully in Chapter 3. The fourth column shows the number of sessions assigned to the Group

[21] Every Fall semester graduate students and staff affiliated with ExCEN canvas classrooms across Georgia State University to inform students about the opportunity to participate in paid research at the Center. Students interested in the opportunity are instructed to go to ExCEN's website - http://excen.gsu.edu/center/ - and register for the online recruiter. After a student is registered they are then eligible to be invited to participate in research projects that take place in ExCEN. Only current students are eligible to participate.

treatment. In the Group treatment students were randomly assigned to exogenously formed groups comprised of two individuals and required to make a joint decision when answering the literacy questions. There are 7 total Group sessions; 2 sessions using the initial labels and 5 sessions using new labels. Beliefs as part of Groups will be the focus of Chapter 4.

Table 2.2 - Session Count by Treatment and Bin Labels

Label	Individual	Internet	Group	Session Count
Initial	16	2	2	20
New	3	3	5	11
Total	19	5	7	31

The proceedings of a session were as follows: a) students were signed into the experiment using their GSU identification card and randomly assigned to one of the 40 seats in the lab; b) students were presented with an introductory text that outlined their time in the lab; c) they would complete a demographic survey, participate in stylized economic experiments; d) paid a flat participation fee and any additional earnings generated from their decisions in the experiments. The complete protocol and instruction sets is documented in Appendix C. Table 2.3 provides an expanded view by session, overall participation, treatment assigned, and the number of responses for each question asked in a given session.

Table 2.3 - Session by Total Participation, Bin Label, Treatment, and by FinQ

Session	Participants	Bin Label	Individual	Internet	Groups	fin1	fin2	fin3	fin4	fin5	fin6	fin7	fin8	fin9	fin10	fin11	fin12	fin13	fin14	fin15	fin16
S01	40	Initial	Yes			14	13	14	13	14	13	40	13	13	14	13	13	14	13	13	13
S02	40	Initial	Yes			13	14	13	13	13	14	40	13	14	13	14	14	13	13	13	13
S03	40	Initial	Yes			13	13	13	14	13	13	40	14	13	13	13	13	13	14	14	14
S04	40	Initial	Yes			14	13	14	13	14	13	40	13	13	14	13	13	14	13	13	13
S05	40	Initial	Yes			14	13	14	13	14	13	40	13	13	14	13	13	14	13	13	13
S06	40	Initial	Yes			13	13	13	14	13	13	40	14	13	13	13	13	13	14	14	14
S07	40	Initial	Yes			13	13	13	14	13	13	40	14	13	13	13	13	13	14	14	14
S08	36	Initial	Yes			13	11	13	12	13	11	36	12	11	13	11	11	13	12	12	12
S09	3	Initial	Yes			1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1
S10	4	Initial	Yes			2	1	2	1	2	1	4	1	1	2	1	1	2	1	1	1
S11	3	Initial	Yes			1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1
S12	36	Initial	Yes			36		36		36		36			36			36			
S13	40	Initial	Yes				40				40	40		40		40	40				
S14	37	Initial	Yes				37				37	37		37		37	37				
S15	36	Initial	Yes			18		18	18	18		36	18		18			18	18	18	18
S16	34	Initial		Yes		34	34			34		34		34	34	34		34	34	34	34
S17	33	New		Yes		33	33			33		33		33	33	33		33	33	33	33
S18	33	Initial	Yes			33	33			33		33		33	33	33		33	33	33	33
S19	33	New	Yes			33	33			33		33		33	33	33		33	33	33	33
S20	32	New		Yes		32	32			32		32		32	32	32		32	32	32	32
S21	16	New			Yes	16	16			16		16		16	16	16		16	16	16	16
S22	22	New			Yes	22	22			22		22		22	22	22		22	22	22	22
S23	22	New			Yes	22	22			22		22		22	22	22		22	22	22	22
S24		N/A																			
S25	39	New		Yes		39	39			39		39		39	39	39		39	39	39	39
S26	28	New			Yes	28	28			28		28		28	28	28		28	28	28	28
S27	34	New			Yes	34	34			34		34		34	34	34		34	34	34	34
S28	34	New	Yes			34	34			34		34		34	34	34		34	34	34	34
S29	39	New	Yes			39	39			39		39		39	39	39		39	39	39	39
S30	32	Initial			Yes	32	32			32		32		32	32	32		32	32	32	32
S31	34	Initial			Yes	34	34			34		34		34	34	34		34	34	34	34
S32	36	Initial		Yes		36	36			36		36		36	36	36		36	36	36	36
Total	976		19	5	7	666	684	165	127	666	183	976	127	684	666	684	183	666	628	628	628

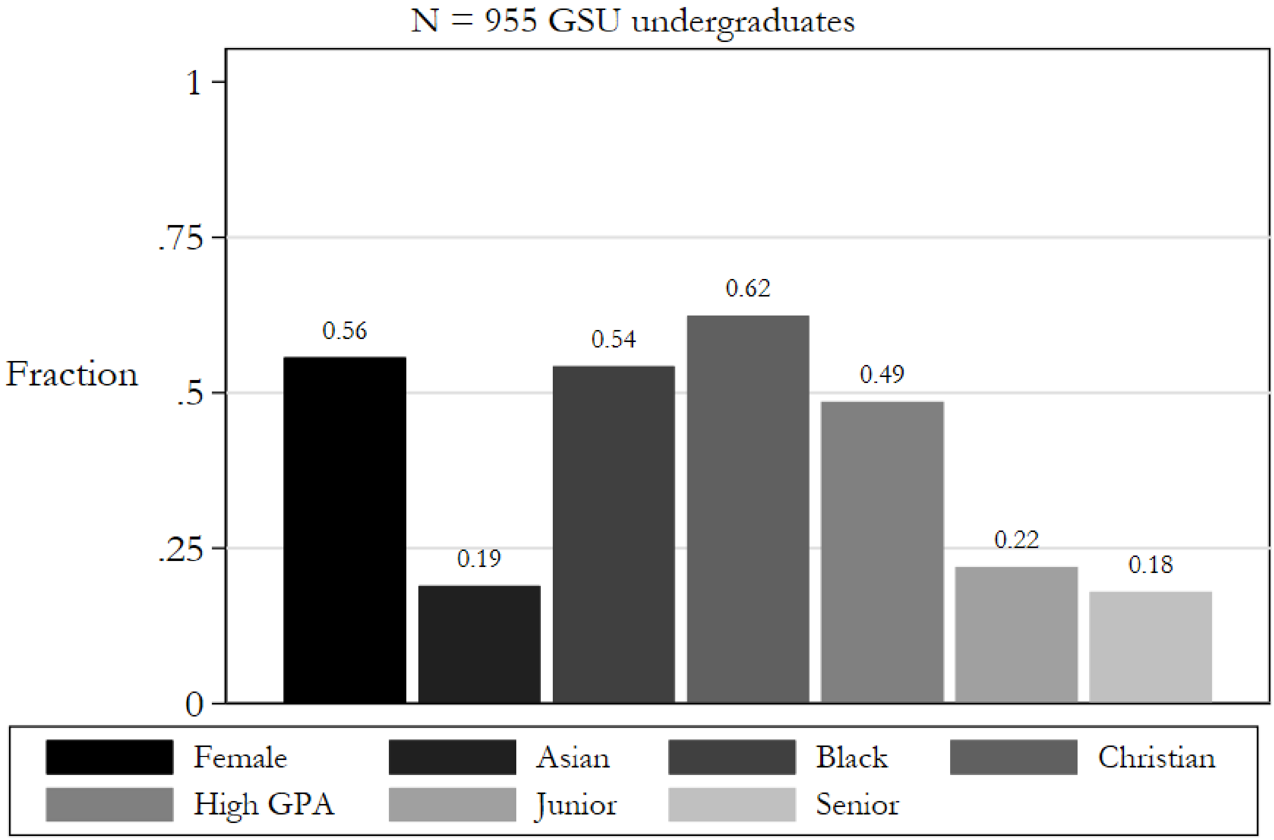
For purposes of this chapter we first start by comparing the responses to the 11 belief questions presented to individuals under both labeling schemes, initial and new, to test whether the label shift had an anchoring effect. For Chapters 3 and 4 we can consider pooling responses across all labels using all Individuals as a control group to compare results with Internet and the Group treatments, respectively, if the effect of the labeling scheme is not statistically significant. Table 2.4 shows the total number of participants in the Individual sessions that responded for each financial literacy question by bin label. For *fin9*, for example, we see that a total of 322 participants responded to that question, 106 were presented the question that used the new labels and 216 with the initial labels.

Table 2.4 - Number of Participants by Question by Bin Label, Individuals

Question	Bin Labels		Total
	Initial	New	
fin1	198	106	304
fin2	216	106	322
fin3	165		165
fin4	127		127
fin5	198	106	304
fin6	183		183
fin7	508	106	614
fin8	127		127
fin9	216	106	322
fin10	198	106	304
fin11	216	106	322
fin12	183		183
fin13	198	106	304
fin14	160	106	266
fin15	160	106	266
fin16	160	106	266
Total	3213	1166	4379

The demographic covariates for the GSU sample are generally self-explanatory, and their averages are displayed in Figure 2.3. Variable “High GPA” denotes subjects reporting a cumulative GPA between 3.75 and 4, which translates into “mostly A’s.” We have more women than men, relatively large shares of Asians and Blacks, and a large fraction of Christians.

Figure 2.3: Average of GSU Sample Demographics, All Sessions



The experiments in Denmark, reported in chapter 5, involved subjects recruited for a field experiment. The recruitment procedures, logistics, and demographics are explained in chapter 5.

2.5 Econometric Specification

The most natural econometric specification for the beliefs data is Interval Regression. This econometric model is essentially the same as Ordinary Least Squares (OLS) regression, except that it allows that data may be in intervals. That is, we may know the strength of someone's belief about how many years a man will live, conditional on being 20, is between 40 and 49, but not know more precisely than that. We might also observe the strength of their belief is 40 years or more, but not know more than that. When the upper and lower bound of the interval are the same, the data collapse to "point data" that is what one normally applies OLS to.

One limitation of Interval Regression, as typically implemented, is that it assumes that the latent distribution being estimated is Gaussian. When beliefs are recorded in terms of probabilities, as we do in some cases, this can be unattractive: the Gaussian distribution implies that the estimated values in this instance could be less than 0 or greater than 1, which is inconsistent with them being a probability that is bounded between 0 and 1. One way to address this issue is to consider an extension of Interval Regression to allow for a latent distribution that is bounded between 0 and 1, and the most flexible of those is the Beta distribution. It is possible to extend the logic of Interval Regression to allow for a Beta distribution, as documented in a lecture by Harrison (2013), reproduced here with permission as Appendix D. These notes also show how one can extend the specification to include a flexible non-Gaussian distribution that is not bounded between 0 and 1, the Gamma distribution.

For most purposes it is acceptable to use the typical Interval Regression specification, which is easy to understand and interpret. As needed, estimates will be provided with the Beta or Gamma Interval Regression model and noted as such.

We estimate EUT and RDU models for each individual, following procedures explained in Harrison and Rutström (2008) and formal econometric models specified in Harrison and Ng (2016), Harrison and Ross (2018) and Harrison and Ulm (2016). We consider the Constant Relative Risk Aversion (CRRA) utility function $U(x) = x^{(1-r)}/(1-r)$ where x is the lottery prize and $r \neq 1$ is a parameter to be estimated. For $r=1$ assume $U(x)=\ln(x)$ if needed. Thus r is the coefficient of CRRA for an EUT individual: $r=0$ corresponds to risk neutrality, $r<0$ to risk loving, and $r>0$ to risk aversion. For the RDU models we consider this CRRA utility function and one of three possible probability weighting functions. The first is the simple "power" probability weighting function proposed by Quiggin (1982), with curvature parameter γ: $\omega(p) = p^\gamma$. [22] The second probability weighting function is the "inverse-S" function popularized by Tversky and Kahneman

[22] So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation. Convexity of the probability weighting function is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \;\; \forall p$ and hence the "RDU EV" weighted by $\omega(p)$ instead of p has to be less than the EV weighted by p.

(1992): $\omega(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma}$. [23] The third probability weighting function is a general functional form proposed by Prelec (1998) that exhibits considerable flexibility. This function is $\omega(p) = \exp\{-\eta(-\ln p)^\varphi\}$, and is defined for $0<p\leq1$, $\eta>0$ and $\varphi>0$.[24] For our purposes, it does not matter which of these probability weighting functions characterize behavior: the only issue is at what statistical confidence level we can (or cannot) reject the EUT hypothesis that $\omega(p) = p$.

If the sole metric for deciding a subject were better characterized by EUT or RDU was the log-likelihood of the estimated model, then there would be no subjects classified as EUT since RDU nests EUT. But if we use metrics of a 10%, 5% or 1% significance level on these test of the EUT hypothesis that $\omega(p) = p$, then we can classify subjects as being RDU subjects if there is statistically significant evidence that they behave as if weighting probability. The null hypothesis is that subjects are better characterized by EUT, and they will be assumed to be better characterized by RDU only if they show statistically significant to reject that null hypothesis. As a general matter the population from which our samples are drawn end up being classified as roughly 50% EUT, and the rest as one or other of the RDU models, with the most popular RDU model being the one with the Prelec probability weighting function (see Harrison and Ng (2016) and Harrison and Ulm (2016)).

Further, the recovery of latent subjective beliefs is undertaken at the level of each individual subject, and as such requires estimates of a model of risk preferences for each individual.[25] In general this is not an issue, since one of the four specifications considered normally solves with *a priori* sensible parameter values for the vast majority of subjects. When estimating such models for a large number of subjects, however, it is

[23] This function exhibits inverse-S probability weighting (optimism for small p, and pessimism for large p) for $\gamma<1$, and S-shaped probability weighting (pessimism for small p, and optimism for large p) for $\gamma>1$.
[24] When $\varphi=1$ this function collapses to the Power function $\omega(p) = p^\eta$. Many apply the Prelec (1998; Proposition 1, part (B)) function with constraint $0 < \varphi < 1$, which requires that the probability weighting function exhibit subproportionality. Contrary to received wisdom, many individuals exhibit estimated probability weighting functions that violate subproportionality, so we use the more general specification from Prelec (1998; Proposition 1, part (C)), only requiring $\varphi > 0$, and let the evidence for an individual determine if the estimates φ lies in the unit interval.
[25] To restate, interval regressions are utilized for the evaluation of the final results throughout the thesis. However, the calibration and recovery of latent subjective beliefs is done utilizing techniques developed by Harrison and Ng (2016) and Harrison and Ulm (2016), which are the most current techniques available at this time.

always possible that some subjects have no such solved model. In that case we adopt an approach with an informally Bayesian flavor. We estimate a pooled RDU model using the Prelec probability weighting function, and use it to *predict parameter estimates for each individual.* This is the most flexible of the four specifications, of course. This model includes a vector of demographic characteristics for each of the parameters of the model of risk preferences, so the predictions reflect the actual values of those characteristics for each subject. Hence the predictions for each subject can be, and are, very different, and can include EUT as a special case if that is the case. We are also able to predict a covariance matrix for the subject, so we can then undertake bootstrap evaluations of the recovered beliefs for that subject. These predicted risk preferences are only used when there is no model for an individual, using the choices of only that individual, that converges with sensible parameter values; the default is to use the estimates based only on the observed choices of that subject. In this GSU sample we use these predicted risk preferences for 84 subjects out of 955, or 8.8%.

2.5.1 How to Interpret the Interval Regression Output of *Stata*

It is worthwhile to carefully go through the regression output generated from *Stata's* **intreg** command using one of the financial literacy questions to articulate and interpret the output that will form the basis for the remainder of our analysis. Here we illustrate an example using the recovered latent subjective beliefs for the pool of control participants[26] for *fin9*, "the budgeting question," with wording as follows:

> *David just found a job with a take-home pay of $2,000 per month. He must pay $900 for rent and $150 for groceries each month. He also spends $250 per month on transportation. If he budgets $100 each month for clothing, $200 for restaurants, and $250 for everything else, how long will it take him to accumulate savings of $600?*

[26] Recall that our control participants are individuals without access to the internet or a group and using the initial labeling scheme.

The bin intervals are exact values which were labeled as: "1 month", "2 months", "3 months", "4 months", "5 months", "6 months", "7 months", "8 months", "9 months", and "10 months" for bins 1 through 10, respectively. For this question and response interval, the true answer is "4 months", which lies in bin 4.

The block below contains the code used and the subsequent estimation in *Stata's* **intreg** for a constant-only model using *fin9*. It is worthwhile being explicit about how one interprets output of this kind, to avoid excessive abstraction in explaining the way the data are evaluated. Immediately following the model is a demonstration on how to recover sigma. Note the highlighted areas end with superscripted indicators that will be used below to further elaborate on what the estimation output is telling us.

```
1 . * demonstrate translation from "lnsigma" to "sigma"
2 . intreg v_lo v_hi[a] [fweight = choiceI][b] if qid=="fin9" & Controls==1, cluster(id)[c]

  Fitting constant-only model[d]:

  Iteration 0:   log pseudolikelihood = -35199.053
  Iteration 1:   log pseudolikelihood = -35197.572
  Iteration 2:   log pseudolikelihood = -35197.572

  Fitting full model[e]:

  Iteration 0:   log pseudolikelihood = -35197.572
  Iteration 1:   log pseudolikelihood = -35197.572

  Interval regression                               Number of obs[g]     =     20,865
                                                       Uncensored      =          0
                                                       Left-censored   =          0
                                                       Right-censored  =        370
                                                       Interval-cens.  =     20,495

                                                    Wald chi2(0)[h]      =          .
  Log pseudolikelihood = -35197.572[f]                Prob > chi2[i]       =          .

                                        (Std. Err. adjusted for 216 clusters in id)
  ------------------------------------------------------------------------------
               |               Robust
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
        _cons[j] |    4.15382   .0811236    51.20   0.000     3.994821    4.312819
  -------------+----------------------------------------------------------------
     /lnsigma[k] |   .2422078   .1051359     2.30   0.021     .0361452    .4482704
  -------------+----------------------------------------------------------------
        sigma[l] |   1.274059   .1339494                      1.036806    1.565602
  ------------------------------------------------------------------------------

3 . nlcom (sigma:

         exp([lnsigma]_cons))[m]

         sigma:  exp([lnsigma]_cons)

  ------------------------------------------------------------------------------
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
        sigma[n] |   1.274059   .1339494     9.51   0.000     1.011523    1.536595
  ------------------------------------------------------------------------------
*updated 6/3/2018 using recovered beliefs
```

Figure 2.4 - *Stata* Output for intreg Command, Constant-only Model Using *fin9*.

a) **vi_lo vi_hi** – The *latent* dependent variable in an interval regression are two numbers telling us the lower bound of an interval and the upper bound of an interval, v_lo and v_hi, respectively. Thus values 0.23 and 0.45 would imply an observation known to lie in the interval (0.23, 0.45). It is possible that intervals overlap, so that another interval might be (0.3, 0.5). There is no requirement that the intervals be the same length. Although specific to the syntax in *Stata*, -∞ is implied by a dot as the lower value, and +∞ is implied by a dot as the upper value. If we have an observation that is exact, as is the case for *fin9*, the upper and lower intervals are the same, and we can mix "exact" and "interval" data freely. In the case of the beliefs data we have non-overlapping intervals by design, and often have intervals including -∞ or +∞.

   i. Note that we stressed the term latent: using the language of most statistical packages, there are two *variables* giving us the lower and upper bounds of each interval, but we are estimating a model of a single *latent variable*.

   ii. Just prior to the table output we are told the types of observations inferred for use in the model. In this case, out of the 216 subjects we have no uncensored observations, no left-censored observations, 370 right-censored observations, and 20,495 interval observations. The word "censored" just means that the lower value is -∞ (left-censored) or the upper value is +∞ (right-censored).

b) **[fweight = choiceI]** – The way the interval regression responses are recorded rests critically on what I call "frequency weights." These are weights on each observation to indicate how many times it should be counted: for example, in a survey of the population it is possible that we know that there are 100 men and 100 women in the population, but the survey has 10 men and 5 women. In this case we might place a weight of 10 on each man in the survey and a weight of 20 on each woman in the survey, so that when the survey observations are weighted they match the population. In this instance, we use the 100 tokens that each subject has in order to apportion weights to each interval. Recall the belief elicitation task gives each subject 100 tokens to apportion across each of the 10 intervals, thus these frequency weights sum to 100 for each subject and each belief question.

c) **cluster(id)** – Given that we have multiple observations from each subject, in fact 100 observations corresponding to each of the 100 tokens that the subject allocates, it is important to recognize when we estimate the model over many individuals that the responses of a given individual will be correlated. This is implemented by the use of "clustering," which is a popular method of allowing for this type of correlation. There are many other ways to allow for this type of correlation, such as "fixed effects" or "random effects" panel estimation.

d) **Fitting constant-only model** – This is the iteration history for fitting the constant only model. This model does not include any predictors and is simply estimating the mean predicted value of the outcome variable. Because the observed values for the outcome variable are intervals, not exact values, the mean predicted value is not simply the mean of the observed values. Instead, the predicted mean is arrived at iteratively by maximizing the log likelihood of the data given a mean predicted value.

e) **Fitting full model** – This is the iteration history for fitting the model including the specified predictors.

f) **Log psuedolikelihood** – This is the log likelihood of the fitted model. It is used in the Likelihood Ratio Chi-Square test of whether all predictors' regression coefficients in the model are simultaneously zero.

g) **Number of obs** – This is the number of observations in the dataset for which all of the predictor variables and at least one of the outcome interval variables is non-missing. In interval regression, one of the interval bounds may be missing. If the upper bound of an interval is missing, then the interval is treated as [lower bound, infinity). If the lower bound of an interval is missing, then the interval is treated as (negative infinity, upper bound]. If both the lower bound and upper bound are missing, then the observation is not included in the model. Also, see note under a) ii above.

h) **LR chi2(0)** – This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero. The number in the parentheses indicates the degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square statistic and is defined by the number of predictors in the model (0).[27]

i) **Prob > chi2** – This is the probability of getting a LR test statistic as extreme as, or more so, than the observed statistic under the null hypothesis; the null hypothesis is that all the regression coefficients are simultaneously equal to zero. The parameter of the chi-square distribution used to test the null hypothesis is defined by the degrees of freedom in the prior line, chi2(0).

j) **Coef.** – These are the regression coefficients. They are interpreted in the same manner as OLS regression coefficients: for a one-unit increase in the predictor variable, the expected value of the outcome variable changes by the regression coefficient, given the other predictor variables in the model are held constant. The first estimate to come from the interval regression model when we have no covariates, such as the gender of the subject, shows us the average value of the estimated distribution of beliefs over all subjects. This is estimated as the mean of the latent distribution of beliefs, and is shown here as 4.154. We observe that the 95% confidence interval of this estimate of the mean of the pooled distribution of subjective beliefs is between 3.995 and 4.313. We can use this estimate of the mean to test the hypothesis that the average belief is the same as the true belief, which in this case is 4 months. We could undertake this hypothesis test explicitly, or simply observe that the true value lies between the 95% confidence intervals, although close to the lower bound.

k) **/lnsigma** – The second estimate to come from the interval regression model shows us the standard deviation of the belief distribution. In this case *Stata* estimates a transformation of the standard deviation, which is the logarithm of the standard deviation. It does this for numerical reasons.

l) **sigma** – Undertakes a nonlinear transformation of the estimate of /lnsigma to recover the estimate of the standard deviation itself. When there is no covariate in the model, as in this case, *Stata* displays this

[27] It is often the case that the estimated interval regression model will not show a value for the Wald statistic, and this is indicated in the case of *Stata* by a dot. All this means is that the numerical algorithm that was being used to find the maximum likelihood estimates was unable to complete an infinitesimal perturbation of the likelihood at the maximum likelihood values. This is often due to numerical issues that are unrelated to whether the estimation method has indeed found the maximum likelihood values. If the Wald statistic is missing, then the p-value associated with it will also be missing.

transformation. So we observe that the estimated standard deviation of the distribution is 1.274, with a 95% confidence interval between 1.037 and 1.566. This estimate of the standard deviation, 1.274, is mathematically $e^{0.2422078}$, which is to say that the estimate of the logarithm of the standard deviation, 0.2422078, is mathematically ln(1.274059).

m) **nlcom** – This mathematical expression demonstrates the translation to convert the logarithm of the standard deviation (from k above) into the recovered standard deviation.

n) **sigma** (recovered) – We then see that the estimate of the standard deviation is 1.274059, which is identical to the estimate automatically provided by *Stata*. There is a slight difference in the 95% confidence intervals that are calculated, but these are usually second-order differences. One reason for the difference is that the nonlinear combination that is explicitly calculated here is in fact an approximation, using what is known in statistics as the "Delta method" (Oehlert 1992). This method provides a way of approximating the correct standard errors or nonlinear transformations of estimated parameters, effectively taking a Taylor series approximation of the true nonlinear transformation. Here we show this transformation explicitly in this instance, even though *Stata* has provided the transformed estimate of the standard deviation of the distribution, because we will be using this explicit transformation extensively to recover detailed coefficients when we have covariates. We will illustrate this momentarily.

The next block of code below is like the one above, however, introduces one covariate to the **intreg** estimation; in this case, whether the participant was a female. This is known as the *total effect* of being female and we will go on to calculate (recover) the marginal effect of the covariate and its standard error, now that *Stata* doesn't automatically provide these. Let us now interpret those results, those of the recovered sigmas, and the standard error of the covariate's incremental effect. Since we are building on the results above, I again highlight the areas of interest which end with superscripted indicators to aid in the discussion that follows.

```
1 . * now do when Stata does not provide this, when there are covariates
2 . intreg v_lo v_hi female° [fweight = choiceI] if qid=="fin9" & Controls==1, cluster(id
> ) het(female)p

  Fitting full model:

  Iteration 0:   log pseudolikelihood = -227411.08
  Iteration 1:   log pseudolikelihood = -36784.947
  Iteration 2:   log pseudolikelihood = -35151.394
  Iteration 3:   log pseudolikelihood =  -35116.73
  Iteration 4:   log pseudolikelihood = -35116.592
  Iteration 5:   log pseudolikelihood = -35116.592

  Interval regression                             Number of obs      =     20,865
                                                     Uncensored      =          0
                                                     Left-censored   =          0
                                                     Right-censored  =        370
                                                     Interval-cens.  =     20,495

                                                  Wald chi2(1)       =       1.21
  Log pseudolikelihood = -35116.592               Prob > chi2        =     0.2710

                                      (Std. Err. adjusted for 216 clusters in id)
  ------------------------------------------------------------------------------
               |               Robust
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
  model        |                                                                   q
        female |  -.1785286   .1621955    -1.10   0.271     -.496426    .1393688
         _cons |   4.262164   .1216165    35.05   0.000       4.0238    4.500528
  -------------+----------------------------------------------------------------
  lnsigma      |
        female |   .0883017   .2348058     0.38   0.707    -.3719092    .5485127
         _cons |   .1846372   .2034445     0.91   0.364    -.2141067    .5833811
  ------------------------------------------------------------------------------

4 . * total effect
5 . nlcom (sigma_cons: exp([lnsigma]_cons)) (sigma_female: exp([lnsigma]_cons + [lnsigma
> ]female))

    sigma_cons:  exp([lnsigma]_cons)
  sigma_female:  exp([lnsigma]_cons + [lnsigma]female)

  ------------------------------------------------------------------------------
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
    sigma_cons |   1.202782   .2446994     4.92   0.000       .72318    1.682384    r
  sigma_female |    1.31382   .1540261     8.53   0.000     1.011934    1.615706
  ------------------------------------------------------------------------------

7 . * marginal effect of covariate

8 . nlcom (sigma_female: exp([lnsigma]_cons + [lnsigma]female) -exp([lnsigma]_cons))

  sigma_female:  exp([lnsigma]_cons + [lnsigma]female) - exp([lnsigma]_cons)

  ------------------------------------------------------------------------------
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
  sigma_female |    .111038   .2891398     0.38   0.701    -.4556655    .6777416    s
  ------------------------------------------------------------------------------

*updated 6/3/2018 using recovered beliefs
```

Figure 2.5 - *Stata* Output for intreg Command, One Covariate Model Using *fin9*.

o) **Female** – In this specification we have included one covariate, a binary indicator that the subject was a female. We include this covariate on the estimate of the mean of the distribution, as well as on the estimate of the standard deviation of the distribution.

p) **het(female)** – "het(.)", asks *Stata* to estimate a model with what is called multiplicative heteroscedasticity. This means that *Stata* allows the estimate of the standard deviation of the normal distribution to vary in a linear fashion with the covariate.

q) The effect on the mean is given by the coefficients, and the effect on the standard deviation is given by the variable in the option "het(.)" We focus initially on the interpretation of the estimates for the mean of the distribution, and come below to the interpretation of the estimates for the standard deviation of the distribution. Since the binary variable female is equal to one if the subject is a female, and equal to zero if the subject is a male, we can infer that the constant term in this estimated model refers to the average of the distribution of beliefs for males: in this case, the average is 4.262 for males. The model tells us that the average for females is 0.179 *smaller* than the average for males, so the average for females is 4.262164 - 0.1785286 = 4.0836354. We observe that the 95% confidence interval for males between 4.024 and 4.500, and we can see that the 95% confidence interval for females, taking care to evaluate with respect to the constant, is between 3.527 (= 4.024 - 0.496) and 4.640 (= 4.500 + 0.139). Further, *Stata's* output provides us the Confidence Interval (CI) for an individual coefficient taking into account that there are other predictors in the model. For a given predictor with a level of 95% confidence, we say that we are 95% confident that the "true" coefficient lies between the lower and upper limit of the interval.

r) We turn now to the effect of the covariate on the estimated standard deviation of the distribution, and interpretation of the constant term and the female term. Note that we are now doing two calculations with the "Delta method," to infer the estimate of the standard deviation for men, 1.203, and the estimate of the standard deviation for women, 1.314. In each case we also have the correct estimates of the standard error and the 95% confidence intervals. In this case the estimate of the standard deviation for men is exactly the same mathematical expression as the previous specification that had no covariates, although with different numerical values: $e^{0.1846372} = 1.203$. But to calculate the standard deviation for females we need to recognize that the mathematical expression becomes $e^{(0.1846 + 0.0883)} = e^{0.2729} = 1.314$.

Notice that the mathematical expression is essentially the same, but the exponent is the linear combination of the coefficient for men (0.1846) and the coefficient for the incremental effect of being female (+0.0883), which is the lnsigma displayed in **intreg** estimation results.

s) We could use the same logic to calculate the incremental effect of being female on the estimated standard deviation of the distribution. This is given by the total effect of being female on the estimated standard deviation of the distribution, which is the same thing as the standard deviation of the distribution for females, minus the standard deviation of the distribution for males. In effect, this is the difference between the two standard deviations calculated in (r) above: 0.111 = 1.319 - 1.203. But we cannot calculate the standard error of this incremental effect in the same way that we can calculate the mean effect itself. We must recognize that the incremental effect of being female is not just a nonlinear transformation of the coefficients directly estimated by *Stata*, but the difference between these two nonlinear transformations. And this requires us to again apply the Delta method, resulting in an estimated standard error of the incremental effect of being female of 0.2891.

It is useful to pause here for a moment to compare the two estimates of the standard deviation for females. One is called the "total effect" of being female on the estimated standard deviation, and is 1.314, shown in (r) above. The total effect is different from the "marginal effect" of being female relative to the estimated standard deviation for males, which is 0.111, shown in (s) above. The terminology "total effect" is a bit awkward in this context, since we can more conveniently refer to the standard deviation of the belief distribution for women, but it is extremely important to differentiate that from the marginal, or incremental, effect of being a woman relative to the total effect of being a man. In this simple setting in which there is just one covariate, these distinctions might seem to be excessive, but as we start to include several covariates they will become increasingly important to correctly interpret results.

Our final block of code below will go through an example when all covariates are introduced into the model estimation.

```
1 . * now for all covariates
2 . intreg v_lo v_hi $demog[t] [fweight = choiceI] if qid=="fin9" & Controls==1, cluster(id
  > ) het($demog)[u]

  Fitting full model:

  Iteration 0:   log pseudolikelihood = -255739.64
  Iteration 1:   log pseudolikelihood = -36696.927
  Iteration 2:   log pseudolikelihood = -33512.381
  Iteration 3:   log pseudolikelihood = -33431.722
  Iteration 4:   log pseudolikelihood = -33431.128
  Iteration 5:   log pseudolikelihood = -33431.128

  Interval regression                             Number of obs     =     20,865
                                                     Uncensored     =          0
                                                     Left-censored  =          0
                                                     Right-censored =        370
                                                     Interval-cens. =     20,495

                                                  Wald chi2(7)      =       4.44
  Log pseudolikelihood = -33431.128               Prob > chi2       =     0.7285

                                    (Std. Err. adjusted for 216 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
model        |                                                                 v
      female |  -.1947403   .1580315    -1.23   0.218    -.5044763    .1149957
       asian |  -.1100327   .1631922    -0.67   0.500    -.4298836    .2098182
       black |  -.1186321   .1705087    -0.70   0.487    -.4528229    .2155587
   christian |   .3144775    .186012     1.69   0.091    -.0500994    .6790544
       gpaHI |   .1177914    .130793     0.90   0.368    -.1385582    .3741409
      junior |   .0104117   .1455112     0.07   0.943     -.274785    .2956083
      senior |   .0793734     .21057     0.38   0.706    -.3333363     .492083
       _cons |   4.070898    .159489    25.52   0.000     3.758305    4.383491
-------------+----------------------------------------------------------------
lnsigma      |
      female |    .175989   .2186728     0.80   0.421    -.2526019    .6045799
       asian |   .1050275   .2926372     0.36   0.720    -.4685309     .678586
       black |    .299265   .2491636     1.20   0.230    -.1890867    .7876168
   christian |   .2115997   .2361546     0.90   0.370    -.2512547    .6744542
       gpaHI |  -.2116856   .1973843    -1.07   0.284    -.5985518    .1751805
      junior |  -.4085053   .2109104    -1.94   0.053     -.821882    .0048714
      senior |    -.266104   .2493592    -1.07   0.286    -.7548391    .2226311
       _cons |  -.0084433   .2964404    -0.03   0.977    -.5894559    .5725692
------------------------------------------------------------------------------

3 . foreach v of global demog {                                                 w
      nlcom (sigma_cons: exp([lnsigma]_cons)) (sigma_`v'_t: exp([lnsigma]_cons
      > + [lnsigma]`v')) (sigma_`v'_m: exp([lnsigma]_cons + [lnsigma]`v') -
      > exp([lnsigma]_cons))
    }

    sigma_cons:  exp([lnsigma]_cons)
  sigma_fema~t:  exp([lnsigma]_cons + [lnsigma]female)                           x
  sigma_fema~m:  exp([lnsigma]_cons + [lnsigma]female) - exp([lnsigma]_cons)

  ------------------------------------------------------------------------------
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
    sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
  sigma_female_t |   1.182399   .3469704     3.41   0.001     .5023498    1.862449
  sigma_female_m |   .1908071   .2426216     0.79   0.432    -.2847225    .6663367
  ------------------------------------------------------------------------------
```

CONTINUED ON NEXT PAGE

```
  sigma_cons:  exp([lnsigma]_cons)
sigma_asia~t:  exp([lnsigma]_cons + [lnsigma]asian)
sigma_asia~m:  exp([lnsigma]_cons + [lnsigma]asian) - exp([lnsigma]_cons)

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_asian_t|   1.101402   .2508395     4.39   0.000     .6097659    1.593039
sigma_asian_m|   .1098101   .3003861     0.37   0.715    -.4789359    .6985561
------------------------------------------------------------------------------

  sigma_cons:  exp([lnsigma]_cons)
sigma_blac~t:  exp([lnsigma]_cons + [lnsigma]black)
sigma_blac~m:  exp([lnsigma]_cons + [lnsigma]black) - exp([lnsigma]_cons)

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_black_t|   1.337526   .3692232     3.62   0.000     .6138618     2.06119
sigma_black_m|   .3459338   .2957258     1.17   0.242     -.233678    .9255457
------------------------------------------------------------------------------

  sigma_cons:  exp([lnsigma]_cons)
sigma_chri~t:  exp([lnsigma]_cons + [lnsigma]christian)
sigma_chri~m:  exp([lnsigma]_cons + [lnsigma]christian) - exp([lnsigma]_cons)

----------------------------------------------------------------------------------
                 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+----------------------------------------------------------------
      sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_christian_t|   1.225264   .3762609     3.26   0.001     .4878063    1.962722
sigma_christian_m|   .2336719   .2727544     0.86   0.392    -.3009168    .7682606
----------------------------------------------------------------------------------

  sigma_cons:  exp([lnsigma]_cons)
sigma_gpaH~t:  exp([lnsigma]_cons + [lnsigma]gpaHI)
sigma_gpaH~m:  exp([lnsigma]_cons + [lnsigma]gpaHI) - exp([lnsigma]_cons)

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_gpaHI_t|   .8024153   .2643792     3.04   0.002     .2842416    1.320589
sigma_gpaHI_m|  -.1891769   .1760838    -1.07   0.283    -.5342948    .1559409
------------------------------------------------------------------------------

  sigma_cons:  exp([lnsigma]_cons)
sigma_juni~t:  exp([lnsigma]_cons + [lnsigma]junior)
sigma_juni~m:  exp([lnsigma]_cons + [lnsigma]junior) - exp([lnsigma]_cons)

-------------------------------------------------------------------------------
              |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
   sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_junior_t|   .6590548   .2211891     2.98   0.003     .2255321    1.092577
sigma_junior_m|  -.3325375   .1826507    -1.82   0.069    -.6905262    .0254513
-------------------------------------------------------------------------------

  sigma_cons:  exp([lnsigma]_cons)
sigma_seni~t:  exp([lnsigma]_cons + [lnsigma]senior)
sigma_seni~m:  exp([lnsigma]_cons + [lnsigma]senior) - exp([lnsigma]_cons)

-------------------------------------------------------------------------------
              |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
   sigma_cons |   .9915922    .293948     3.37   0.001     .4154647     1.56772
sigma_senior_t|   .7599161    .253219     3.00   0.003     .2636159    1.256216
sigma_senior_m|  -.2316762   .2179309    -1.06   0.288    -.6588128    .1954605
-------------------------------------------------------------------------------

*updated 6/3/2018 using recovered beliefs
```

Figure 2.6 - *Stata* Output for intreg Command, Full Covariate Model Using *fin9*.

t) **$demog** – In this specification we have included the full complement of covariates, which were defined in Section 2.4. All covariates were included for the estimate of the mean of the distribution, as well as for the estimate of the standard deviation of the distribution.

u) **het($demog)** – again "het(.)" asks *Stata* to estimate a model with multiplicative heteroscedasticity. In the full demographic model, this means that *Stata* allows the estimate of the logarithm of the standard deviation of the normal distribution to vary in a linear fashion with each covariate.

v) The effect on the mean is given by the coefficients, and the effect on the logarithm of the standard deviation is given by the variable in the option "het(.)". For the covariates of the mean, we can interpret the estimate in the same manner as estimates of OLS regression coefficients: for a one unit increase in the predictor variable, the expected value of the outcome variable changes by the regression coefficient, given the other predictor variables in the model are held constant. By setting all binary covariates in the model equal to zero one can evaluate the model constant, which is the omitted variable that the other covariates are compared to. For this model the constant term is interpreted as White males in either their freshman or sophomore year of college (underclassmen) [28], with a mean report of 4.071 and a 95% CI of 3.758 to 4.383. Let us again focus on the covariate "female" to narrate this example now that the model is accounting for a richer set of demographics. By setting the binary variable female equal to one the model tells us that the average for females is 0.195 *smaller* than the average for White male underclassmen, so the average for females is 4.070898 - 0.1947403 = 3.8761577. We can calculate the 95% CI for females, again taking care to evaluate with respect to the constant, and it is between 3.254 (= 3.758 - 0.504) and 4.498 (= 4.383 + 0.115). The remainder of the covariates can be evaluated in the

[28] Underclassman is a term used colloquially in the United States to refer to a freshman or sophomore student in a high school or university setting. The *Oxford English Dictionary* defines upperclassman, as a special use noun in the United States, as a junior or senior student in high school or college. The distinction was made to be able to identify those that generally haven't started course work in their major area of study.

same manner. It should be noted here that the estimates provided in this section (v) will also be interpreted later as *bias*.

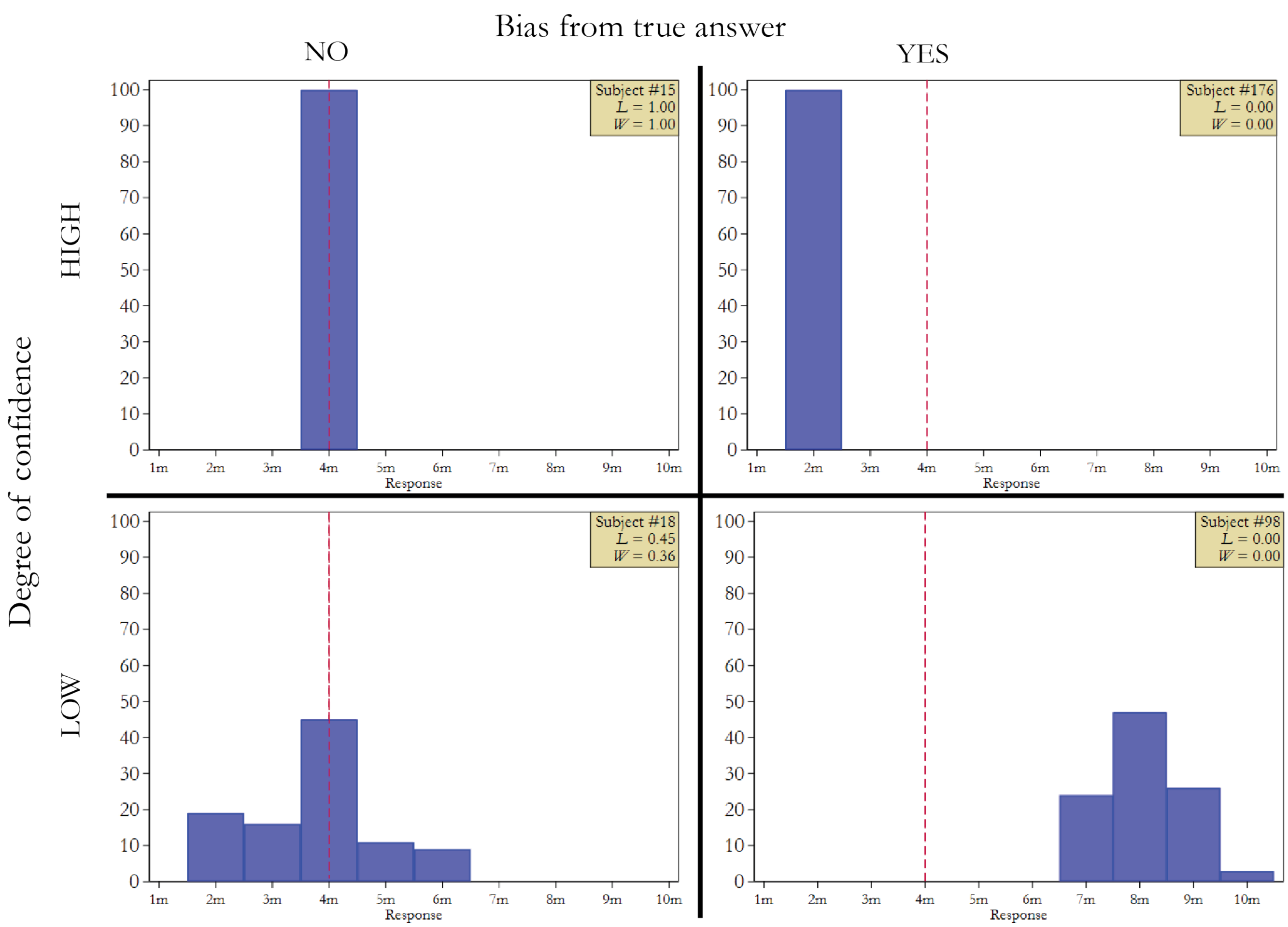
w) Is *Stata* code that loops over all the covariates of the logarithm of the standard deviation ("lnsigma") in the model to recover the estimate of the standard deviation for the constant, and the total effect and marginal effect for each covariate.

x) This section of results should be interpreted comparably to sections (r) and (s) earlier. Here the "sigma_cons" term is the estimate of the standard deviation for White male underclassmen, 0.992. The "sigma_female_t" is the overall effect of being female on the estimated standard deviation of the distribution, 1.182: this estimate accounts for the *total effect* of being female after taking into account all of the covariate values of females in the sample. Here we can read off the *p*-value on this statistic as 0.001, thus only looking the *total effect* of being a female on the standard deviation of the underlying belief distribution is statistically significant at the 0.1 percentage point significance level. On the other hand, "sigma_female_m" is the incremental effect of being female on the estimated standard deviation of the distribution, 0.191, and accounts for the *marginal effect* of being female when controlling for the other covariates in the model. The *p*-value on this statistic is 0.432, thus the *marginal effect* of being a female on the standard deviation of the underlying belief distribution is not statistically significant at any traditional significance levels. The estimates provided in this section (x) will also be interpreted later as *confidence*.

2.5.2 Characterizing Responses Using Bias and Confidence

Section 1.2.1 introduced concepts that allow us to build a framework for characterizing responses using the terms bias and degree of confidence in reference to some known true answer. Section 2.5.1 then presented the output generated from *Stata* using interval regression and detailed how to interpret those estimates with respect to bias and confidence. Figure 2.7 shows these concepts in practice by illustrating four participants' raw token allocations for the "budget" question fin9 in four panels that represent varying degrees of confidence and bias from the true answer. Recall that the true answer is 4 months, and this is displayed by the dashed red line in each panel. The two panels on the top of Figure 2.7 exhibit a high degree of confidence, and the two panels on the bottom exhibit a low degree of confidence. The two panels on the left exhibit no bias from the true answer, whereas the two panels on the right exhibit bias.

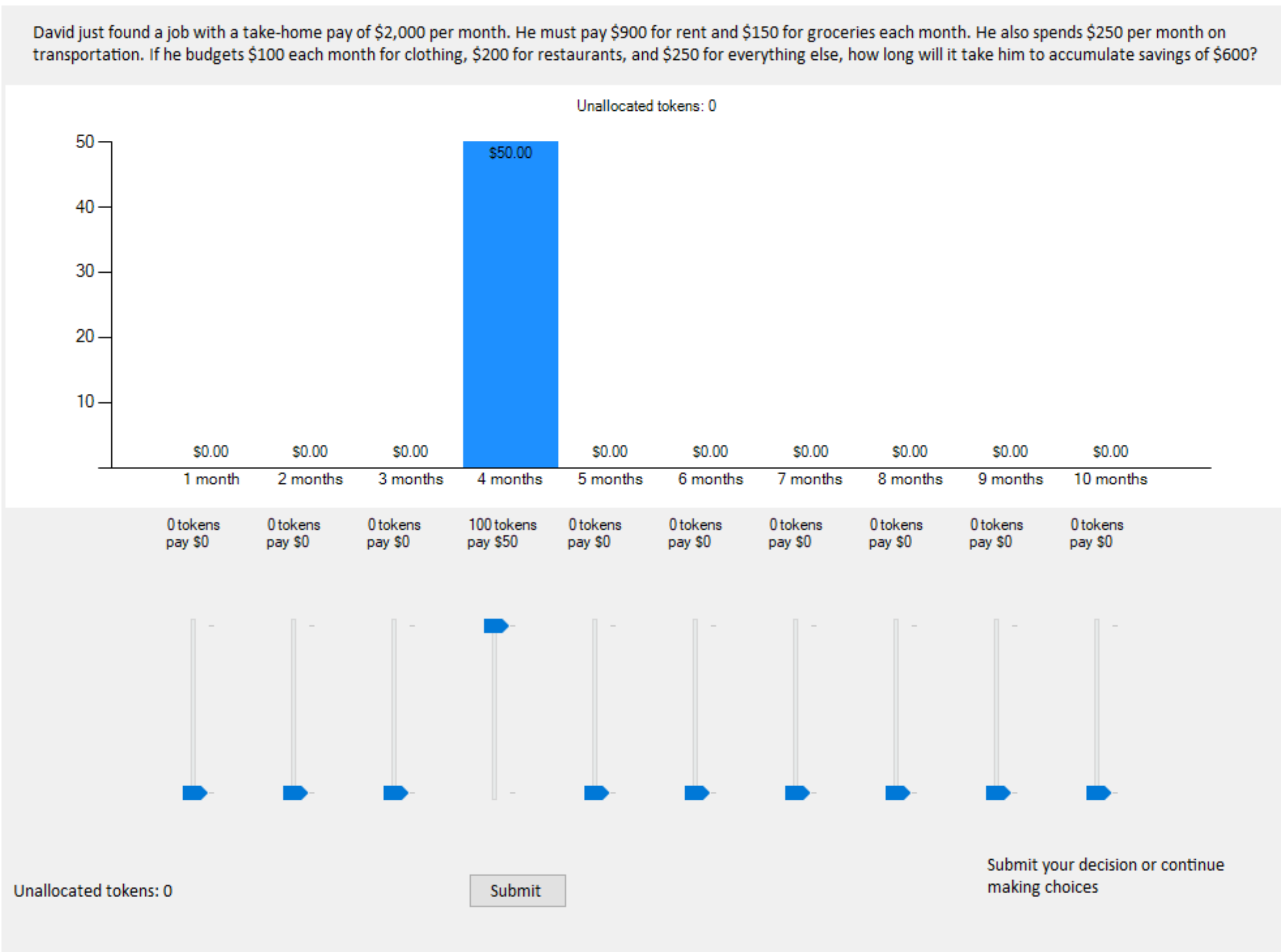
Figure 2.7 - Examples of Bias and Confidence shown with *fin9*



No Bias, High Confidence

The top-left panel of Figure 2.7 is the token allocation given by Subject #15. This subject exhibits no bias from the true answer and a high degree of confidence by allocating all 100 tokens to the bin that contained the true answer. This is the ideal scenario for a decision maker to be in and the one that generates the maximum earnings under the quadratic scoring rule for this question, which was $50. Figure 2.8 displays the earning outcomes for this participant. This participant has the capability to plan a budget and, given the income and expenditures presented in the question, knows it will take exactly 4 months to accumulate savings of $600. This is important when planning to save meaningful sums of money and knowing how long it will take to reach a targeted amount.

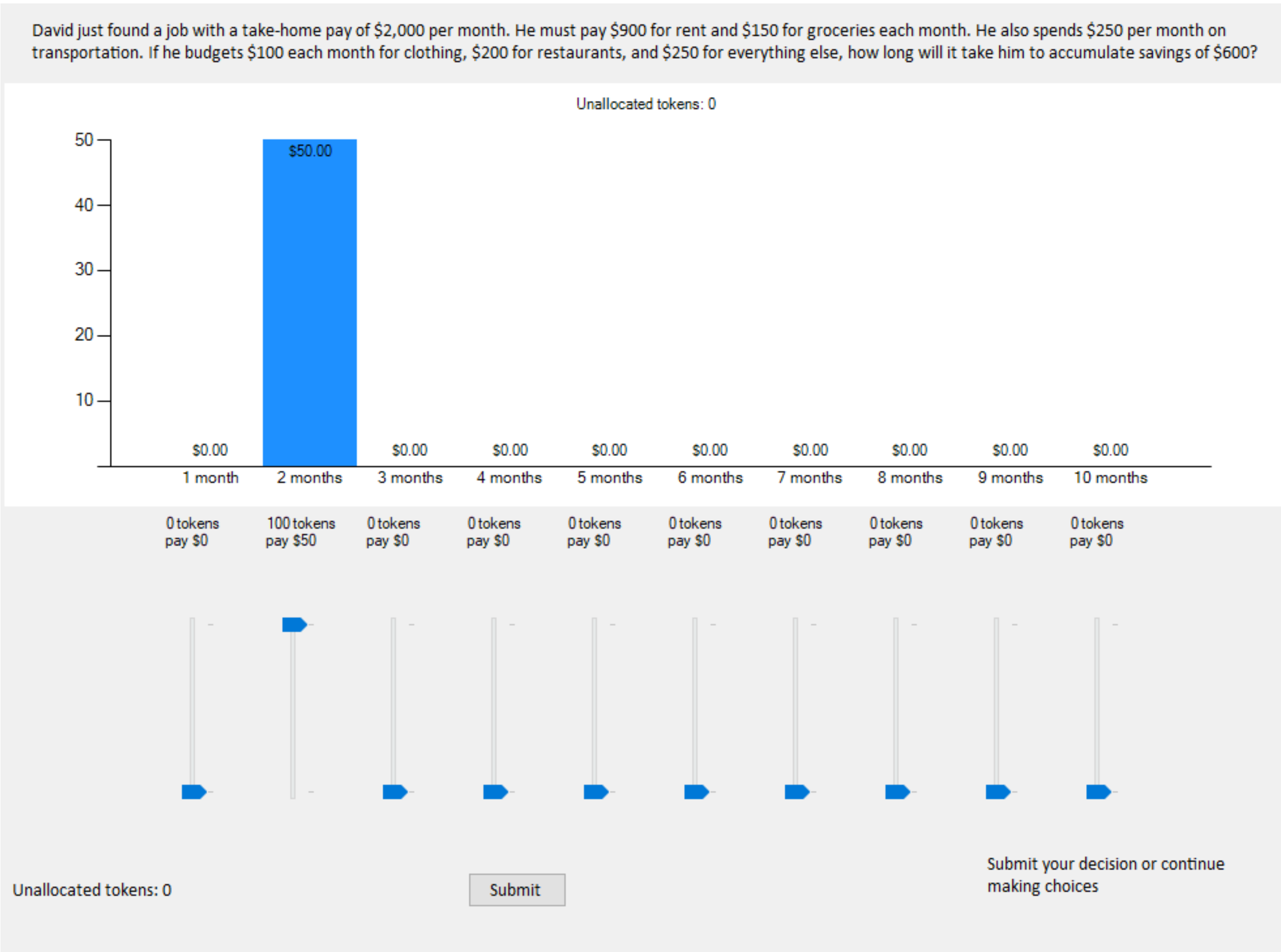
Figure 2.8 - Potential Earning Outcome with No Bias and High Confidence, *fin9*



Bias, High Confidence

The top-right panel of Figure 2.7 is the token allocation given by Subject #176. This subject exhibits bias away from the true answer and, unfortunately, a high degree of confidence by allocating all 100 tokens to a bin that did not contain the true answer. Here the participant was certain the true answer was 2 months to accumulate a savings of $600, when in fact it would take 4 months. This is one of the worst scenarios for a decision maker to be in and generates no earnings for this question. Figure 2.9 displays the earning outcomes for this participant. This participant does not seem capable at planning a budget and if given the task to do so they would systematically underestimate the time it takes to accumulate savings.

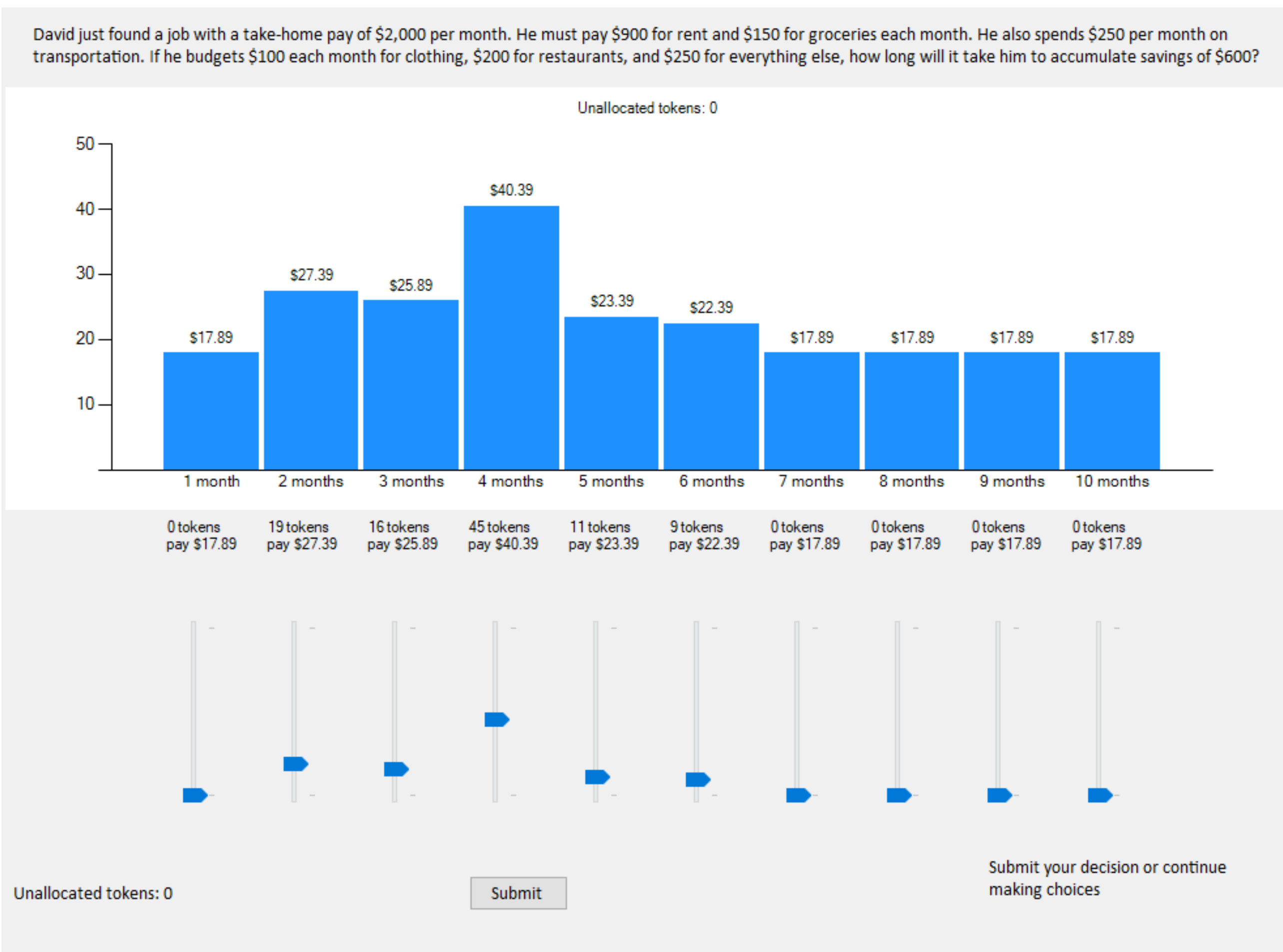
Figure 2.9 - Potential Earning Outcome with Bias and High Confidence, *fin9*



No Bias, Low Confidence

The lower-left panel of Figure 2.7 is the token allocation given by Subject #18. This subject exhibits no bias from the true answer and a low degree of confidence by allocating tokens across 5 different response bins, allocating 45 tokens to the bin that contained the true answer. Here the participant had a rough sense that the true answer was 4 months, whereby allocating the largest percentage of their overall tokens to that bin. However, they were less confident than Subject #15 and allocated their remaining 19, 16, 11, and 9 tokens across the response bins of 2, 3, 5, and 6 months, respectively. This scenario does not generate the maximum earnings for this question, but did afford the participant a potential payout over all bins, with actual earnings of $40.39. Figure 2.10 displays the earning outcomes for this participant. Recall, from Section 2.2.2, that the intuition as to why there are positive earnings over all bins including those with zero token allocation has to do with the incentivizing nature of the QSR that allows it to reward reports that are precise as possible. This participant is reasonably capable at planning a budget and reports diffused beliefs from 2 months to 6 months in order to accumulate savings of $600. It is significant that the subject is at least aware that they do not know the exact, correct answer.

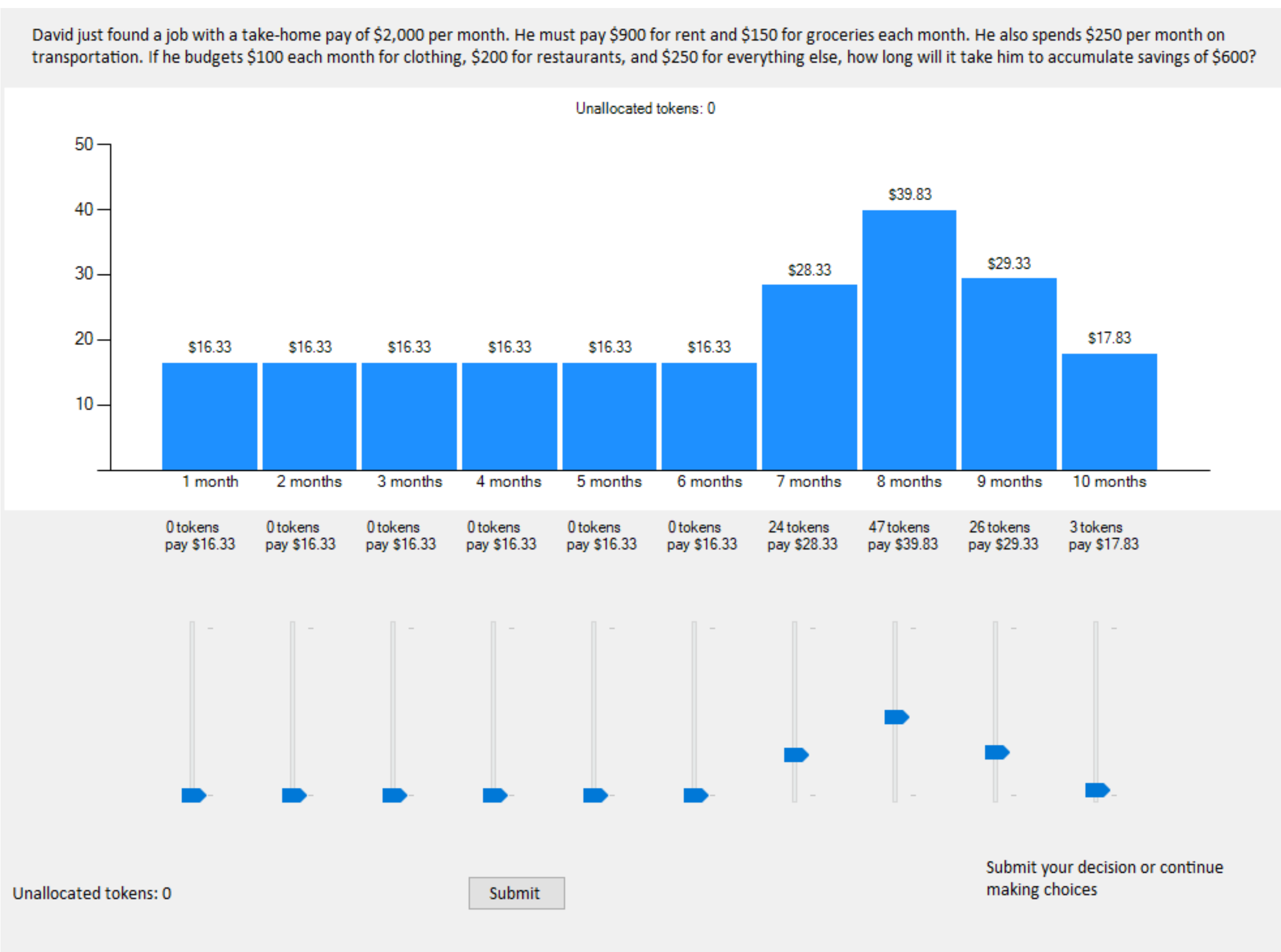
Figure 2.10 - Potential Earning Outcome with No Bias and Low Confidence, *fin9*



Bias, Low Confidence

Lastly, the lower-right panel of Figure 2.7 is the token allocation given by Subject #98. This subject exhibits a bias from the true answer and a low degree of confidence by allocating tokens across 4 different response bins, and no tokens to the bin that contained the true answer. Here the participant didn't believe that the true answer was 4 months, and instead allocated 24, 47, 26, and 3 tokens across the response bins of 7, 8, 9, and 10 months, respectively. This scenario does not generate the maximum earnings for this question, but did afford the participant a potential payout over all bins, with actual earnings of $16.33. Figure 2.10 displays the earning outcomes for this participant. This participant systematically overestimated the time it takes to accumulate savings of $600. Again, the fact that the subject is aware of their lack of precise knowledge in this instance may help mitigate the consequences of their bias.

Figure 2.11 - Potential Earning Outcome with Bias and Low Confidence, *fin9*



2.5.3 Constructing Measures of Literacy

For each subject in Figure 2.7 there are also 2 different literacy measures displayed, *L* and *W*. The measure *L* is defined as the percentage of 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. Subject #15 allocated all their tokens to the correct bin, thus $L = 1.0$ (= 100/100). Subject #18 allocated 45 tokens to the correct bin, thus $L = 0.45$ (= 45/100). This constructed *L* is data that we can use directly when undertaking estimation (i.e., it can be used as a covariate since it is data and not an estimate).

On the other hand, the *W* shown in Figure 2.7 is defined as the expected earnings for the bin containing the true answer divided by the maximum earning amount. It is motivated by the "efficiency" measure widely used for over 50 years in experimental economics. Recall the maximum earnings a participant could earn in this task is $50, which is the denominator for all subjects. Using Subject #18 as an example, we'll calculate the numerator and their measure of *W*. The numerator for this subject is equal to the actual payoff, $40.39, multiplied by the percentage of raw token allocation that is placed into the true bin, 0.45, which is the expected earnings of $18.18. Thus Subject #18's measure of $W = 0.36$ (= $18.18/$50). This measure gets at expected earnings, knowing what the true answer is, hence the focus on the true bin only. However, it has the weakness that the expectation is taken with respect to the share of tokens, which is only the subjective probability of the subject if the subject is risk neutral (or all bins with tokens have exactly the same number of tokens).

There are two other versions of efficiency measures that could be calculated given the data. Note that the denominator used in all efficiency measures use the maximum earnings a participant could earn in this task, $50. The first variation of efficiency measure, *W1*, is the percentage of actual earnings realized by the participant. Using subject #18 we see their actual payoff at $40.39, thus $W1 = 0.81$ (= $40.36/$50). Efficiency measure *W2* is reported as *W*, and explained above. The other variation of efficiency measure, *W3*, is another way to go beyond *W1* to get at expectation, and reflects the subject's uncertainty over all bins, not just the true bin as given in *W2* (reported as *W* above). To calculate the *W3* measure for subject #18, we sum every payoff amount over all 10 bins and multiply it by the respective percentage of tokens allocated to each bins. This total is $32.11 and is the expected earnings over all bins, which would then be divided by $50. Thus *W3* = 0.64 (= $32.11/$50) in this instance.

2.6 Results

The evaluation of results in this chapter focuses on individuals and utilizes the correction techniques of Harrison and Ulm (2016) in order to correct every participant's raw token allocations of reported subjective belief distributions to "recover" the true latent subjective belief distribution for all financial questions asked; the technique is described in Section 2.2.2. Each subject is typed at the individual level by their responses to the incentivized risk aversion task. At the 5% significance level, the overall frequency of each decision making model employed in recovering the true latent subject belief distribution is as follows: 40% of our subjects were typed as EUT, 5% were RDU using an inverse-S probability weighting function, 1.1% were RDU using a power probability weighting function, and the remainder 53.9% were RDU using a Prelec probability weighting specification.

The initial evaluation of results will look at pooled beliefs across all individuals using both initial and new labels to test the effect of shifting the labels. We then report on pooled beliefs across individuals using the new labels. Finally, we will look at beliefs when controlling for observable demographic characteristics, in order to evaluate hypotheses about how literacy could vary across demographics using the new labels.

2.6.1 Effect of Different Labeling Schemes

The first evaluation of results focuses on the pooled beliefs of individuals across both initial and new labels in order to test whether the shift in labels had a statistically significant effect on the average token reports of participants. Table 2.3 lists the sessions which individuals participated in using the initial labeling scheme as sessions 1 through 15 and session 18, and those using the new labeling scheme as sessions 19, 28, and 29. Table 2.5 presents interval regression estimates and evaluates results for two models: a constant model only and another model, M1, that introduces the covariate new_label, controlling for a change in the mean estimate due to the new label dummy, for all financial questions asked under both labeling schemes. The first column of Table 2.5 lists the financial question. The second column lists the number of individuals that responded by each question. The next column lists the model type, either the constant only model or the model M1. The next three columns are the parameter name with mean and standard deviation of the estimates. The next two columns are results from a hypothesis test, H0, that the coefficient on the model constant is equal to the true answer for each financial question. This is interpreted as a test of bias away from the true answer. The final two columns are testing the hypothesis, H1, that the covariate new_label has a statistically significant marginal effect at the 5% significance level. In other words, it is testing if the label shifts between initial and new labeling schemes had an effect on responses.

It is helpful to undertake a detailed example of the analysis for question fin1 using Table 2.5. We see that a total 304 individuals responded to this question, which included both initial and new labels. The coefficient on the constant in the constant only model is $108.70 with a standard deviation of $4.33, and we reject the hypothesis that the coefficient on the model constant is equal to the true answer at the 5% significance level. Recall that the true answer is $110.41. Stated another way, the subjects' average response was biased significantly biased from the true answer. Now consider the results for the one covariate model, M1. For question fin1 we see that the coefficient and standard deviation, respectively, for the model constant is $107.64 and $4.40 and for the covariate new_label is $3.10 and $3.38. The average response for participants presented with the new label scheme can be calculated by adding the estimated mean of the constant to the mean of the new_label, here $110.74 (= $107.64 + $3.10). We then perform a hypothesis test that the average response when controlling for the new labeling scheme is equal to the true answer, and here we *cannot* reject the hypothesis at the 5% significance level. Finally, for question fin1, we see that the *p*-value on the parameter estimate new_label is statistically significant at a level <0.000. From this result we infer that the different labeling scheme caused a statistically significant difference in average responses, such that biased responses became unbiased responses.

Reviewing Table 2.5, we can quickly see that for model M1 that a test of the null hypothesis of no bias could not be rejected at a significance level of 5% only for fin1. Thus responses to questions fin1 under the new labeling scheme were not biased from the true answer. All other questions failed that test and the average responses are statistically biased away from the true answer. Further, testing the marginal effect of the different labeling schemes shows that only two questions, fin13 and fin14, could not be rejected at a significance level of 5%.

This section broadly covers the need to analyze the different labeling schemes separately. Throughout the remainder of Chapter 2 and across Chapters 3 and 4 the results will be presented solely using the new labeling scheme. A detailed exposition on the labeling schemes used can be found in Appendix A.

2.6.2 Pooled Beliefs of Individuals with New Labels

The first evaluation of results focuses on pooled beliefs of individuals across new labels. Table 2.6 and Figure 2.12 illustrate the results for the question on a savings account with 2% interest, one of the core questions asked in the financial literacy field. The average belief is $110.74, with a 95% confidence interval between $110.27 and $111.20. The correct answer is $110.41, so these beliefs are generally unbiased. The literacy index *L* shows that only 53% of the tokens were correctly allocated to the bin containing the true answer. And the literacy index *W* shows that 50% of the possible earnings was attained, from perfect literacy (all 100 tokens allocated to the correct bin). For a subject pool of college undergraduates, these outcomes are a bit disappointing, but broadly consistent with findings from the *National Financial Capability Study in the United States: Data from the 2015 State-by-State Survey* (2015) that asks the same question, although in a multiple choice format. That study's choice architecture listed the following as options: a) More than $102; b) Exactly $102; c) Less than $102; d) Don't know; and e) Prefer not to say. The correct answer lies in option a) More than \$102, and 75% of their national sample and 71% of the Georgia sample answered correctly. It makes sense that we see a larger number of participants answering a multiple choice correctly over 5 outcomes instead of 10 outcomes.

However, we quickly see that literacy levels are even lower for other types of questions. Table 2.7 and Figure 2.13 show the results for the question about the earliest age at which Social Security benefits begin, which is actually 62. The average response is biased, with a 95% CI that does not span the true value, the imprecision of beliefs as shown by the wide dispersion of average beliefs across intervals lowers literacy by either measure. Table 2.8 and Figure 2.14 ask a comparable question about the age at which Medicare eligibility begins, which is 65. There is evidence of bias here as well, with the average response being 57.3 and the 95% CI being below the true answer. In addition, there is are more diffuse beliefs being reported here than to the Social Security question, resulting in even lower literacy indices.

Returning to more narrow economic literacy, Table 2.9 and Figure 2.15 shows results for the real interest rate question, also a staple of the general financial literacy literature. The modal token allocation for the individuals is in the correct bin. However, there is statistically significant evidence bias in quantitative terms, the average is \$101 and the correct answer is \$99.98, and the imprecision around this average response is large, resulting in low literacy indices of $L = 0.28$ and $W = 0.29$. These are lower with respect to the indices for the interest rate question that is also widely used in the financial literacy literature.

Table 2.10 and Figure 2.16 show results for the savings horizon question, an important dimension of literacy for retirement decisions. The mode of the token allocation is in the bin that contains the correct answer. Again, there is statistical evidence of bias, with the average response being 5 months and the 95% CI including 4.54 to 5.36 months, when the true answer is 4.0 months. The imprecision of beliefs is relatively low, leading to a literacy index L of 0.77 and a literacy index W of 0.74. These results suggest that individuals are only slightly biased, however, and appear to do well in constructing and calculating a budget in order to find the time it takes to accumulate a certain amount of savings, here \$600 in 4 months.

Table 2.11 and Table 2.12, and Figure 2.17 and Figure 2.18, consider paired questions on stolen credit cards and debit cards, respectively. The modal response in both cases was $0, or, more precisely, the bin representing the lowest interval between $0 and $5 (the mid-point to the $10 bin) for the stolen credit card and the interval between $0 and $25 (the midpoint to the $50 bin) for the stolen debit card, respectively. As it happens, the true answer is quite low for a credit card, at $50, and quite high for a debit card, at $500. Although the non-zero responses are "in the right direction" in the sense that they are more significant for the debit card question, they are striking evidence of imprecision. The responses for both questions are significantly biased and we see very low literacy indices.

Table 2.13 and Table 2.14, and Figure 2.19 and Figure 2.20, show results for two interest rate questions. These topics elicit some of the highest literacy rates, with some subjects knowing the correct answer with virtual certainty. Table 2.13, for question fin13, shows evidence of bias, with the average response of 4.45% and the 95% CI not including the true answer of 5%. Table 2.14, for fin14, shows evidence of slight bias, with the average response $101.74 and the 95% CI not including the true answer of $102. For both questions the modal response is the correct answer.

Table 2.15 and Table 2.16, and Figure 2.21 and Figure 2.22, close out with two basic question on longevity risk for men and women. Both are fundamental components of literacy with respect to retirement planning. In both cases there is statistically significant bias, with the averages being below the true values, and neither had a 95% confidence interval containing the correct answer. For the question about men the bias is 9.5 years, and for the question about women the bias is 11.7 years from the true answer. The imprecision is greater for the question about men, 9.7 years versus 9.53 years.

2.6.3 Demographics

The next evaluation of results looks at beliefs across all individuals when we control for observable demographic characteristics, in order to evaluate hypotheses about how literacy is associated with different demographics. In effect these results allow an insight into the pooled results, to see if there are any systematic demographic correlates of literacy, or if literacy tends to be the same across the sample. Results are still interpreted against the "omitted variable" of the average response for White male underclassmen.

Table 2.17 and Table 2.18 relate to fin1, the savings account question with 2% interest. The results listed in Table 2.17 are similar to the results listed in Table 2.5 and Table 2.6, but with one covariate "true" that is equal to the true answer to the question and then a series of demographics defined earlier. There is no constant term in the first part of the estimation. If the responses for the omitted variable are unbiased the coefficient on the parameter true in the first part of the estimation should equal 1, since there is no constant.[29] If there is bias we would see the parameter "true" grow larger or smaller than 1 and we can interpret this as a percent adjustment of the true answer. For example, the "true" parameter estimate listed of 1.01 for fin1 should be interpreted as a percent adjustment of the true answer to obtain the average response for White male underclassmen. Here the true parameter of 1.01 would be evaluated as 101.28% of the true answer of \$110.41, thus \$111.83, which is the average response reported for the omitted category. We see that the coefficient on the true parameter is 1.01 in Table 2.17 with a 95% confidence interval between 1.00 and 1.02. The *p*-value = 0.013 in the text above the table is reporting on a model test that the coefficient on the true parameter is equal to 1; here we reject the hypothesis that the true parameter is equal to 1 and is statistically significant at the 5% significance level.[30] Back inside the table text, we see the coefficient on the covariate for those with high grade point averages is statistically significant with an average of -1.52 and a confidence interval between -2.40 and -0.64. To evaluate the average response for those with high grades, we deduct \$1.52 from \$111.83 (calculated previously), which is \$110.31 and only \$0.10 less than the correct answer.

Recall from Table 2.6 that we found what might appear to be a slightly different result: statistically significant evidence of unbiasedness. Different results can occur and the reason, of course, is that the

---

[29] In Table 2.5 and Table 2.6, by contrast, the coefficient on the constant would have been equal to the true value if there was no bias.

[30] The *p*-value in the table is, as customary, testing the hypothesis that the model coefficient is equal to zero.

estimation in Table 2.6 implicitly assumed that all individuals held the same beliefs, whereas Table 2.17 allows for sampling variability in terms of the observed demographics under the new labeling scheme. These estimations are therefore "answering different inferential questions," and may thus lead to different answers.

The main objective of Table 2.18, however, is to study the effects of demographics in terms of bias and confidence. The latter is listed in the bottom panel of the table as the additional imprecision of beliefs. The top panel shows the bias from the correct answer for each demographic characteristic taken by itself. We observe that females, Asians, Blacks, Christians, and those with senior class standing hold statistically and quantitatively biased beliefs relative to the correct answer of \$110.41 at the 5% significance level. Conversely, participants that reported a high grade point average and those with junior class standing held unbiased beliefs. Illustrating an example using the covariate "gpaHI", those students with a high grade point average, we see their estimated bias from the correct answer at -\$0.10 and a *p*-value = 0.39, therefore we cannot reject the hypothesis that those with high grade point averages reported values statistically different than the correct answer. Note this is the same -\$0.10 amount that we calculated earlier in the example from Table 2.17. We could do these calculations by hand using the values supplied in Table 2.17, however, Table 2.18 allows a more natural interpretation of these effects.

Of course, that is just part of the literacy evaluation. Even if there is no bias, or to be precise no evidence of statistically significant bias, there could be "costly" imprecision in beliefs. The word costly is in quotation marks, because imprecision is directly costly in terms of the foregone potential earnings from the scoring rule, but it is possible that there is a beneficial imprecision when the individual senses that they might have a biased belief. Recall the earlier discussion of the pooled beliefs about the longevity risk of women. Hence the second panel of Table 2.18 shows the additional imprecision of beliefs by each demographic, compared to the average imprecision of the pooled sample. In this case the estimates reflect the standard deviation of beliefs, not the log of the standard deviation (referred to as "LnSigma" in Table 2.6 and Table 2.17). We infer significantly lower levels of imprecision for females, those with a high grade point average, and those with junior or senior class standing. However, only juniors and those with high grade point averages exhibited no evidence of statistically significant bias, thus the extra precision of their beliefs is enhancing their literacy measures.

Table 2.19 shows significant bias for the social security start question and the 95% CI for the true parameter ranging between 1.04 and 1.09. The top panel of Table 2.20 shows that all demographics reported a statistically significant bias from the correct answer. Evidence of statistically significant bias ranged from an average low of 2.67 years for Asians to an average high of 4.67 for females. The bottom panel of Table 2.20 shows that upperclassmen are the only demographics that had more precise beliefs compared to the average level of imprecision. Since both classes of upperclassmen, juniors and seniors, exhibited bias from the correct answer, this additional precision does not increase levels of literacy. Thus we have an instance where "everybody gets it wrong," as distinct from there being identifiable demographic groups that exhibit worse literacy than others.

Table 2.21 and Table 2.22 consider the Medicare eligibility question, where we see very low levels of literacy from the pooled responses. Table 2.21 informs us that White underclassmen underestimated the correct answer with the 95% CI of the true parameter ranging between 0.88 and 0.97. Table 2.22 allow us to say that the bias in literacy for this question is shared by all demographics across the board (top panel), and no demographic group exhibits significantly different precision than the average (bottom panel) at the 5% significance level. Again, we see "everybody gets it wrong" when considering the age at which a citizen of the US is normally eligible for Medicare.

Table 2.23 shows that there is significant bias away from the true answer for the question about the real interest rate across the pooled response with a p-value = 0.041. The top panel of Table 2.24 shows that students with a high grade point average and students in their junior year of study do better on this question, in the sense that they do not have a statistically significant bias from the correct answer, $98.98, at a 5% significance level. The bottom panel of Table 2.24 shows that no demographic exhibits different levels of precision compared to the average at the 5% significance level. However, at the 10% significance level, we see evidence of decreased imprecision for those with junior class standing; thus enhancing literacy measures for this group.

Table 2.25 shows the first example of an insignificant bias away from the true answer for the savings horizon question across pooled response. Stated another way, White underclassmen are unbiased with respect to the correct answer. However, Black and Christians exhibit statistically significant differences at the 5% significance level. The top panel of Table 2.26 shows that all demographics except Christians do not have a bias from the correct answer at a 5% significance level. In the bottom panel of Table 2.26 we see that all of those demographics listed have statistically significantly different levels of imprecision relative to the average level of imprecision. The corrections for each demographic are in fact diminishing the average imprecision, and getting closer towards exhibiting zero additional imprecision. Given that females, Asians, Blacks, those with a high grade point average, and students in their junior or senior year of study were unbiased from the correct answer, the effect the extra precision of their beliefs is enhancing their literacy measures.

Table 2.27 and Table 2.28 refer to the question about the stolen credit card. Table 2.27 reports a *p*-value = 0.730 in the text above the table, is reporting on a model test that the coefficient on the true parameter is equal to 1. Here we cannot reject the hypothesis that the true parameter is equal to 1 and is statistically significant at the 5% significance level. However, we see from the text inside the table that the mean estimate for the true parameter is 1.78 with a *p*-value = 0.43, which implies that we cannot reject the hypothesis that there is significant difference from an estimate equal to *zero*. This finding could be further evidence of diffuse beliefs being reported across all bin levels, but with the average belief being near the correct answer. Table 2.28 confirms a bias from the correct answer for only Christians. In each case of unbiased responses relative to the correct answer, literacy measures did not benefit from increased precision. Table 2.29 and Table 2.30 consider a similar question relating to the stolen debit card. The interpretation of results is similar as for the stolen credit card. The variation can be found in Table 2.30 that shows that all demographics are unbiased from the correct answer (top panel) and that all demographics, apart from Christians, hold more precise beliefs relative to the average (bottom panel).

Table 2.31 shows another example of an insignificant bias away from the true answer for the nominal interest rate question across the pooled response. Participants with junior class standing exhibit statistically significant differences at the 5% significance level. As there is insignificant bias away from the true answer over the pooled response, this finding suggests that students that are juniors have a different underlying belief distribution then the average White underclassmen. This finding could lead to different levels of literacy between the two demographics. The top panel of Table 2.32 show that Asians, Blacks, those with a high grade point average, and students in their junior or senior years of study are not biased relative to the correct answer at 5% significance level. The bottom panel of the table lends further support to enhanced literacy levels through the reduction of imprecision relative to the average imprecision for students in their junior year at the 5% significance level.

Table 2.33 shows the last example of an insignificant bias away from the true answer for the interest rate question across the pooled response. Table 2.34 reports that only Christians exhibit a statistically significant bias relative to the correct answer for this question (top panel), and females, Blacks, those with high grade point averages, and upperclassmen exhibit significantly different precision than the average (bottom panel) at the 5% significance level. Thus we have an instance where "most everybody gets it right," with Christians exhibiting slightly worse literacy than the others.

Table 2.35 and Table 2.36 consider the question about the remaining life for men conditional on reaching 20 years old. Table 2.35 shows lower levels of literacy from the pooled responses and statistically significant differences for females at the 5% significance level. Table 2.36 confirms the bias from the correct answer at the 5% significance level for everyone except students with high grade point averages and those in their junior year. In the bottom panel of the table we see that of the two unbiased demographics, only those with a high grade point average exhibited less imprecision in relation to the average imprecision. Since those with high grade point averages are statistically unbiased from the correct answer, this increase in precision implies a higher literacy measure for that demographic.

Table 2.37 and Table 2.38 round out the analysis and consider the question about the remaining life for women conditional on reaching 20 years old. The results are strikingly similarly to the results for the previous question asking the same longevity question about men. Table 2.37 shows lower levels of literacy from the pooled responses and a statistically significant difference for females and those with higher grade point averages at the 5% significance level. Table 2.38 shows that those with a high grade point average are the only demographic that do not exhibit bias relative to the correct answer, and that is statistically significant at the 5% level. These results vanish at the 10% level.

2.7 Conclusions

This chapter began with a presentation of questions identified in the current literature that are being used to measure financial literacy. A total of 16 questions are initially identified, which were also uniquely suited to be asked in a subjective belief elicitation setting. These questions span various topics such as interest and inflation, financial rules, regulations, and procedures, budgeting, and longevity risk. Of the overall 16 questions identified, a subset of 11 questions were used in the final analysis.

The chapter then covered the statistical foundations necessary to present a quantitative framework for measuring financial literacy using subjective belief elicitation. This research utilizes scoring rules and methods operationalized by Harrison and Phillips (2014), Harrison and Ulm (2016) and Harrison et al. (2017). The elicitation method was selected for several reasons. First it allows us to incentivize responses using a scoring rule that pays subjects for performing better. With knowledge of the risk preferences of subjects, these responses are also incentive-compatible, in the sense that the experiments can recover true latent beliefs from observed token allocations under the scoring rule. The elicitation format also permits us to infer a rich characterization of an individual's entire subjective belief distribution. Using that information, we develop a construct to characterize responses using concepts of confidence and bias. The elicitation format also enables a definition of several measure of literacy, such as the percent of raw tokens that were allocated to the correct answer or other efficiency measures that analyze an individual's overall earnings.

After the theoretical framework of subjective belief elicitation is developed the experimental tasks that the subjects experienced were defined. Each subject completed a battery of binary choices over risky lotteries, a battery of token allocations for beliefs, and a standard socio-demographic survey. Each task addresses a component required by the theoretical framework.

The next topic covered was the econometric specification of the model. There was some discussion of the various methods available, and interval regression was discussed in detail. A thorough discussion explains how to correctly interpret the output of these regressions, along with other constructed measures of literacy, bias, and confidence.

Finally, we put the theory, planning, experimentation, and econometric specification into practice and examine results for a subset of individual participants under different labeling schemes and controlling for heterogeneity of observable demographics. The techniques covered in this chapter form the foundation for the analyses of treatments in subsequent chapters. The results presented here for individuals serve also as a "control" group when testing literacy under different scaffolds in later chapters. Table 2.39 displays a summary view of the financial literacy questions and the associated pooled L and W indices across the results previously displayed in this chapter's tables shown as standalones. It further classifies the questions into 3 categories: numeracy, procedural, and longevity risk, to allow a more nuanced evaluation of which scaffolds perform better under what circumstances in the next two chapters.

Table 2.5: Evaluating the Effect of New Labels

Question	Answer	N	Model	Parameter	mean	sd	H0: (mean=true) stat. sig. diff. at 0.05 level	p-value	H1: new_label is stat. sig. diff. at 0.05 level	p-value
fin1	$110.41	304	Constant	_cons	108.70	4.33	YES	<0.001		
			M1	_cons	107.64	4.40				
			M1	new_label	3.10	3.38	NO	0.171	YES	<0.001
fin2	age 62	322	Constant	_cons	63.13	4.18	YES	<0.001		
			M1	_cons	61.84	3.84				
			M1	new_label	3.92	3.58	YES	<0.001	YES	<0.001
fin5	age 65	304	Constant	_cons	59.92	5.03	YES	<0.001		
			M1	_cons	61.34	4.27				
			M1	new_label	-3.99	5.30	YES	<0.001	YES	<0.001
fin7	$98.98	614	Constant	_cons	99.76	2.22	YES	<0.001		
			M1	_cons	99.50	2.13				
			M1	new_label	1.48	2.24	YES	<0.001	YES	<0.001
fin9	4 months	322	Constant	_cons	4.42	1.69	YES	<0.001		
			M1	_cons	4.15	1.27				
			M1	new_label	0.80	2.25	YES	<0.001	YES	0.000
fin10	$50.00	304	Constant	_cons	238.56	515.70	YES	<0.001		
			M1	_cons	270.96	601.32				
			M1	new_label	-99.96	283.08	YES	<0.001	YES	0.018
fin11	$500.00	322	Constant	_cons	625.70	1105.79	YES	0.021		
			M1	_cons	457.09	803.44				
			M1	new_label	484.71	1495.55	YES	0.001	YES	0.001
fin13	5%	304	Constant	_cons	4.56	1.17	YES	<0.001		
			M1	_cons	4.62	1.13				
			M1	new_label	-0.16	1.24	YES	<0.001	NO	0.167
fin14	$102.00	266	Constant	_cons	101.80	1.35	YES	0.002		
			M1	_cons	101.84	1.40				
			M1	new_label	-0.11	1.25	YES	0.003	NO	0.395
fin15	57.1 years	266	Constant	_cons	49.19	12.53	YES	<0.001		
			M1	_cons	50.26	14.05				
			M1	new_label	-2.68	9.70	YES	<0.001	YES	0.030
fin16	61.7 years	266	Constant	_cons	52.59	12.35	YES	<0.001		
			M1	_cons	54.10	13.77				
			M1	new_label	-3.77	9.53	YES	<0.001	YES	0.002

* Note H0: refers to the hypothesis that the coefficient on the model constant is equal to the true answer for Model = Constant, and is evaluated as (_cons + new_label) for Model = M1

* Note H1: refers to the p-value for the parameter estimate new_label

Table 2.6: Interval Regression Estimates for
the Question about the Savings Account with 2% Interest, New Labels

Average: \$110.7 Standard deviation: \$3.4
Correct answer: \$110.41 N=106 GSU undergraduates
Literacy indices $L = 0.53$ and $W = 0.50$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	110.74	<0.001	110.27	111.20
		LnSigma		
constant	1.22	<0.001	1.09	1.34
		Sigma		
sigma	3.38		2.97	3.84

Figure 2.12: Elicited Beliefs about
The Savings Account with 2% Interest Question,
Individual Responses, New Labels

Average: \$110.7 Standard deviation: \$3.4
Correct answer: \$110.41 N=106 GSU undergraduates
Literacy indices $L = 0.53$ and $W = 0.50$

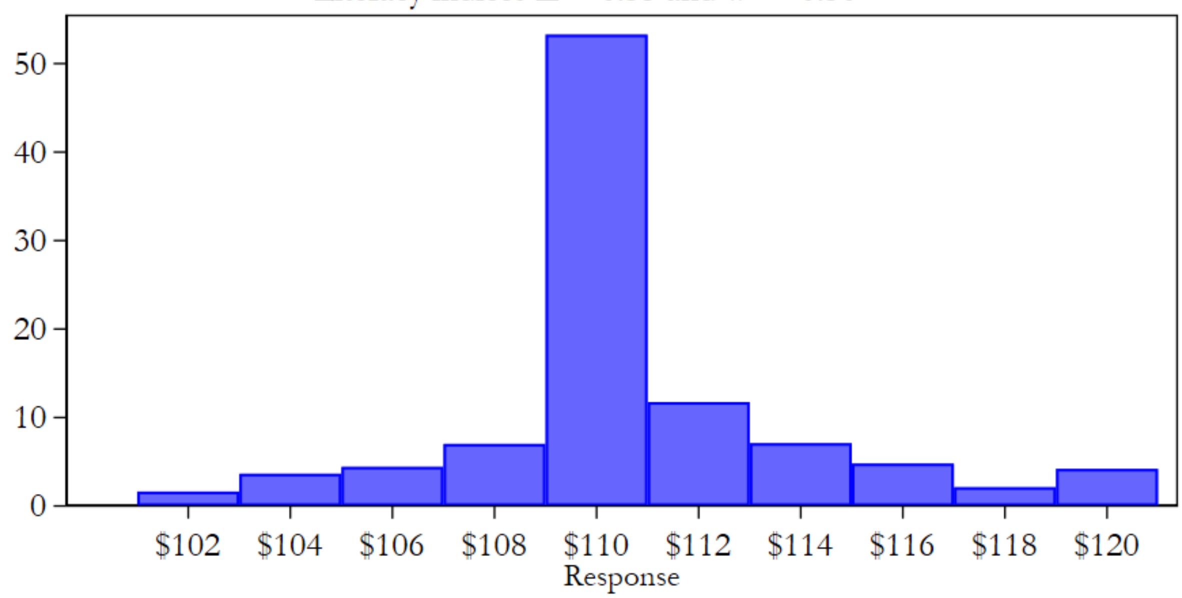


Table 2.7: Interval Regression Estimates for the Question about the Social Security Start Age, New Labels

Average: 65.8 years   Standard deviation: 3.6 years
Correct answer: 62   N=106 GSU undergraduates
Literacy indices $L = 0.26$ and $W = 0.21$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	65.75	<0.001	65.24	66.26
		LnSigma		
constant	1.27	<0.001	1.16	1.39
		Sigma		
sigma	3.58		3.19	4.01

Figure 2.13: Elicited Beliefs about The Social Security Start Question, Individual Responses, New Labels

Average: 65.8 years Standard deviation: 3.6 years
Correct answer: 62 N=106 GSU undergraduates
Literacy indices $L = 0.26$ and $W = 0.21$

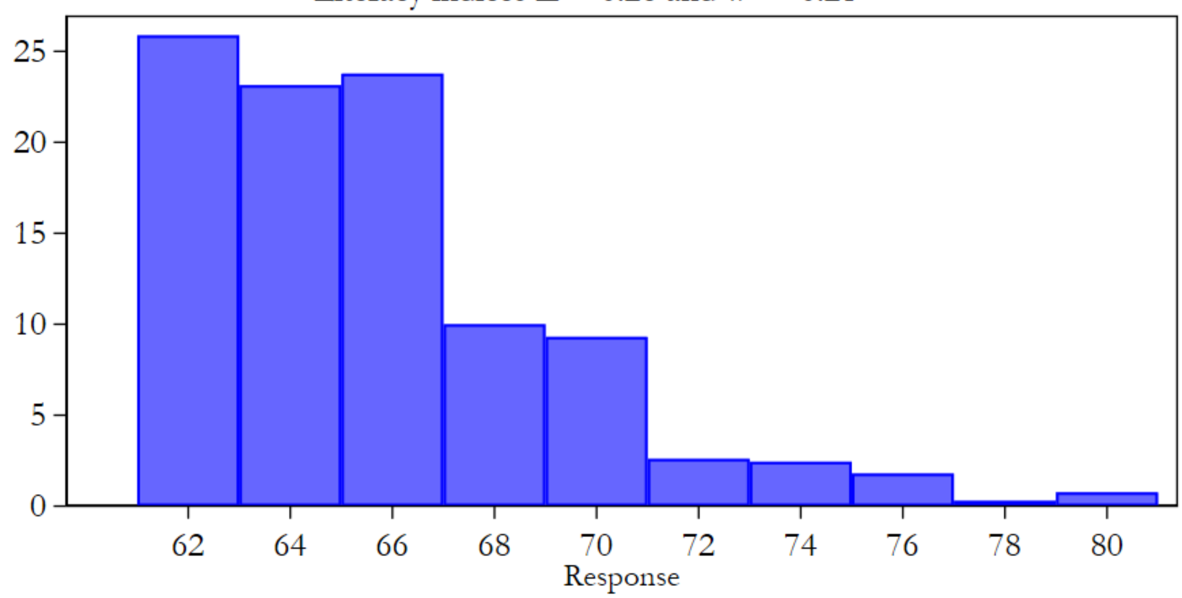


Table 2.8: Interval Regression Estimates for the Question about Medicare Eligibility, New Labels

Average: 57.3 years Standard deviation: 5.3 years
Correct answer: 65 N=106 GSU undergraduates
Literacy indices $L = 0.16$ and $W = 0.16$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	57.35	<0.001	56.46	58.23
		LnSigma		
constant	1.67	<0.001	1.61	1.73
		Sigma		
sigma	5.30		4.98	5.63

Figure 2.14: Elicited Beliefs about The Medicare Eligibility Question, Individual Responses, New Labels

Average: 57.3 years Standard deviation: 5.3 years
Correct answer: 65 N=106 GSU undergraduates
Literacy indices $L = 0.16$ and $W = 0.16$

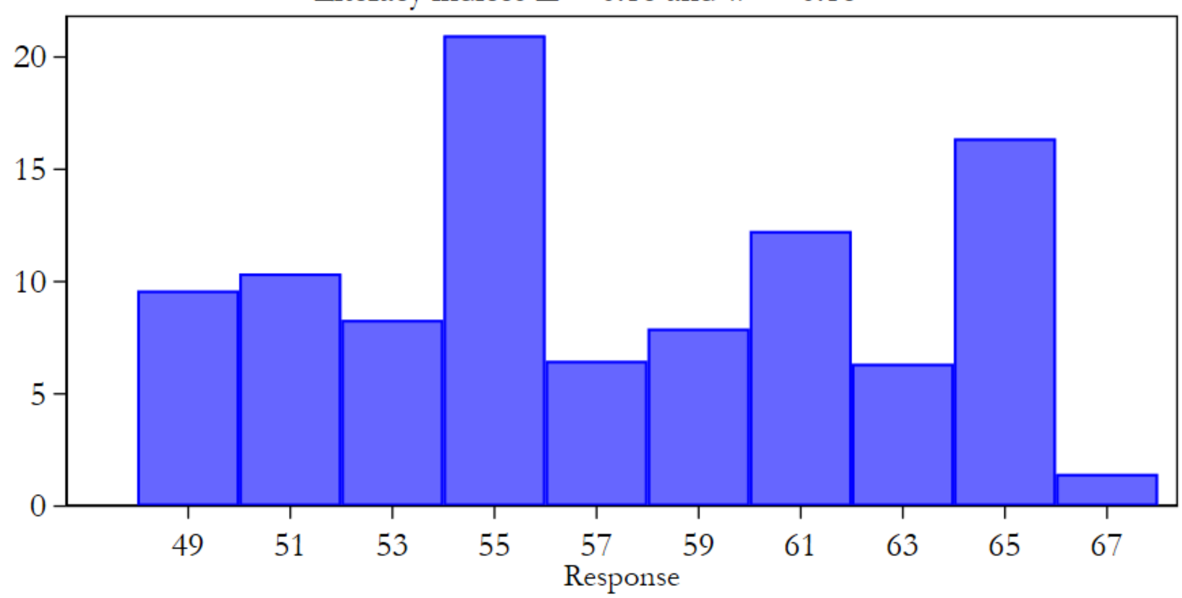


Table 2.9: Interval Regression Estimates for the Question about the Real Interest Rate, New Labels

Average: \$101.0 Standard deviation: \$2.2
Correct answer: \$98.98 N=106 GSU undergraduates
Literacy indices L = 0.28 and W = 0.29

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	100.98	<0.001	100.62	101.35
		LnSigma		
constant	0.81	<0.001	0.73	0.88
		Sigma		
sigma	2.24		2.08	2.41

Figure 2.15: Elicited Beliefs about The Real Interest Rate Question, Individual Responses, New Labels

Average: \$101.0 Standard deviation: \$2.2
Correct answer: \$98.98 N=106 GSU undergraduates
Literacy indices L = 0.28 and W = 0.29

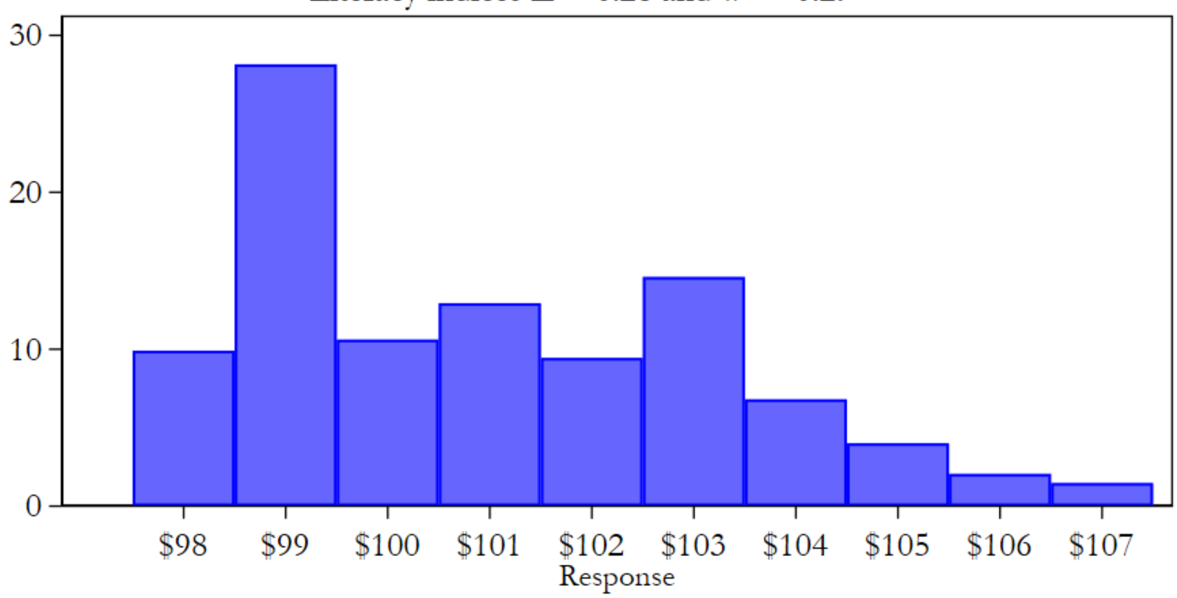


Table 2.10: Interval Regression Estimates for the Question about the Savings Horizon, New Labels

Average: 5.0 months Standard deviation: 2.2 months
Correct answer: 4 months (4m) N=106 GSU undergraduates
Literacy indices $L = 0.77$ and $W = 0.74$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	4.95	<0.001	4.54	5.36
		LnSigma		
constant	0.81	<0.001	0.57	1.05
		Sigma		
sigma	2.25		1.76	2.86

Figure 2.16: Elicited Beliefs about The Savings Horizon Question, Individual Responses, New Labels

Average: 5.0 months Standard deviation: 2.2 months
Correct answer: 4 months (4m) N=106 GSU undergraduates
Literacy indices $L = 0.77$ and $W = 0.74$

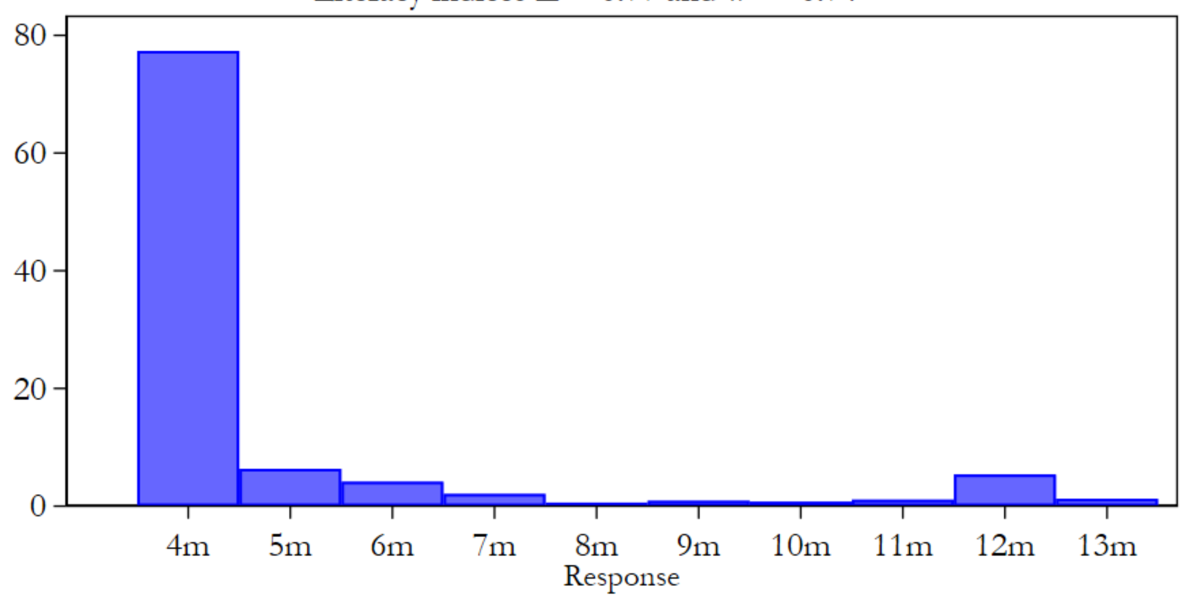


Table 2.11: Interval Regression Estimates for
the Question about the Stolen Credit Card, New Labels

Average: \$171.0 Standard deviation: \$283.1
Correct answer: \$50 N=106 GSU undergraduates
Literacy indices $L = 0.06$ and $W = 0.06$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	171.00	<0.001	120.00	222.00
		LnSigma		
constant	5.65	<0.001	5.47	5.82
		Sigma		
sigma	283.08		237.52	337.37

Figure 2.17: Elicited Beliefs about
The Stolen Credit Card Question,
Individual Responses, New Labels

Average: \$171.0 Standard deviation: \$283.1
Correct answer: \$50 N=106 GSU undergraduates
Literacy indices $L = 0.06$ and $W = 0.06$

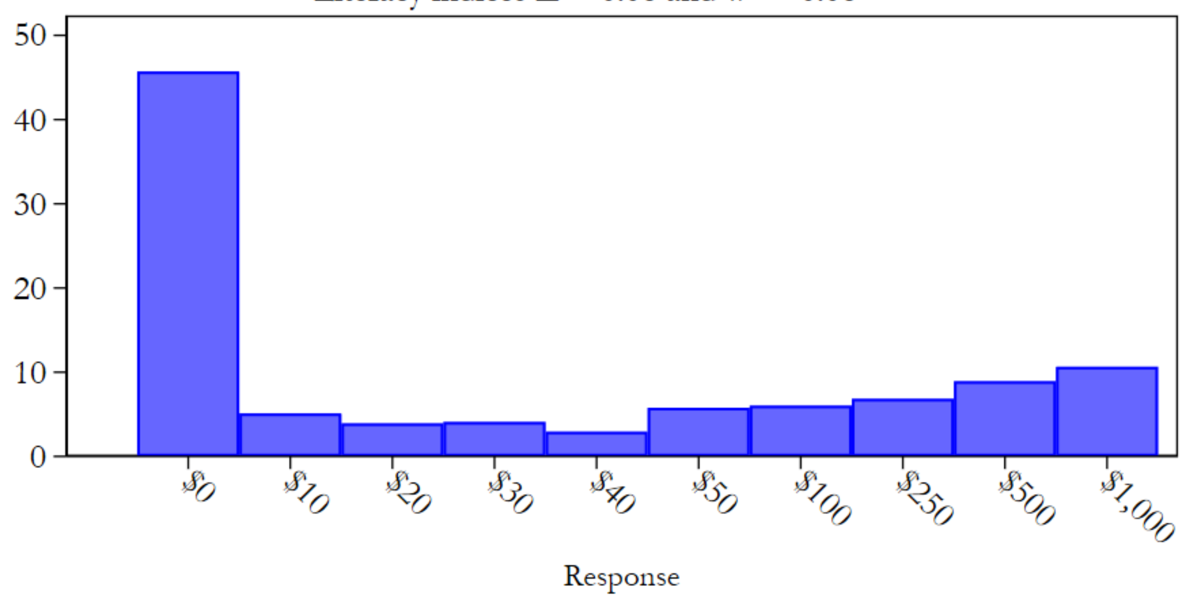


Table 2.12: Interval Regression Estimates for the Question about the Stolen Debit Card, New Labels

Average: \$941.8 Standard deviation: \$1495.6
Correct answer: \$500 N=106 GSU undergraduates
Literacy indices $L = 0.16$ and $W = 0.12$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	941.80	<0.001	679.38	1204.22
		LnSigma		
constant	7.31	<0.001	6.97	7.66
		Sigma		
sigma	1495.55		1059.22	2111.63

Figure 2.18: Elicited Beliefs about The Stolen Debit Card Question, Individual Responses, New Labels

Average: \$941.8 Standard deviation: \$1495.6
Correct answer: \$500 N=106 GSU undergraduates
Literacy indices $L = 0.16$ and $W = 0.12$

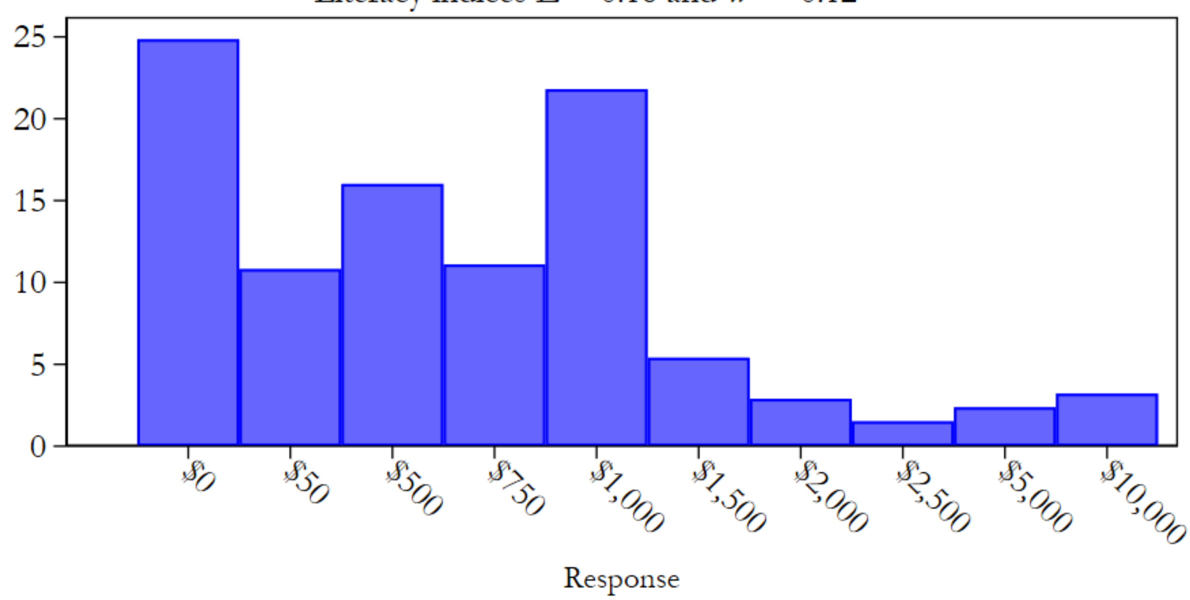


Table 2.13: Interval Regression Estimates for the Question about Nominal Interest, New Labels

Average: 4.5% Standard deviation: 1.2%
Correct answer: 5% N=106 GSU undergraduates
Literacy indices L = 0.80 and W = 0.76

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	4.45	<0.001	4.26	4.65
		LnSigma		
constant	0.21	0.03	0.02	0.40
		Sigma		
sigma	1.24		1.02	1.50

Figure 2.19: Elicited Beliefs about The Nominal Interest Question, Individual Responses, New Labels

Average: 4.5% Standard deviation: 1.2%
Correct answer: 5% N=106 GSU undergraduates
Literacy indices L = 0.80 and W = 0.76

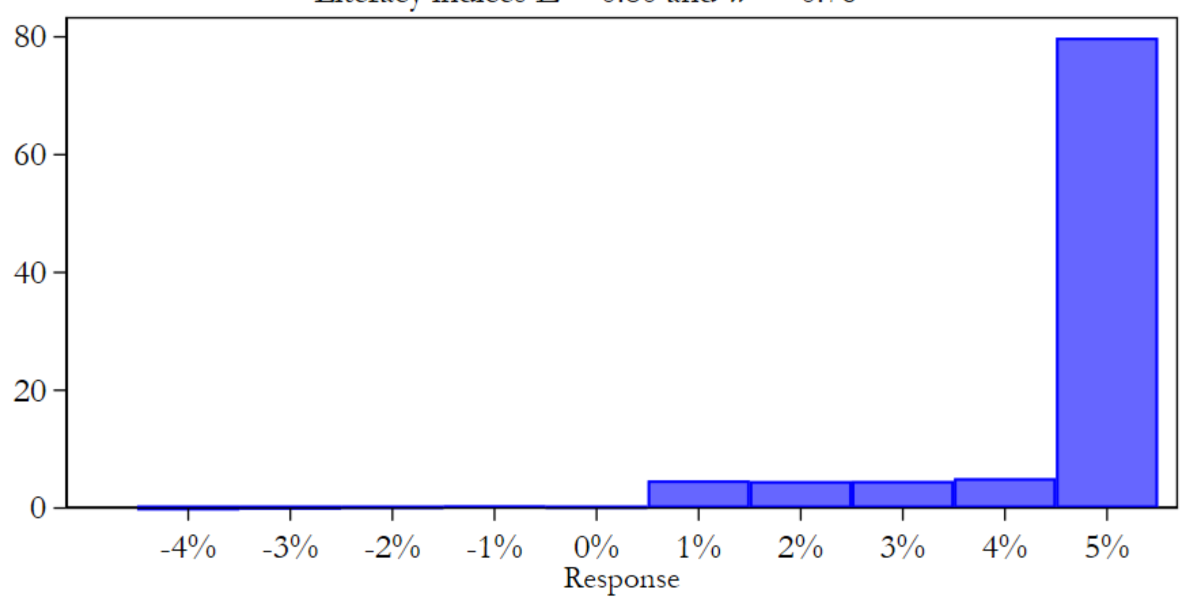


Table 2.14: Interval Regression Estimates for
the Question about the Interest Rate, New Labels

Average: \$101.7 Standard deviation: \$1.3
Correct answer: \$102.00 N=106 GSU undergraduates
Literacy indices $L = 0.74$ and $W = 0.68$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	101.74	<0.001	101.56	101.91
		LnSigma		
constant	0.23	0.02	0.04	0.41
		Sigma		
sigma	1.25		1.04	1.51

Figure 2.20: Elicited Beliefs about
The Interest Rate Question,
Individual Responses, New Labels

Average: \$101.7 Standard deviation: \$1.3
Correct answer: \$102.00 N=106 GSU undergraduates
Literacy indices $L = 0.74$ and $W = 0.68$

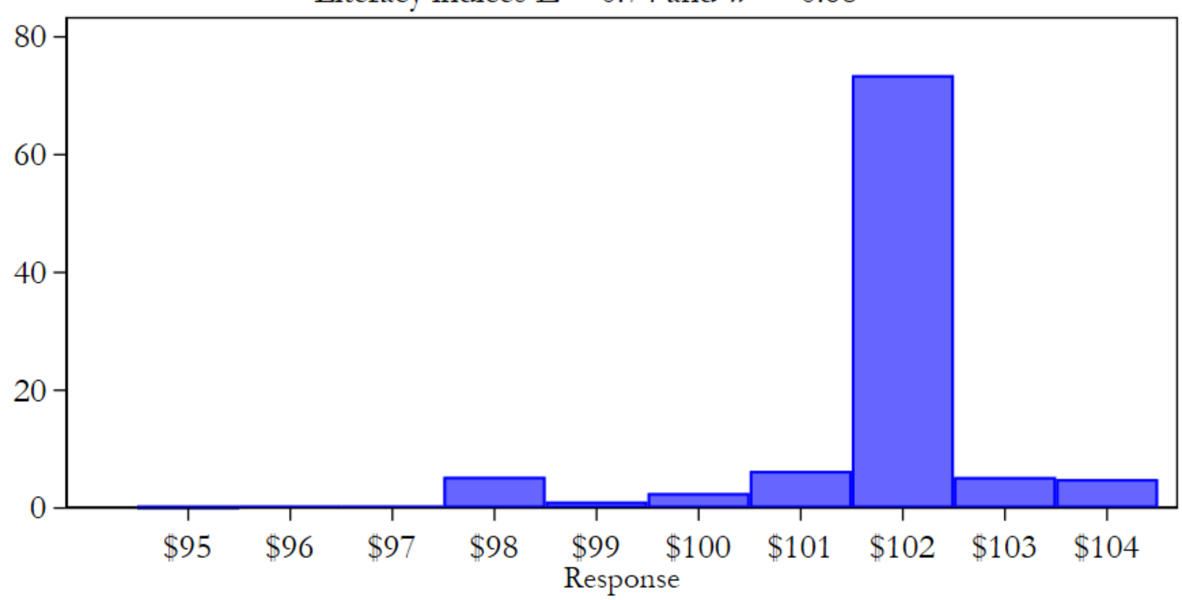


Table 2.15: Interval Regression Estimates for the Question about the Remaining Life for Men, New Labels

Average: 47.6 years Standard deviation: 9.7 years
Correct answer: 57.1 N=106 GSU undergraduates
Literacy indices L = 0.22 and W = 0.18

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	47.58	<0.001	45.94	49.22
		LnSigma		
constant	2.27	<0.001	2.15	2.40
		Sigma		
sigma	9.70		8.56	10.98

Figure 2.21: Elicited Beliefs about The Remaining Life for Men Question, Individual Responses, New Labels

Average: 47.6 years Standard deviation: 9.7 years
Correct answer: 57.1 N=106 GSU undergraduates
Literacy indices L = 0.22 and W = 0.18

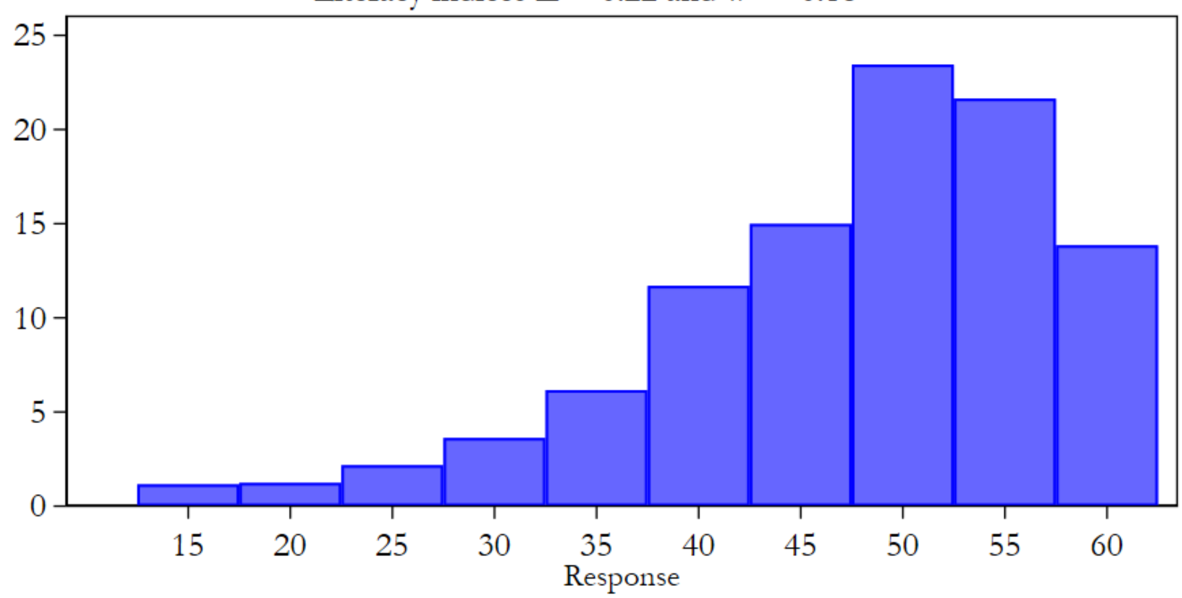


Table 2.16: Interval Regression Estimates for the Question about the Remaining Life for Women, New Labels

Average: 50.3 years Standard deviation: 9.5 years
Correct answer: 61.7 N=106 GSU undergraduates
Literacy indices $L = 0.27$ and $W = 0.22$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	50.34	<0.001	48.79	51.88
		LnSigma		
constant	2.25	<0.001	2.12	2.39
		Sigma		
sigma	9.53		8.29	10.95

Figure 2.22: Elicited Beliefs about The Remaining Life for Women Question, Individual Responses, New Labels

Average: 50.3 years Standard deviation: 9.5 years
Correct answer: 61.7 N=106 GSU undergraduates
Literacy indices $L = 0.27$ and $W = 0.22$

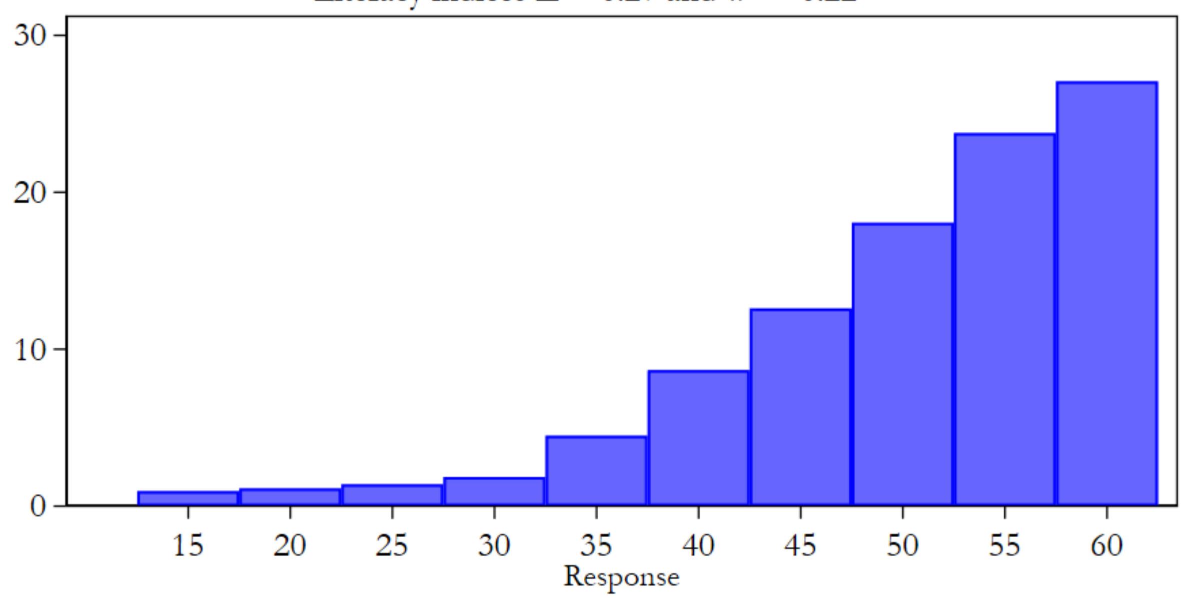


Table 2.17: Interval Regression Estimates, With Controls for Demographics, for the Question about the Savings Account with 2% Interest, New Labels

Correct answer: \$110.41 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.013

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	1.01	<0.001	1.00	1.02
female	-0.09	0.83	-0.90	0.72
asian	0.60	0.23	-0.37	1.57
black	-0.12	0.77	-0.93	0.69
christian	-0.08	0.85	-0.88	0.73
gpaHI	-1.52	<0.001	-2.40	-0.64
junior	-0.85	0.09	-1.85	0.15
senior	-0.34	0.45	-1.20	0.53
		LnSigma		
female	-0.06	0.72	-0.37	0.25
asian	0.30	0.16	-0.12	0.71
black	0.50	<0.001	0.12	0.87
christian	0.20	0.18	-0.09	0.48
gpaHI	-0.16	0.26	-0.44	0.12
junior	-0.40	0.06	-0.82	0.01
senior	-0.18	0.38	-0.58	0.22
constant	0.82	<0.001	0.41	1.22

Table 2.18: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Account with 2% Interest, New Labels

Bias is relative to the correct answer: $110.41
Additional imprecision is relative to the average imprecision: 3.38

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
Bias from the True Response				
female	1.33	0.02	0.23	2.42
asian	2.02	<0.001	0.76	3.27
black	1.30	0.05	0.05	2.55
christian	1.34	<0.001	0.45	2.23
gpaHI	-0.10	0.39	-1.40	1.19
junior	0.57	0.26	-0.63	1.77
senior	1.08	0.04	0.09	2.08
Additional Imprecision of Beliefs				
female	-1.24	0.02	-2.26	-0.22
asian	-0.33	0.34	-1.52	0.86
black	0.34	0.35	-1.02	1.70
christian	-0.62	0.15	-1.51	0.26
gpaHI	-1.45	<0.001	-2.37	-0.53
junior	-1.86	<0.001	-2.61	-1.11
senior	-1.49	<0.001	-2.37	-0.61

Table 2.19: Interval Regression Estimates, With Controls for Demographics, for the Question about the Social Security Start Age, New Labels

Correct answer: 62 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is < 0.001

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	1.07	<0.001	1.04	1.09
female	0.59	0.26	-0.43	1.61
asian	-1.41	0.05	-2.83	0.02
black	-1.05	0.08	-2.22	0.12
christian	0.25	0.64	-0.79	1.29
gpaHI	0.47	0.35	-0.51	1.45
junior	-0.04	0.95	-1.24	1.17
senior	-0.69	0.20	-1.76	0.38
		LnSigma		
female	-0.08	0.50	-0.30	0.15
asian	-0.27	0.22	-0.70	0.16
black	0.11	0.48	-0.18	0.40
christian	-0.03	0.83	-0.28	0.23
gpaHI	0.27	0.01	0.06	0.49
junior	-0.40	<0.001	-0.63	-0.16
senior	-0.30	<0.001	-0.51	-0.09
constant	1.26	<0.001	0.89	1.63

Table 2.20: Literacy Bias and Imprecision, by Demographics, for the Question about the Social Security Start Age, New Labels

Bias is relative to the correct answer: 62
Additional imprecision is relative to the average imprecision: 3.58

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
Bias from the True Response				
female	4.67	<0.001	3.25	6.08
asian	2.67	<0.001	1.04	4.30
black	3.03	<0.001	1.60	4.46
christian	4.33	<0.001	2.98	5.68
gpaHI	4.55	<0.001	2.77	6.32
junior	4.04	<0.001	2.70	5.38
senior	3.38	<0.001	1.91	4.86
Additional Imprecision of Beliefs				
female	-0.31	0.36	-1.56	0.94
asian	-0.89	0.13	-2.06	0.28
black	0.35	0.34	-0.84	1.53
christian	-0.14	0.39	-1.32	1.03
gpaHI	1.06	0.19	-0.66	2.77
junior	-1.21	0.01	-2.09	-0.32
senior	-0.95	0.05	-1.86	-0.05

Table 2.21: Interval Regression Estimates, With Controls for Demographics, for the Question about Medicare Eligibility, New Labels

Correct answer: 65 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is < 0.001

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.92	<0.001	0.88	0.97
female	0.95	0.36	-1.09	2.99
asian	-1.06	0.55	-4.53	2.42
black	-2.19	0.16	-5.21	0.83
christian	-1.42	0.20	-3.58	0.74
gpaHI	-1.05	0.27	-2.91	0.81
junior	-1.23	0.40	-4.08	1.62
senior	0.75	0.58	-1.91	3.41
		LnSigma		
female	-0.02	0.83	-0.19	0.15
asian	0.04	0.77	-0.26	0.35
black	0.12	0.38	-0.15	0.39
christian	0.01	0.94	-0.17	0.19
gpaHI	-0.05	0.53	-0.19	0.10
junior	0.10	0.59	-0.25	0.44
senior	0.16	0.14	-0.05	0.37
constant	1.52	<0.001	1.21	1.83

Table 2.22: Literacy Bias and Imprecision, by Demographics, for the Question about Medicare Eligibility, New Labels

Bias is relative to the correct answer: 65
Additional imprecision is relative to the average imprecision: 5.30

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-4.05	0.03	-7.50	-0.59
asian	-6.05	<0.001	-8.99	-3.11
black	-7.19	<0.001	-9.74	-4.63
christian	-6.42	<0.001	-9.15	-3.68
gpaHI	-6.05	<0.001	-9.47	-2.63
junior	-6.23	<0.001	-9.74	-2.71
senior	-4.25	0.01	-7.44	-1.05
		Additional Imprecision of Beliefs		
female	-0.81	0.18	-2.05	0.44
asian	-0.51	0.28	-1.68	0.66
black	-0.13	0.39	-1.64	1.37
christian	-0.69	0.22	-1.92	0.54
gpaHI	-0.93	0.17	-2.33	0.48
junior	-0.26	0.38	-2.01	1.49
senior	0.07	0.40	-1.36	1.49

Table 2.23: Interval Regression Estimates, With Controls for Demographics, for the Question about the Real Interest Rate, New Labels

Correct answer: $98.98 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.041

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	1.02	<0.001	1.00	1.03
female	0.72	0.03	0.08	1.36
asian	0.30	0.67	-1.06	1.65
black	-0.06	0.93	-1.31	1.19
christian	0.75	0.05	0.01	1.49
gpaHI	-0.68	0.05	-1.36	-0.01
junior	-0.92	0.02	-1.69	-0.16
senior	-0.11	0.82	-1.08	0.86
		LnSigma		
female	0.00	0.97	-0.18	0.19
asian	-0.06	0.73	-0.40	0.28
black	0.01	0.97	-0.26	0.27
christian	-0.04	0.73	-0.29	0.20
gpaHI	-0.23	<0.001	-0.40	-0.06
junior	-0.42	0.02	-0.79	-0.06
senior	-0.05	0.71	-0.29	0.20
constant	0.95	<0.001	0.57	1.34

Table 2.24: Literacy Bias and Imprecision, by Demographics, for the Question about the Real Interest Rate, New Labels

Bias is relative to the correct answer: $98.98
Additional imprecision is relative to the average imprecision: 2.24

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	2.28	<0.001	0.74	3.83
asian	1.86	<0.001	0.72	3.00
black	1.50	<0.001	0.58	2.42
christian	2.31	<0.001	0.97	3.66
gpaHI	0.88	0.20	-0.57	2.33
junior	0.64	0.28	-0.85	2.12
senior	1.45	0.04	0.14	2.76
		Additional Imprecision of Beliefs		
female	0.36	0.29	-0.51	1.24
asian	0.20	0.35	-0.60	1.00
black	0.37	0.27	-0.44	1.17
christian	0.24	0.33	-0.54	1.03
gpaHI	-0.17	0.37	-1.02	0.67
junior	-0.54	0.08	-1.14	0.05
senior	0.23	0.31	-0.43	0.89

Table 2.25: Interval Regression Estimates, With Controls for Demographics, for the Question about the Savings Horizon, New Labels

Correct answer: 4 months (4m) N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.123

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.95	<0.001	0.89	1.01
female	0.05	0.70	-0.21	0.31
asian	0.26	0.15	-0.10	0.62
black	0.52	0.01	0.11	0.92
christian	0.55	<0.001	0.23	0.88
gpaHI	-0.04	0.80	-0.35	0.27
junior	0.49	0.28	-0.39	1.37
senior	0.29	0.19	-0.15	0.73
		LnSigma		
female	-0.09	0.71	-0.59	0.41
asian	0.82	0.01	0.17	1.47
black	1.39	<0.001	0.95	1.83
christian	1.04	<0.001	0.52	1.55
gpaHI	-0.16	0.45	-0.58	0.26
junior	1.25	<0.001	0.58	1.92
senior	0.62	<0.001	0.20	1.03
constant	-1.63	<0.001	-2.36	-0.89

Table 2.26: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Horizon, New Labels

Bias is relative to the correct answer: 4 months (4m)
Additional imprecision is relative to the average imprecision: 2.25

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-0.15	0.24	-0.45	0.15
asian	0.06	0.38	-0.28	0.40
black	0.31	0.15	-0.13	0.76
christian	0.35	0.03	0.06	0.65
gpaHI	-0.24	0.14	-0.57	0.09
junior	0.29	0.33	-0.65	1.23
senior	0.09	0.37	-0.40	0.58
		Additional Imprecision of Beliefs		
female	-2.07	<0.001	-2.19	-1.95
asian	-1.80	<0.001	-2.14	-1.46
black	-1.46	<0.001	-2.00	-0.91
christian	-1.69	<0.001	-1.98	-1.40
gpaHI	-2.08	<0.001	-2.20	-1.96
junior	-1.56	<0.001	-2.21	-0.91
senior	-1.88	<0.001	-2.18	-1.59

Table 2.27: Interval Regression Estimates, With Controls for Demographics, for the Question about the Stolen Credit Card, New Labels

Correct answer: $50 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.730

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	1.78	0.43	-2.66	6.22
female	10.46	0.86	-104.51	125.43
asian	9.61	0.93	-205.73	224.96
black	55.01	0.37	-66.01	176.04
christian	82.67	0.30	-74.29	239.63
gpaHI	11.55	0.83	-94.16	117.26
junior	-71.17	0.46	-261.27	118.93
senior	-50.31	0.63	-252.65	152.03
		LnSigma		
female	0.14	0.57	-0.33	0.61
asian	0.11	0.88	-1.31	1.52
black	0.33	0.45	-0.54	1.21
christian	0.24	0.53	-0.51	0.98
gpaHI	-0.00	1.00	-0.42	0.42
junior	-0.14	0.73	-0.95	0.66
senior	-0.03	0.94	-0.88	0.82
constant	5.14	<0.001	3.69	6.58

Table 2.28: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Credit Card, New Labels

Bias is relative to the correct answer: $50
Additional imprecision is relative to the average imprecision: 283.08

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	49.49	0.38	-246.37	345.35
asian	48.64	0.29	-72.11	169.39
black	94.04	0.23	-82.22	270.30
christian	121.69	0.03	18.34	225.04
gpaHI	50.58	0.37	-185.47	286.62
junior	-32.14	0.35	-149.10	84.81
senior	-11.28	0.39	-149.16	126.59
		Additional Imprecision of Beliefs		
female	-88.13	0.35	-427.18	250.92
asian	-94.16	0.12	-214.01	25.69
black	-45.69	0.36	-238.40	147.02
christian	-68.02	0.30	-248.35	112.31
gpaHI	-113.26	0.24	-337.39	110.87
junior	-135.51	0.08	-285.77	14.74
senior	-118.13	0.13	-270.94	34.69

Table 2.29: Interval Regression Estimates, With Controls for Demographics, for the Question about the Stolen Debit Card, New Labels

Correct answer: \$500   N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.837

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.91	0.04	0.04	1.78
female	-217.12	0.20	-550.17	115.92
asian	-2.83	0.99	-535.82	530.16
black	320.13	0.20	-173.31	813.56
christian	380.83	0.01	80.97	680.69
gpaHI	-7.75	0.96	-292.49	276.98
junior	-17.94	0.91	-343.12	307.24
senior	459.74	0.04	10.52	908.96
		LnSigma		
female	-0.30	0.24	-0.80	0.20
asian	-0.41	0.35	-1.27	0.45
black	0.17	0.67	-0.61	0.95
christian	0.70	0.02	0.14	1.27
gpaHI	0.17	0.37	-0.21	0.55
junior	0.04	0.90	-0.58	0.66
senior	0.94	<0.001	0.46	1.43
constant	6.34	<0.001	5.53	7.15

Table 2.30: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Debit Card, New Labels

Bias is relative to the correct answer: $500
Additional imprecision is relative to the average imprecision: 1495.55

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-262.98	0.24	-778.73	252.76
asian	-48.68	0.39	-422.08	324.71
black	274.27	0.09	-41.44	589.98
christian	334.97	0.07	-19.02	688.96
gpaHI	-53.61	0.39	-671.26	564.04
junior	-63.79	0.38	-463.76	336.18
senior	413.88	0.16	-195.72	1023.48
		Additional Imprecision of Beliefs		
female	-1075.55	<0.001	-1526.49	-624.60
asian	-1120.63	<0.001	-1296.03	-945.22
black	-823.68	<0.001	-1148.38	-498.98
christian	-349.52	0.24	-1020.08	321.05
gpaHI	-821.32	0.01	-1444.50	-198.14
junior	-906.41	<0.001	-1475.53	-337.30
senior	-42.25	0.40	-1230.72	1146.23

Table 2.31: Interval Regression Estimates, With Controls for Demographics, for the Question about Nominal Interest, New Labels

Correct answer: 5%   N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.169

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.92	<0.001	0.80	1.04
female	-0.24	0.15	-0.57	0.09
asian	-0.17	0.51	-0.67	0.33
black	-0.02	0.93	-0.35	0.32
christian	-0.26	0.06	-0.54	0.02
gpaHI	0.17	0.18	-0.08	0.42
junior	0.38	0.04	0.01	0.76
senior	0.31	0.15	-0.12	0.74
		LnSigma		
female	0.14	0.58	-0.36	0.65
asian	0.32	0.38	-0.40	1.05
black	-0.00	0.99	-0.48	0.48
christian	0.46	0.12	-0.12	1.04
gpaHI	-0.40	0.05	-0.81	0.00
junior	-0.75	<0.001	-1.20	-0.30
senior	-0.22	0.44	-0.79	0.34
constant	0.02	0.97	-0.80	0.83

Table 2.32: Literacy Bias and Imprecision, by Demographics, for the Question about Nominal Interest, New Labels

Bias is relative to the correct answer: 5%
Additional imprecision is relative to the average imprecision: 1.24

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-0.66	0.03	-1.21	-0.10
asian	-0.59	0.09	-1.25	0.08
black	-0.43	0.06	-0.87	0.01
christian	-0.68	0.01	-1.18	-0.18
gpaHI	-0.25	0.28	-0.81	0.32
junior	-0.03	0.39	-0.47	0.41
senior	-0.11	0.35	-0.51	0.30
		Additional Imprecision of Beliefs		
female	-0.06	0.40	-1.00	0.87
asian	0.17	0.38	-0.99	1.33
black	-0.22	0.34	-0.97	0.53
christian	0.38	0.27	-0.47	1.23
gpaHI	-0.56	0.07	-1.15	0.03
junior	-0.75	<0.001	-1.15	-0.35
senior	-0.42	0.16	-1.04	0.19

Table 2.33: Interval Regression Estimates, With Controls for Demographics, for the Question about the Interest Rate, New Labels

Correct answer: \$102.00 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.678

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	1.00	<0.001	1.00	1.00
female	-0.02	0.84	-0.22	0.17
asian	-0.38	0.11	-0.85	0.09
black	-0.07	0.47	-0.25	0.12
christian	-0.27	0.02	-0.48	-0.05
gpaHI	0.08	0.37	-0.09	0.24
junior	0.06	0.56	-0.15	0.28
senior	0.14	0.24	-0.09	0.38
		LnSigma		
female	0.29	0.14	-0.09	0.67
asian	0.78	0.02	0.10	1.45
black	0.06	0.78	-0.39	0.52
christian	0.81	<0.001	0.36	1.26
gpaHI	-0.18	0.35	-0.56	0.20
junior	-0.50	0.08	-1.06	0.05
senior	-0.25	0.28	-0.70	0.20
constant	-0.60	0.04	-1.18	-0.02

Table 2.34: Literacy Bias and Imprecision, by Demographics, for the Question about the Interest Rate, New Labels

Bias is relative to the correct answer: \$102.00
Additional imprecision is relative to the average imprecision: 1.25

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-0.06	0.33	-0.26	0.14
asian	-0.42	0.09	-0.90	0.06
black	-0.11	0.22	-0.31	0.09
christian	-0.31	0.05	-0.60	-0.02
gpaHI	0.04	0.38	-0.18	0.25
junior	0.02	0.39	-0.19	0.24
senior	0.10	0.31	-0.19	0.39
		Additional Imprecision of Beliefs		
female	-0.52	0.04	-1.01	-0.03
asian	-0.06	0.39	-0.86	0.74
black	-0.67	<0.001	-0.98	-0.36
christian	-0.02	0.40	-0.55	0.51
gpaHI	-0.80	<0.001	-1.07	-0.52
junior	-0.92	<0.001	-1.18	-0.66
senior	-0.83	<0.001	-1.07	-0.59

Table 2.35: Interval Regression Estimates, With Controls for Demographics, for the Question about the Remaining Life for Men, New Labels

Correct answer: 57.1 N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is 0.016

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.93	<0.001	0.87	0.99
female	-3.95	0.01	-7.06	-0.84
asian	-2.91	0.26	-7.94	2.12
black	-3.32	0.09	-7.11	0.48
christian	-1.25	0.45	-4.49	2.00
gpaHI	2.43	0.12	-0.62	5.47
junior	1.00	0.62	-3.00	5.01
senior	-1.94	0.30	-5.61	1.74
		LnSigma		
female	0.27	0.02	0.05	0.48
asian	0.28	0.29	-0.24	0.80
black	0.22	0.21	-0.12	0.56
christian	0.16	0.29	-0.13	0.45
gpaHI	-0.33	<0.001	-0.55	-0.11
junior	0.09	0.68	-0.32	0.49
senior	0.23	0.09	-0.04	0.51
constant	1.80	<0.001	1.42	2.19

Table 2.36: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Men, New Labels

Bias is relative to the correct answer: 57.1
Additional imprecision is relative to the average imprecision: 9.70

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
female	-8.19	<0.001	-13.06	-3.32
asian	-7.15	<0.001	-11.36	-2.94
black	-7.56	<0.001	-11.16	-3.95
christian	-5.49	<0.001	-8.93	-2.04
gpaHI	-1.81	0.28	-6.04	2.41
junior	-3.24	0.15	-7.79	1.31
senior	-6.18	0.02	-10.98	-1.37
		Additional Imprecision of Beliefs		
female	-1.78	0.22	-4.93	1.38
asian	-1.68	0.25	-5.14	1.79
black	-2.17	0.08	-4.53	0.20
christian	-2.60	0.05	-5.06	-0.15
gpaHI	-5.34	<0.001	-7.16	-3.53
junior	-3.08	0.08	-6.45	0.29
senior	-2.03	0.20	-5.44	1.39

Table 2.37: Interval Regression Estimates, With Controls for Demographics, for the Question about the Remaining Life for Women, New Labels

Correct answer: 61.7   N=106 GSU undergraduates
p-value for test of hypothesis that true is 1 is < 0.001

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
true	0.87	<0.001	0.81	0.94
female	-3.74	0.01	-6.64	-0.83
asian	-4.76	0.16	-11.43	1.91
black	-1.75	0.33	-5.25	1.75
christian	-0.84	0.58	-3.77	2.10
gpaHI	3.29	0.03	0.33	6.25
junior	0.67	0.70	-2.73	4.07
senior	-0.96	0.59	-4.49	2.57
		LnSigma		
female	0.26	0.03	0.03	0.49
asian	0.55	0.06	-0.03	1.13
black	0.16	0.33	-0.16	0.48
christian	0.15	0.30	-0.13	0.44
gpaHI	-0.31	0.02	-0.56	-0.05
junior	-0.01	0.96	-0.37	0.35
senior	0.24	0.10	-0.05	0.52
constant	1.78	<0.001	1.45	2.10

Table 2.38: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Women, New Labels

Bias is relative to the correct answer: 61.7
Additional imprecision is relative to the average imprecision: 9.53

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
Bias from the True Response				
female	-11.55	<0.001	-16.33	-6.78
asian	-12.58	<0.001	-18.37	-6.78
black	-9.57	<0.001	-13.32	-5.81
christian	-8.65	<0.001	-11.78	-5.53
gpaHI	-4.53	0.07	-9.27	0.21
junior	-7.15	<0.001	-11.08	-3.21
senior	-8.78	<0.001	-13.40	-4.16
Additional Imprecision of Beliefs				
female	-1.86	0.17	-4.61	0.89
asian	0.68	0.39	-4.38	5.74
black	-2.59	0.02	-4.68	-0.51
christian	-2.65	0.02	-4.81	-0.48
gpaHI	-5.18	<0.001	-7.06	-3.31
junior	-3.67	<0.001	-5.80	-1.54
senior	-2.04	0.19	-5.31	1.23

Table 2.39: Pooled Measures of L and W Indices for Individual "Control" Subjects

Question	Type	Correct Answer	Individual Literacy Measures L	W
fin1 - Savings Account 2%	Numeracy	$110.41	0.53	0.50
fin2 - Social Security Start Age	Procedural	62	0.26	0.21
fin5 - Medicare Eligibility	Procedural	65	0.16	0.16
fin7 - Real Interest Rate	Numeracy	$98.98	0.28	0.29
fin9 - Savings Horizon	Numeracy	4 months	0.77	0.74
fin10 - Stolen Credit Card	Procedural	$50	0.06	0.06
fin11 - Stolen Debit Card	Procedural	$500	0.16	0.12
fin13 - Nominal Interest	Numeracy	5%	0.80	0.76
fin14 - Interest Rate	Numeracy	$102	0.74	0.68
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22	0.18
fin16 - Remaining Life for Women	Longevity Risk	61.7 years	0.27	0.22

*Pooled literacy measures of L and W are initally reported for each question in their respective section.

Appendix A: Label and other Biases and their Effects on Measurements of Literacy

One can argue that if a participant doesn't know an answer for certain that there could be a tendency for them to favor the first bin, the middle bin, or the last bin as a heuristic for selecting where to place their 100 tokens. What is the effect of shifting the labels so that the interval that contains the true answer moves from bin 1 to bin 5, or bin 7 to bin 3, for example? This is just a shifting of the 10 bins we have subjects allocate tokens over and re-labeling them. We're calling this the **Label treatment** to test the effect of the new labels.

Table A1 below shows the bin labels from the **Control** group using the initial label scheme and Table A2 the bin labels from the **Label treatment** using the new labeling scheme. In each of the tables we place an asterisk next to the value and highlight the bin that contains the true answer. Comparing the labels of *fin1* across the two tables we see that the bin that contained the true answer in the control was bin 10 and then shifted to bin 5 in the label treatment, a shift of 5 bins. Table A3 shows the absolute shift that occurred for all the financial questions.

Table A 1 - Interval Labels for Control Group with Initial Labeling Scheme

QuestionID	Correct Answer	Bin1	Bin2	Bin3	Bin4	Bin5	Bin6	Bin7	Bin8	Bin9	Bin10
fin1	$110.41	92	94	96	98	100	102	104	106	108	110*
fin2	62	54	56	58	60	62*	64	66	68	70	72
fin5	65	55	57	59	61	63	65*	67	69	71	73
fin7	$98.98	95	96	97	98	99*	100	101	102	103	104
fin9	4 months	1	2	3	4*	5	6	7	8	9	10
fin10	$50.00	0	25	50*	100	250	500	1000	2500	5000	10000
fin11	$500.00	0	25	50	100	250	500*	1000	2500	5000	10000
fin13	5%	0%	1%	2%	3%	4%	5%*	6%	7%	8%	9%
fin14	$102.00	98	99	100	101	102*	103	104	105	106	107
fin15	57.1 years	0	10	20	30	40	50	60*	70	80	90
fin16	61.7 years	0	10	20	30	40	50	60*	70	80	90

Note: * denotes the bin containing the true answer.

Table A 2 - Interval Labels for Treatment Group with New Labeling Scheme

QuestionID	Correct Answer	Bin1	Bin2	Bin3	Bin4	Bin5	Bin6	Bin7	Bin8	Bin9	Bin10
fin1	$110.41	102	104	106	108	110*	112	114	116	118	120
fin2	62	62*	64	66	68	70	72	74	76	78	80
fin5	65	49	51	53	55	57	59	61	63	65*	67
fin7	$98.98	98	99*	100	101	102	103	104	105	106	107
fin9	4 months	4*	5	6	7	8	9	10	11	12	13
fin10	$50.00	0	10	20	30	40	50*	100	250	500	1000
fin11	$500.00	0	50	500*	750	1000	2500	5000	7500	1000	2500
fin13	5%	-4%	-3%	-2%	-1%	0%	1%	2%	3%	4%	5%*
fin14	$102.00	95	96	97	98	99	100	101	102*	103	104
fin15	57.1 years	15	20	25	30	35	40	45	50	55*	60
fin16	61.7 years	15	20	25	30	35	40	45	50	55	60*

Note: * denotes the bin containing the true answer.

Table A 3 - Absolute Bin Shift between Control and Treatment Groups

QuestionID	Mean Shift of True Answer Bin Control vs Treatment
fin1	5
fin2	4
fin5	3
fin7	3
fin9	3
fin10	3
fin11	3
fin13	4
fin14	3
fin15	2
fin16	3

There could also be a widening or narrowing of the bin response interval labels and we would have to think through what effect that would have on literacy. One can show that as the response intervals widen for the bins, they would naturally contain more responses within a bin's response space. The wider the response interval labels, the more we allow a participant with lower literacy to get "close to" the bin that contains the true answer without knowing it outright. Conversely, the more narrow or tight we make the bin response interval labels, the more literate we require the subject to be to select the bin that contains the true answer. This is akin to asking someone to throw a dart and hit the dartboard (widening) versus asking them to hit the bullseye on the dartboard (narrowing). This is just a re-labeling of the 10 bins we have subjects allocate tokens over and allow some variation in the bin width.

While this is a separate bias related to precision of belief and could have had its own treatment, I unfortunately did not put this into the design of the experiment prior to running. The following questions have had their bin widths adjusted conjoint with the **Label treatment**: *fin10*, *fin11*, *fin15*, *fin16*. Table A4 below shows the difference within a bin's response space from the control less the Label treatment. Let's illustrate an example using *fin1*. From Table A1 above we see that the control labels for bin 1 and bin 2 are 92 and 94, a difference of 2. Likewise, from Table A2 the Label treatment labels for bin 1 and bin 2 are 102 and 104, also a difference of 2. Thus, looking at the difference of the label widths between the control and Label treatment is 2 minus 2, or 0, which is the number shown in Table A4 for "Bin 1&2" for *fin1* below.

Table A 4 - Differences in Label Widths Between Control and Treatment Groups

QuestionID	Bin 1&2	Bin 2&3	Bin 3&4	Bin 4&5	Bin 5&6	Bin 6&7	Bin 7&8	Bin 8&9	Bin 9&10
fin1	0	0	0	0	0	0	0	0	0
fin2	0	0	0	0	0	0	0	0	0
fin5	0	0	0	0	0	0	0	0	0
fin7	0	0	0	0	0	0	0	0	0
fin9	0	0	0	0	0	0	0	0	0
fin10	15	15	40	140	240	450	1350	2250	4500
fin11	-25	-425	-200	-100	-1250	-2000	-1000	9000	3500
fin13	0	0	0	0	0	0	0	0	0
fin14	0	0	0	0	0	0	0	0	0
fin15	5	5	5	5	5	5	5	5	5
fin16	5	5	5	5	5	5	5	5	5

For *fin10* we can see that the range of bin response intervals varied from 15 dollars to 4500 dollars and for *fin11* the bin response intervals varied from -25 to 3500 dollars over bins 1 through 10 for the control compared to the Label treatment response, respectively. While these are not zero, I feel the amounts are small and will introduce only a slight confound in the analysis for these questions, however, are not overly concerning.

Appendix B: Recruitment Protocol

The email invitation text sent to potential participants was as follows:

You have been invited to the following experiment:

Date: XXX
Time: XXX
Location: Andrew Young School of Policy Studies, room 447

This experiment is on a first register, first to participate basis. All recruits who take part in the experiment are guaranteed a $5.00 fee; those who show up on time, but cannot participate will get a $5.00 show up fee and priority in later experiments.

Please login to the Experiment Recruiter at http://excen.gsu.edu/recruiter/
to confirm or decline your participation in this experiment. Replying to this message will NOT confirm or decline participation in the experiment.

While you may login to the Recruiter and Accept or Decline this invitation at your earliest convenience, subjects will be confirmed on a first come, first served basis.

You must bring your Student ID Card in order to participate in the experiment.

Thank you.

Full participation was usually confirmed within 24 hours of sending out the recruitment emails.

Appendix C: Experimental Instructions

1. General Introduction

Welcome to the experiment today. We are going to ask you to make two types of decisions, and to answer some survey questions in-between. Any earnings you accumulate will be in addition to the show-up fee you have received just for being here.

We will read through the instructions for the first task in a moment. After you have completed those decisions an experimenter will come around to your booth and select one decision for payment and make a record of your earnings. You will then be asked to answer some survey questions. When everybody has caught up, we will read through the instructions for the second task. When you have completed those decisions an experimenter will again come around to your booth, select one decision for payment, and make a record of your total earnings. We will then ask one of you at a time to come up to be paid.

Are there any questions?

2. Demographic Questions

Q1. What is your AGE?

1. 16 (1)
2. 17 (2)
3. 18 (3)
4. 19 (4)
5. 20 (5)
6. 21 (6)
7. 22 (7)
8. 23 (8)
9. 24 (9)
10. 25 (10)
11. 26 (11)
12. 27 (12)
13. 28 (13)
14. 29 (14)
15. 30 (15)
16. 31 (16)
17. 32 (17)
18. 33 (18)
19. 34 (19)
20. 35 (20)
21. 36 (21)
22. 37 (22)
23. 38 (23)
24. 39 (24)
25. 40 (25)
26. 41 (26)
27. 42 (27)
28. 43 (28)
29. 44 (29)
30. 45 (30)
31. 46 (31)
32. 47 (32)
33. 48 (33)
34. 49 (34)
35. 50 (35)
36. 51 (36)
37. 52 (37)
38. 53 (38)
39. 54 (39)
40. 55 (40)
41. 56 (41)
42. 57 (42)
43. 58 (43)
44. 59 (44)
45. 60 (45)
46. 61 (46)
47. 62 (47)
48. 63 (48)
49. 64 (49)
50. 65 (50)
51. 66 (51)
52. 67 (52)
53. 68 (53)
54. 69 (54)
55. 70 (55)
56. 71 (56)
57. 72 (57)
58. 73 (58)
59. 74 (59)
60. 75 (60)
61. 76 (61)
62. 77 (62)
63. 78 (63)
64. 79 (64)
65. 80 (65)
66. 81 (66)
67. 82 (67)
68. 83 (68)
69. 84 (69)
70. 85 (70)
71. 86 (71)
72. 87 (72)
73. 88 (73)
74. 89 (74)
75. 90 (75)
76. 91 (76)
77. 92 (77)
78. 93 (78)
79. 94 (79)
80. 95 (80)
81. 96 (81)
82. 97 (82)
83. 98 (83)
84. 99 (84)

Q2. What is your sex?

85. Male (1)
86. Female (2)
87. Other (please specify) (3) ____________________

Q2_TEXT. What is your sex?

[Text Entry for: Other (please specify)]

Q3. Which of the following categories best describes you?

88. White/Caucasian (1)
89. Black/African-American (2)
90. African (3)
91. Asian-American (4)
92. Asian (5)
93. Hispanic-American (6)
94. Hispanic (7)
95. Mixed Race (8)
96. Other (please specify) (9) ____________________

Q4. What is your major? (select all that apply)

97. Accounting (1)
98. Economics (2)
99. Finance (3)
100. Business Administration, other than Accounting, Economics, or Finance (4)
101. Education (5)
102. Engineering (6)
103. Health Professions (7)
104. Public Affairs or Social Services (8)
105. Biological Sciences (9)
106. Math, Computer Sciences, or Physical Sciences (10)
107. Social Sciences or History (11)
108. Humanities (12)
109. Psychology (13)
110. Other Fields (please elaborate) (14)
111. Other Fields (please elaborate) (14_text)
112. Does not apply (15)

Q5. What is your class standing?

113. Freshman (1)
114. Sophomore (2)
115. Junior (3)
116. Senior (4)
117. Masters (5)
118. Doctoral (6)
119. Does not apply (7)

Q6. What is the highest level of education you expect to complete?

120. Bachelor's Degree (1)
121. Master's Degree (2)
122. Doctoral Degree (3)
123. First Professional Degree (4)
124. High School Diploma or GED (5)
125. Less than High School (6)

Q7. What was the highest level of education that your father (or male guardian) completed?

126. Less than High School (1)
127. GED or High School Equivalency (2)
128. High School (3)
129. Vocational or Trade School (4)
130. College or University (5)
131. Don't Know (6)

Q8. What was the highest level of education that your mother (or female guardian) completed?

132. Less than High School (1)
133. GED or High School Equivalency (2)
134. High School (3)
135. Vocational or Trade School (4)
136. College or University (5)
137. Don't Know (6)

Q9. What is your citizenship status in the United States?

138. U.S. Citizen (1)
139. Resident Alien (2)
140. Non-Resident Alien (3)
141. Other Status (please elaborate) (4) ____________________

Q9_TEXT. What is your citizenship status in the United States?

[Text entry for: Other Status (please elaborate)]

Q10. Are you a foreign student on a Student Visa?

142. Yes (1)
143. No (2)

Q11. Are you currently...?

144. Single and never married? (1)
145. Married? (2)
146. Separated, divorced, or widowed? (3)

Q12. On a 4-point scale, what is your current GPA if you are doing a Bachelor's degree, or what was it when you did a Bachelor's degree? This GPA should refer to all of your coursework, not just the current year. (please select one)

147. Between 3.75 and 4.0 GPA (mostly A's) (1)
148. Between 3.25 and 3.74 GPA (about half A's and half B's) (2)
149. Between 2.75 and 3.24 GPA (mostly B's) (3)
150. Between 2.25 and 2.74 GPA (about half B's and half C's) (4)
151. Between 1.75 and 2.24 GPA (mostly C's) (5)
152. Between 1.25 and 1.74 GPA (about half C's and half D's) (6)
153. Less than 1.25 GPA (mostly D's or below) (7)
154. Have not taken course for which grades are given (8)

Q13. We are interested in knowing what kind of background you have in Economics. From the following choices, please select all of the Economics courses that you have taken. (select all that apply)

155. The Global Economy (ECON 2100) (1)
156. Principles of Macroeconomics (ECON 2105)Finance (2)
157. Principles of Microeconomics (ECON 2106) (3)
158. Macroeconomics - CTW (ECON 3900) (4)
159. Microeconomics (ECON 3910) (5)
160. Other (please elaborate) (6)
161. Text entry for: Other (please elaborate) (6_text)

Q14. Where do you live now? That is, where do you stay most often?

162. Your own place (apartment, house, condo, etc.) (1)
163. Parent or Guardian's home (2)
164. Another's home (non-parental relative's or non-related adult's home) (3)
165. Group living arrangement (dormitory, barracks, group home, etc.) (4)
166. Homeless (no regular place to stay) (5)
167. Other (please elaborate) (6) ____________________

Q14_TEXT. Where do you live now? That is, where do you stay most often?

[Text entry for: Other (please elaborate)]

Q15. How many people live in your household? Include yourself, your spouse, and any dependents. Do not include your parents or roommates unless you claim them as dependents. (regardless of your living situation, always include yourself as "1")

168. 1 (1)
169. 2 (2)
170. 3 (3)
171. 4 (4)
172. 5 (5)
173. 6 (6)
174. 7 (7)
175. 8 (8)
176. 9 (9)
177. 10 (10)
178. 11 (11)
179. 12 (12)
180. 13 (13)
181. 14 (14)
182. 15 (15)
183. 16 (16)
184. 17 (17)
185. 18 (18)
186. 19 (19)
187. 20 (20)

Q16. Please select the category below that best describes the total amount of INCOME earned last year by the people in YOUR HOUSEHOLD (as "household" is defined in the previous question). Consider all forms of income, including salaries, tips, interest and dividend payments, scholarship support, student loans, parental support, social security, alimony, child support, and others.

188. $15,000 or under (1)
189. $15,001 - $25,000 (2)
190. $25,001 - $35,000 (3)
191. $35,001 - $50,000 (4)
192. $50,001 - $65,000 (5)
193. $65,001 - $80,000 (6)
194. $80,001 - $100,000 (7)
195. $100,001 - $150,000 (8)
196. Over $150,000 (9)
197. Prefer to not answer (10)
198. Don't Know (11)

Q17. Please select the category below that best describes the total amount of INCOME earned last year by YOUR PARENTS. Again, consider all forms of income, including salaries, tips, interest and dividend payments, scholarship support, student loans, parental support, social security, alimony, child support, and others.

199. $15,000 or under (1)
200. $15,001 - $25,000 (2)
201. $25,001 - $35,000 (3)
202. $35,001 - $50,000 (4)
203. $50,001 - $65,000 (5)
204. $65,001 - $80,000 (6)
205. $80,001 - $100,000 (7)
206. $100,001 - $150,000 (8)
207. Over $150,000 (9)
208. Prefer to not answer (10)
209. Don't Know (11)

Q18. Do you work for pay part-time, full-time, or neither?

210. Part-time (1)
211. Full-time (2)
212. Neither (3)

Q19. How much money do you typically spend each day using cash and your debit card (in dollars)?

[Text entry]

Q20. Do you currently smoke cigarettes?

213. Yes (1)
214. No (2)

Q21. Answered if: *Yes* was selected on Q20. If you do smoke cigarettes, approximately how many cigarettes do you smoke per day?

[Text entry]

Q22. 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?

[Text entry]

Q23. If it takes 5 machines 5 minutes to make 5 widgets, how many minutes would it take 100 machines to make 100 widgets?

[Text entry]

Q24. 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 many days would it take for the patch to cover half of the lake?

[Text entry]

Q25_1. How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please select and option on the scale, where 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks'.

	0 (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 (6)	6 (7)	7 (8)	8 (9)	9 (10)	10 (11)	
Not at all willing to take risks	215.	216.	217.	218.	219.	220.	221.	222.	223.	224.	225.	Very willing to take risks

Q26. How would you characterize your religious beliefs? Please select the option that best describes your beliefs.

226. Atheism (1)
227. Buddhism (2)
228. Christianity - Baptist (3)
229. Christianity - Catholic (4)
230. Christianity - Lutheran (5)
231. Christianity - Methodist (6)
232. Christianity - Other (7)
233. Hinduism (8)
234. Islam (9)
235. Judaism (10)
236. Nonreligious or Agnostic (11)
237. Prefer to not answer (12)
238. Other (please elaborate) (13)

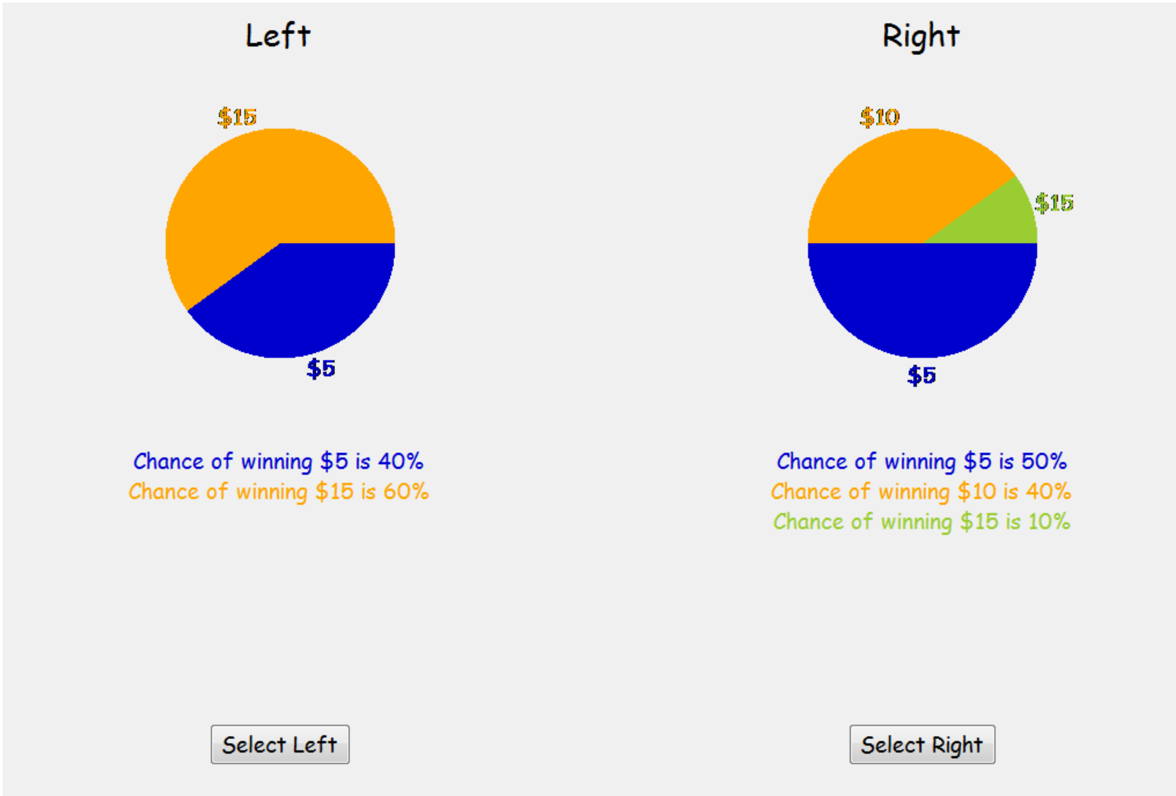
Q26_TEXT. How would you characterize your religious beliefs? Please select the option that best describes your beliefs.

[Text entry for: Other (please elaborate)]

3. Instructions for 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. There are 50 pairs in the series. For each pair of prospects, you should choose the prospect you prefer. You will actually get the chance to play **one** of these prospects, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer.

Here is an example of what the computer display of 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.

In the above example the left prospect pays five dollars ($5) if the number drawn is between 1 and 40, and pays fifteen dollars ($15) if the number is between 41 and 100. The blue color in the pie chart corresponds to 40% of the area and illustrates the chances that the number drawn will be between 1 and 40 and your prize will be $5. The orange area in the pie chart corresponds to 60% of the area and illustrates the chances that the number drawn will be between 41 and 100 and your prize will be $15. When you select the lottery to be played out the computer will confirm what die rolls correspond to the different prizes.

Now look at the pie chart on the right. It pays five dollars ($5) if the number drawn is between 1 and 50, ten dollars ($10) if the number is between 51 and 90, and fifteen dollars ($15) if the number is between 91 and 100. As with the prospect on the left, the pie slices represent the percentage of the possible numbers which yield each payoff. For example, the size of the $15 pie slice is 10% of the total pie, and is thus 10 numbers out of 100.

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

After you have worked through all of the pairs of prospects, raise your hand and an experimenter will come over as soon as they are available. You will then roll two 10-sided dice to determine which pair of prospects will be played out. You roll the die until a number between 1 and 50 comes up. Since there is a chance that any of your 50 choices could be played out for real earnings, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will again roll the two ten-sided dice to determine the outcome of the prospect you chose.

For instance, suppose you picked the prospect on the left in the above example and it was the pair chosen to be played. If the random number from your rolls of the dice was 37, you would win $5; if it was 93, you would win $15. If you picked the prospect on the right and drew the number 37, you would win $5; if it was 93, you would win $15.

Therefore, your payoff is determined by three things:

- **which prospect you selected, the left or the right, for each of these 50 pairs;**
- **which prospect pair is chosen to be played out in the series of 50 pairs using the two 10-sided dice; and**
- **the outcome of that prospect when you roll the two 10-sided dice again.**

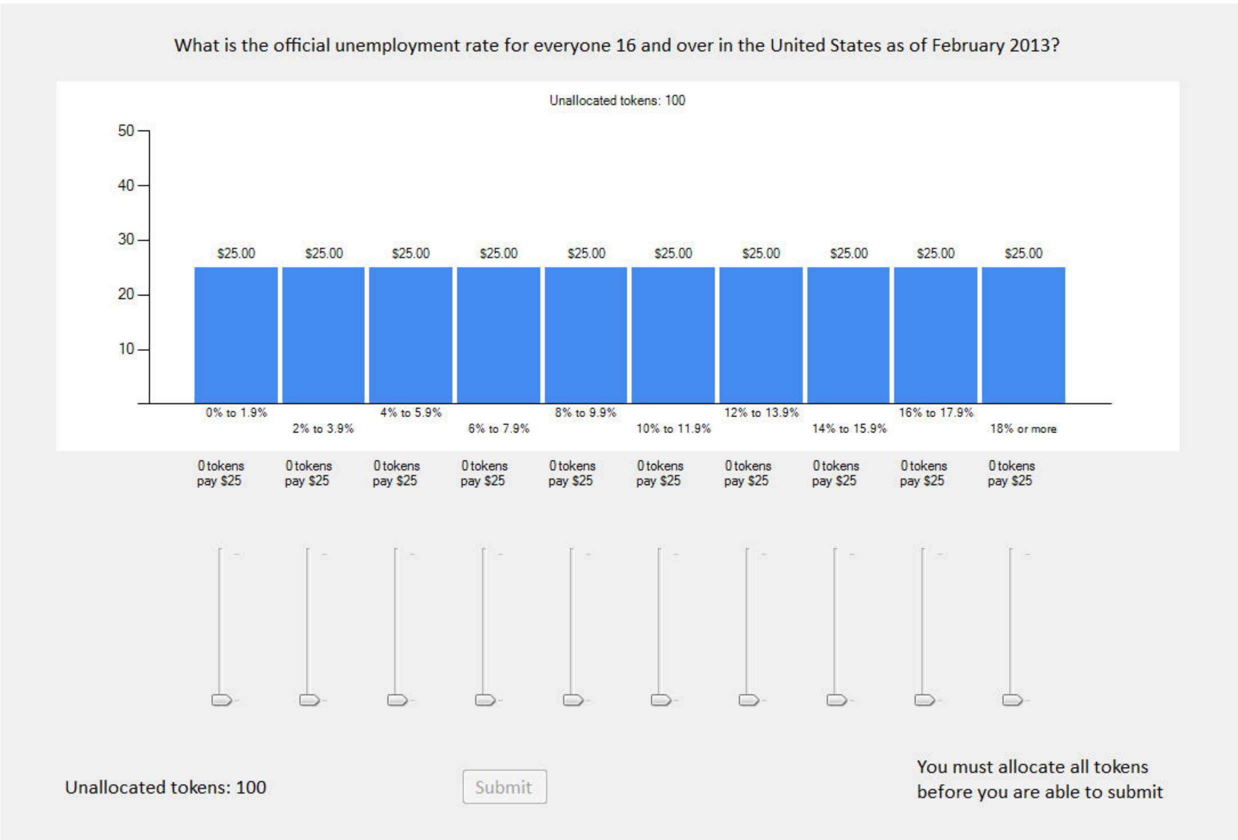
Which prospects you prefer is a matter of personal choice. The people next to you may be presented with different prospects, and may have different preferences, so their responses should not matter to you or influence your decisions. Please work silently, and make your choices by thinking carefully about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

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

4. Instructions for Belief Elicitation

This is a task where you will be paid according to how accurate your beliefs are about certain things. You will be presented with 15 questions and asked to place some bets on your beliefs about the answers to each question. You will actually 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 everyone has made all decisions, and that process is explained below.

Here is an example of what the computer display of a question might look like.



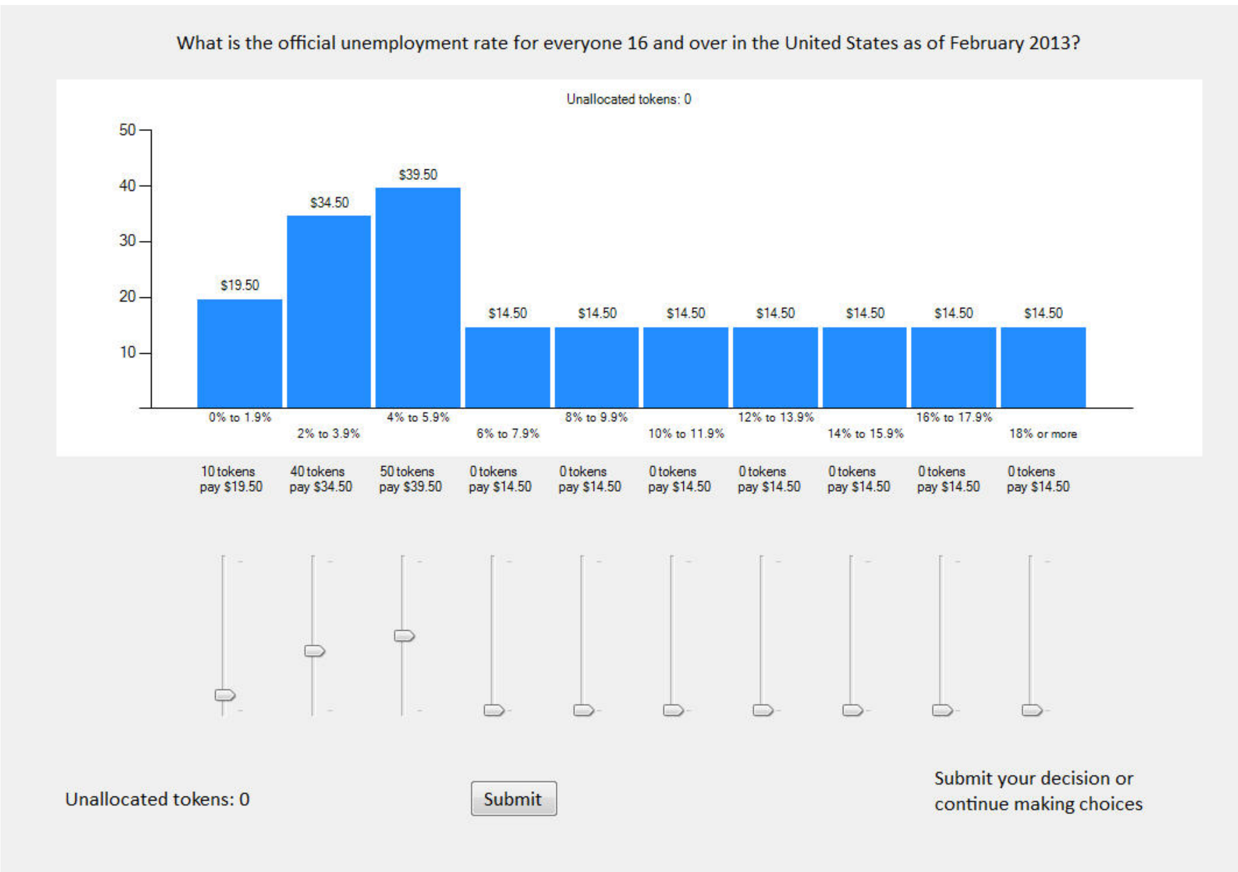
The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate across the sliders. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate all 100 tokens, and in this example we start with 10 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. Please note that the bars above each slider correspond to that particular slider. In the above example, the tokens you allocate to each bar will naturally reflect your beliefs about the official unemployment rate for everyone 16 and over in February 2013. The first bar corresponds to your belief that the unemployment rate is between 0% and 1.9%. The second bar corresponds to your belief that the unemployment rate is between

2% and 3.9%, and so on. Each bar shows the amount of money you could earn if the official unemployment rate is in the interval shown under the bar.

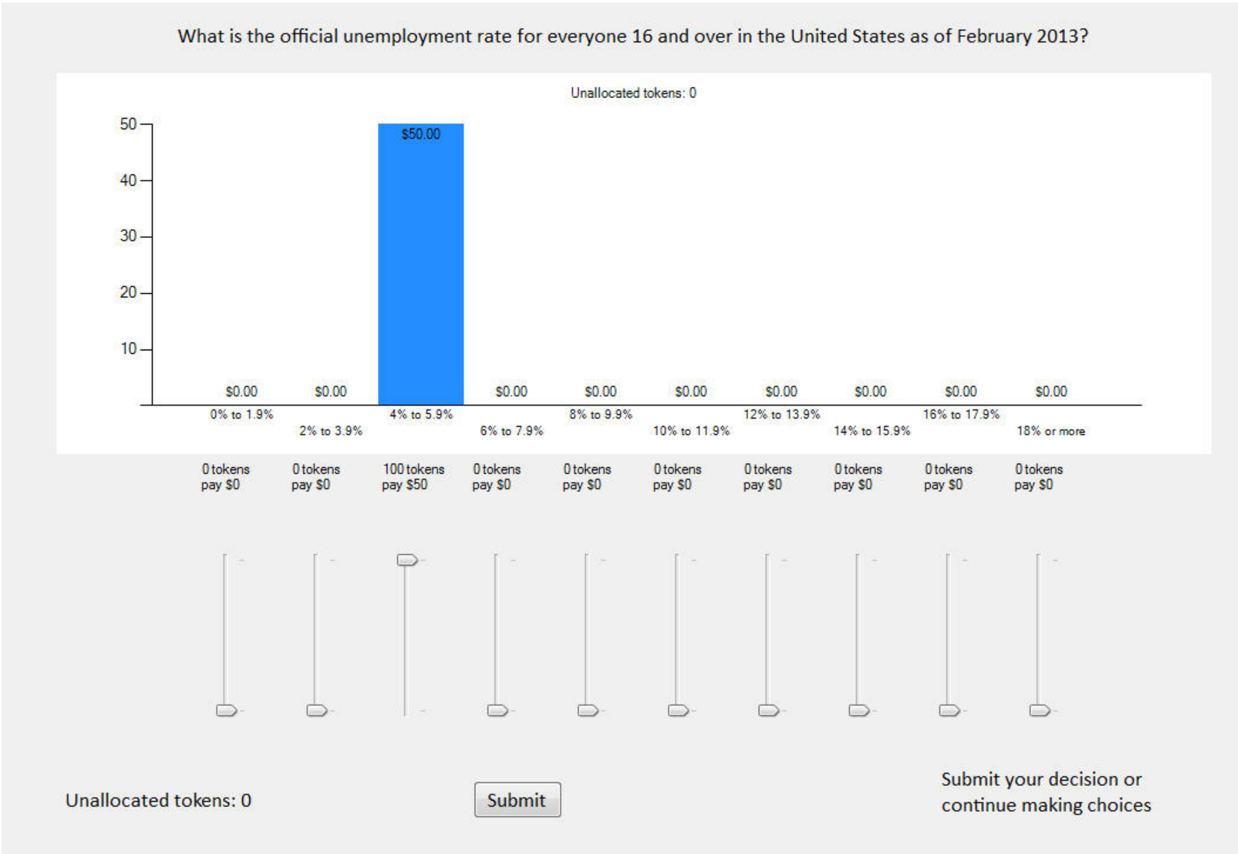
To illustrate how you use these sliders, suppose you think there is a fair chance the unemployment rate is just under 5%. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 4% to 5.9%, 40 tokens to the interval 2% to 3.9%, and 10 tokens to the interval 0% to 1.9%. So you can see in the picture below that if indeed the unemployment rate is between 4% and 5.9% you would earn \$39.50. You would earn less than \$39.50 for any other outcome. You would earn \$34.50 if the unemployment rate is between 2% and 3.9%, \$19.50 if it is between 0% and 1.9%, and for any other unemployment rate you would earn \$14.50.



You can adjust the allocation as much as you want to best reflect your personal beliefs about the unemployment rate.

Your earnings depend on your reported beliefs and, of course, the true answer. For instance, suppose you allocated your tokens as in the figure shown above. The true unemployment rate is actually 7.7%, according to the *Bureau of Labor Statistics*. So if you had reported the beliefs shown above, 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 unemployment rates between 4% and 5.9%. Then you would have faced the earnings outcomes shown below.



Note the "good news" and "bad news" here. If the unemployment rate is indeed between 4% and 5.9%, you earn the maximum payoff, shown here as $50. But the true unemployment rate is 7.7%, so 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:

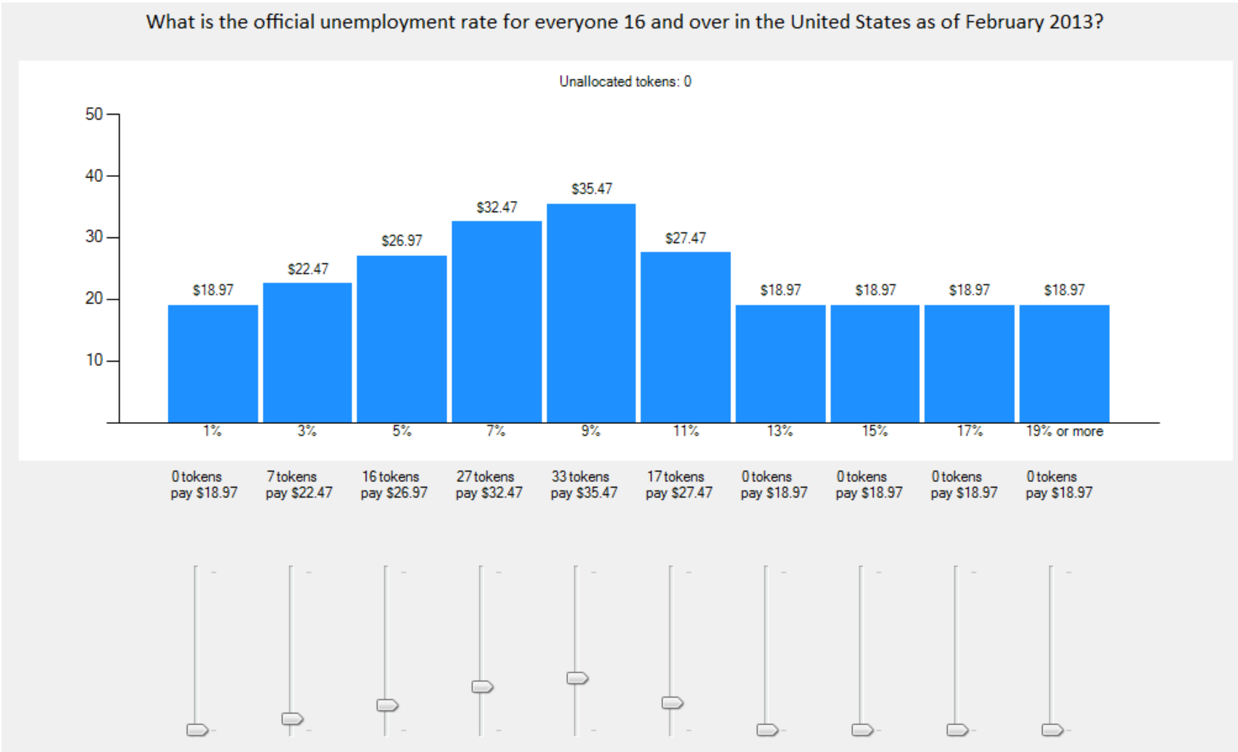
- **Your belief about the correct answer to each question is a personal judgment that depends on the information you have about the topic of the question.**

- **Depending on your choices and the correct answer you can earn up to $50.**

- **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.

For some of the questions we will round the correct answer to the nearest amount shown under each bar. For example, the decision screen for the unemployment question might have shown unemployment rates of 1%, 3%, 5%, 7%, 9%, 11%, 13%, 15%, 17% and 19% or more, as shown below.



In this case, the correct answer of 7.7% would have been rounded to 7% rather than rounded to 9%, and the payment would have been $32.47.

When you are satisfied with its decisions, you should click on the **Submit** button and confirm the choices. When everyone is finished we will come to you and roll a 20-sided die until a number between 1 and 11 comes up to determine which question will be played out. The experimenter will record your individual 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.

Are there any questions?

Appendix D: Class Notes on Beta and Gamma Interval Regression

By Glenn Harrison (with permission)

December 2013

Much of the behavioral data we consider is not obviously best modeled as a Normal distribution. Two alternatives are considered here, building on data that can be better characterized with the Beta distribution or the Gamma distribution.

1. Beta Interval Regression

The normal distribution is not an attractive statistical assumption for data *a priori* constrained to lie in a known interval, such as elicited probabilities. The same point applies to data *a priori* constrained to be greater or less than some known value, such as age; we refer to such data as left-censored intervals or right-censored intervals. It is a relatively simple matter to replace the assumption of a normal distribution with a beta distribution, providing considerable flexibility as well automatically satisfying the constraint to lie in the open unit interval.[31] The beta distribution need not be symmetric, allowing left or right skew towards the endpoints. Brehm and Gates (1993) discuss direct use of the beta distribution for maximum likelihood (ML) regression models, and also point to re-parameterization from King (1998) to facilitate regression analysis in terms of "mean effects" and "variance effects." Paolino (2001, p.336) also offers yet another attractive re-parameterization for beta regression.

The most popular re-parameterization of the beta distribution was offered in the statistics literature by Ferrari and Cribari-Neto (2004). The "original" probability density of the beta distribution can be written

$$f(x, a, b) = \{\Gamma(a+b)/[\,\Gamma(a)\Gamma(b)\,]\}\ x^{a-1}\ (1-x)^{b-1} \qquad \text{(D1)}$$

where $a, b > 0$ and $0 \leq x \leq 1$ for virtually all values of the distributional parameters a and b. Let $F(x, a, b)$ denote the cumulative beta distribution. The proposed re-parameterization has parameters μ and φ, with $\mu = a/(a+b)$ and $\varphi = a+b$, so that $a = \mu\varphi$ and $b = (1-\mu)\varphi$. The advantage of this re-parameterization is that we can directly specify the mean and variance of the dependent variable x as $E(x) = \mu$ and $Var(x) = \mu(1-\mu)/(1+\varphi)$. The latter is a "direct" expression for variance in the sense that for given μ, the variance of x is larger as φ is smaller; so, in the usual statistical parlance, φ can be viewed as a precision parameter. The obvious constraints on these parameters are that $0< \mu<1$ and $\varphi>0$. Hence, given values of μ and φ we can directly evaluate $f(\cdot)$ and $F(\cdot)$.

The log-likelihood for the beta regression model for a vector of observations y with typical element y_i can then be directly specified as

$$L_i(\mu, \varphi \mid y_i) = \ln f(y_i, \mu\varphi, (1-\mu)\varphi) \qquad \text{(D2)}$$

and maximized by numerical methods.[32] *Stata* software to implement this beta regression model is available, and can be obtained by searching within *Stata* for **`betafit`**.

---

[31] Known intervals can be accommodated with trivial re-normalizations to the unit interval.

[32] If the dependent variable was not bounded in the unit interval one would add a "link function" following Ferrari and Cribari-Neto (2004, p.803).

An important extension of the beta regression model allows for the dependent variable to take on values 0 and 1, but with a different data-generating process than is assumed for intermediate values strictly between 0 and 1. This is akin to a "double-hurdle" model, and specifies logit functions for the two endpoints, and the beta model for the open unit interval. It is particularly appropriate for models where one expects a "spike" at 0 or 1, such as portfolio allocation data and applications in empirical finance (e.g., Cook, Kieschnick, and McCullough (2008)). This is also known as the zero-one inflated beta regression model; *Stata* software to implement this model is also available, and can be obtained by searching within *Stata* for **`zoib`**.

We offer a further extension of the beta regression model, to consider data that is observed in interval form. The example of interest, of course, is the response to a belief elicitation task such as considered here, where the dependent variable comes in 10 "bins" that might reflect (uncensored or censored) intervals. Specifically, and following the notation for the "interval regression" model in *Stata*, we have four classes of responses:

1. for "certain" observations $i \in C$ we observe point data y_i;
2. for observations $i \in L$ we observe left-censored data where we only know that the value is less than or equal to some known $y_{L\,i}$;
3. for observations $i \in R$ we observe right-censored data where we only know that the value is greater than or equal to some known $y_{R\,i}$; and
4. for observations $i \in I$ we observe interval data where we only know that the value is in the closed interval $[y_{1i}, y_{2i}]$.

The log-likelihood can then be expressed in general form as

$$L(\mu, \varphi \mid y) = \sum_{i \in C} \ln\{\Pr(y_i)\} + \sum_{i \in L} \ln\{\Pr(y_i \leq y_{L\,i})\} + \sum_{i \in R} \ln\{\Pr(y_i \geq y_{R\,i})\} + \sum_{i \in I} \ln\{\Pr(y_{1i} \leq y_i \leq y_{2i})\} \qquad \text{(D3)}$$

where $\Pr(\cdot)$ denotes probability. Each of the components can be expressed in terms of the probability density and cumulative density of the beta distribution:

5. $\Pr(y_i)$ is just $f(y_i, \mu\varphi, (1-\mu)\varphi)$ from (D2);
6. $\Pr(y_i \leq y_{L\,i})$ is $F(y_{L\,i}, \mu\varphi, (1-\mu)\varphi)$;
7. $\Pr(y_i \geq y_{R\,i})$ is $1-F(y_{R\,i}, \mu\varphi, (1-\mu)\varphi)$; and
8. $\Pr(y_{1i} \leq y_i \leq y_{2i})$ is $F(y_{2\,i}, \mu\varphi, (1-\mu)\varphi) - F(y_{1\,i}, \mu\varphi, (1-\mu)\varphi)$.

The density in (D2) and corresponding cumulative distributions may be directly evaluated with the intrinsic functions **`betaden`** and **`ibeta`** in *Stata*. Thus it is a relatively simple matter to specify the likelihood for an interval beta regression model, following the documentation in (Harrison and Rutström 2008).

2. Gamma Interval Regression

Even when the distribution of the variable being modeled can take on values between $\pm\infty$, one might want to consider non-normal specifications of the interval regression model. This is attractive when there are no covariates that might render the error distribution normal, which occurs when we are simply using the interval regression model to estimate the mean and variance of the distribution. It might also be attractive when we do not want to check in every case that the error distribution, after allowing for non-normal covariates, is appropriately distributed. For these reasons we extend the beta interval regression model to allow for the gamma distribution.

There are several "gamma distributions" in common use, and one has to be careful to specify which one. We use the simple gamma distribution with two shape parameters a and b, and one location-shift parameter g, following *Stata* (version 13):

$$f(x, a, b, g) = \{1/[\, \Gamma(a)b^a\,]\}\ (x\text{-}g)^{a-1}\exp(-(x\text{-}g)/b) \qquad \text{(D4)}$$

where a > 0, b > 0 and $-\infty < g < \infty$. The mean of this density is ab-g, and the variance is ab^2. It is possible to come up with re-parameterizations so that one can directly specific the mean as a linear function of covariates, but then the variance is difficult to tease apart in the same manner. In this instance we prefer to just use the "raw" distribution and generate estimates of mean and standard deviation by non-linear combinations of all estimated parameters.[33] Using (D4) instead of (D1) in (D3) leads to a log-likelihood for the gamma interval regression model; the density function (D4) may be directly evaluated with the intrinsic function **`gammaden`** in *Stata*, and the corresponding cumulative distribution function evaluated with the function **`gammap`**.

An obvious extension of some value would be to use the Generalized Gamma distribution, as illustrated in health economics by Manning, Basu, and Mullahy (2005). This distribution adds an extra shape parameter, and collapses in convenient form to the standard gamma (D4), the Weibull distribution, the Exponential distribution, and Log-normal distribution. Thus it provides a way to check for the "best" distribution across a popular range of alternatives.

3. *Stata* Implementation

The program listed below implements the standard interval regression model with normal errors, the beta interval regression model, and the gamma interval regression model. It uses data available online, to replicate published examples, and also employs user-written routines to verify special cases with pre-existing code. These may be accessed online from *Stata* in the usual manner. A machine-readable version of this program is available on request.

```
* ML interval regression demonstrates interval regression with normal, gamma and beta

* log text file
log using "ML interval regression.log", replace

* initializations
version 13.1
about
set more off

* flag for internet access
global internet "y"

* get the data for tests
if "$internet" == "y" {

        * get Stata data for the manual example for -intreg-
        use http://www.stata-press.com/data/r13/womenwage
        by wagecat: keep if _n==1
        generate wage1 = wagecat[_n-1]
        keep wagecat wage1
        save lagwage, replace
        use http://www.stata-press.com/data/r13/womenwage
        merge m:1 wagecat using lagwage
        erase lagwage.dta
```

[33] This comes at a minor numerical cost: one then has to include each covariate in all three parameters.

```
        generate wage2 = wagecat
        replace wage2 = . if wagecat == 51
        sort age, stable
        save stata_intreg_data1, replace

        * get data to test beta regression
        use http://fmwww.bc.edu/repec/bocode/c/citybudget.dta, clear
        save stata_intreg_data2, replace

}

* read in data for -intreg- test
use stata_intreg_data1, clear

* replicate the Stata command
intreg wage1 wage2 age c.age#c.age nev_mar rural school tenure

* define the interval regression commands
program define MLintreg_normal

        args lnf mu LNsigma

        tempvar sigma vlo vhi type zlo zhi

        quietly {

                * transform sigma
                generate double `sigma’ = exp(`LNsigma’)

                * read in the data
                generate double `vlo’ = $ML_y1
                generate double `vhi’ = $ML_y2

                * determine the type of censored data
                generate int `type’ = .
                replace `type’ = 0 if `vlo’ == `vhi’
                replace `type’ = 1 if `vlo’ == . & `vhi’ ~= .
                replace `type’ = 2 if `vlo’ ~= . & `vhi’ == .
                replace `type’ = 3 if `vlo’ ~= . & `vhi’ ~= . & `vhi’ > `vlo’

                * get the standardized normals for the intervals
                generate double `zlo’ = (`vlo’ - `mu’)/`sigma’
                generate double `zhi’ = (`vhi’ - `mu’)/`sigma’

                * get the likelihood contributions
                replace `lnf’ = ln(normalden(`vlo’,`mu’,`sigma’)) if `type’ == 0
                replace `lnf’ = ln(normal(`zhi’)) if `type’ == 1
                replace `lnf’ = ln(1-normal(`zlo’)) if `type’ == 2
                replace `lnf’ = ln(normal(`zhi’)-normal(`zlo’)) if `type’ == 3

        }

end

* define the interval regression commands
program define MLintreg_beta

        args lnf muK LNphi

        tempvar mu phi a b vlo vhi type

        quietly {

                * transform parameters
                generate double `mu’ = 1/(1+exp(`muK’))
                generate double `phi’ = exp(`LNphi’)
                generate double `a’ = `mu’*`phi’
                generate double `b’ = (1-`mu’)*`phi’

                * read in the data
                generate double `vlo’ = $ML_y1
                generate double `vhi’ = $ML_y2

                * determine the type of censored data
```

```
                generate int `type' = .
                replace `type' = 0 if `vlo' == `vhi'
                replace `type' = 1 if `vlo' == . & `vhi' ~= .
                replace `type' = 2 if `vlo' ~= . & `vhi' == .
                replace `type' = 3 if `vlo' ~= . & `vhi' ~= . & `vhi' > `vlo'

                * get the likelihood contributions
                replace `lnf' = ln(betaden(`a',`b',`vlo')) if `type' == 0

*               code from -betafit-
*               replace `lnf' = lngamma(exp(`LNphi')) - lngamma(invlogit(`muK')*exp(`LNphi')) - ///
                                lngamma((1-invlogit(`muK'))*exp(`LNphi')) +                     ///
                                (invlogit(`muK')*exp(`LNphi')-1)*ln(`vlo') +                    ///
                                ((1-invlogit(`muK'))*exp(`LNphi')-1)*ln(1-`vlo') if `type' == 0

                replace `lnf' = ln(ibeta(`a',`b',`vhi')) if `type' == 1
                replace `lnf' = ln(1-ibeta(`a',`b',`vlo')) if `type' == 2
                replace `lnf' = ln(ibeta(`a',`b',`vhi')-ibeta(`a',`b',`vlo')) if `type' == 3

        }

end

* define the interval regression commands
program define MLintreg_gamma

        args lnf LNa LNb g

        tempvar a b vlo vhi type

        quietly {

                * transform parameters
                generate double `a' = exp(`LNa')
                generate double `b' = exp(`LNb')

                * read in the data
                generate double `vlo' = $ML_y1
                generate double `vhi' = $ML_y2

                * determine the type of censored data
                generate int `type' = .
                replace `type' = 0 if `vlo' == `vhi'
                replace `type' = 1 if `vlo' == . & `vhi' ~= .
                replace `type' = 2 if `vlo' ~= . & `vhi' == .
                replace `type' = 3 if `vlo' ~= . & `vhi' ~= . & `vhi' > `vlo'

                * get the likelihood contributions
                replace `lnf' = ln(gammaden(`a',`b',`g',`vlo')) if `type' == 0
                replace `lnf' = ln(gammap(`a',(`vhi'-`g')/`b')) if `type' == 1
                replace `lnf' = ln(1-gammap(`a',(`vlo'-`g')/`b')) if `type' == 2
                replace `lnf' = ln(gammap(`a',(`vhi'-`g')/`b')-gammap(`a',(`vlo'-`g')/`b'))     ///
                                if `type' == 3

        }

end

* replicate the Stata command with no covariates
intreg wage1 wage2
local mu = [model]_cons
local lns = [lnsigma]_cons

* see if this replicates Stata with no covariates
ml model lf MLintreg_normal (mu: wage1 wage2 = ) (LNsigma: ), missing maximize difficult     ///
                                                              init(`mu' `lns', copy)
ml display
nlcom (sigma: exp([LNsigma]_cons))

* see if this replicates Stata with covariates
ml model lf MLintreg_normal (mu: wage1 wage2 = age c.age#c.age nev_mar rural school tenure ) ///
                            (LNsigma: ), missing maximize
ml display
```

```
* now gamma with no covariates
ml model lf MLintreg_gamma (LNa: wage1 wage2 = ) (LNb: ) (g: ), missing maximize difficult
ml display
nlcom (a: exp([LNa]_cons)) (b: exp([LNa]_cons)) (g: [g]_cons)                             ///
       mean: (exp([LNa]_cons)*exp([LNb]_cons)) + [g]_cons)                                ///
      (sigma: sqrt(exp([LNa]_cons)*exp([LNb]_cons)^2))

* gamma with covariates: need to have covariates in all three parameters with this "raw"
parameterization
ml model lf MLintreg_gamma (LNa: wage1 wage2 = age c.age#c.age nev_mar rural school tenure ) ///
                           (LNb: age c.age#c.age nev_mar rural school tenure )              ///
                           (g: age c.age#c.age nev_mar rural school tenure),                ///
                            missing maximize difficult
ml display
nlcom (a: exp([LNa]_cons)) (b: exp([LNa]_cons)) (g: [g]_cons)                             ///
      (mean: (exp([LNa]_cons)*exp([LNb]_cons)) + [g]_cons)                                ///
      (sigma: sqrt(exp([LNa]_cons)*exp([LNb]_cons)^2))

* read in data for beta interval model test
use stata_intreg_data2, clear

* drop missing LHS since interval regression needs the "missing" option
drop if governing==.

* replicate with covariates and without
betafit governing, mu(minorityleft noleft houseval popdens) alternative
betafit governing, alternative
local mu = [mu]_cons
local LNphi = [ln_phi]_cons

* test without covariates and then with covariates
ml model lf MLintreg_beta (muK: governing governing = ) (LNphi: ), missing maximize        ///
                                 difficult technique(nr) init(`mu' `LNphi', copy)
ml display
nlcom (mu: 1/(1+exp([muK]_cons))) (phi: exp([LNphi]_cons))

ml model lf MLintreg_beta (muK: governing governing = minorityleft noleft houseval popdens ) ///
                          (LNphi: ), missing maximize difficult technique(nr) continue
ml display
nlcom (mu: 1/(1+exp([muK]_cons))) (phi: exp([LNphi]_cons))
```

Since the parameterization of the gamma interval regression model does not directly generate estimates of the mean and standard deviation of the latent index, it may be useful to show how one can extend the application of the "delta method" to correctly calculate the marginal effects of covariates on the mean and standard deviation. In the pseudo-code below the variables x1 and x2 are covariates of some previously estimated model, and the first part of the expression for the marginal effect is the total effect of the covariate (if that is desired):

```
* first calculate the mean and standard deviation for the constant
nlcom (mean: (exp([LNa]_cons)*exp([LNb]_cons)) + [g]_cons)
nlcom (sigma: sqrt(exp([LNa]_cons)*exp([LNb]_cons)^2))

* now the marginal effects of the covariates on the mean and standard deviation
foreach v in x1 x2 {
    nlcom (mean_`v': ((exp([LNa]_cons+[LNa]`v')*exp([LNb]_cons+[LNb]`v'))+([g]_cons+[g]`v'))   ///
                   - ((exp([LNa]_cons)*exp([LNb]_cons)) + [g]_cons))
    nlcom (sigma_`v': (sqrt(exp([LNa]_cons+[LNa]`v')*exp([LNb]_cons+[LNb]`v')^2))              ///
                   -  (sqrt(exp([LNa]_cons)*exp([LNb]_cons)^2)) )
}
```

Chapter 3 - On Measuring Financial Literacy: Applications of Subjective Belief Elicitations and Extended Scaffolds

3.1 Introduction

This chapter looks at "extended" literacy, which is, literacy that is partly achieved and maintained through access to scaffolds. Results will be presented for beliefs over questions considered in Chapter 2. The key treatment here is that subjects have access to the Internet, and can be compared with the control group described in Chapter 2.

3.2 Literature Review

There is an obvious sense in which access to the Internet is a scaffold. But this particular treatment allows one to consider an aspect of scaffolds that is often neglected: that the value of one scaffold can depend on the existence of another scaffold. Consider roads. They are a wonderful scaffold for driving if we have cars, traffic lights and lane markings. But take away traffic lights and lane markings and roads can become a very poor scaffold.[34] Similarly, access to the Internet without the ability to structure the query correctly can make the Internet an unreliable scaffold. Internet search engines will always spit out some answer, but if the question is poorly posed, that may generate a false sense of confidence. An immediate example is the longevity risk question about how many more years a male will live in the United States, conditional on reaching the age of 20. If that question is posed to the Internet as "how long will a man live in the United States?" then the answer might be reliable if someone then deducts 20 from it. If the direct response from the Internet is used, there will be evidence of significant bias, overstating how long men are expected to live conditional on reaching the age of 20. Note that this bias arises from using the scaffold incorrectly.

[34] Proof: me driving in Dar es Salaam, Tanzania. Of course this is much more of a problem for me than for the Tanzanians. There they know how to use alternative scaffolding, a range of signaling conventions, that I do not.

3.3 Experimental Design

The data presented in Chapters 2, 3, and 4 are obtained in the same manner from experiments conducted with students at Georgia State University (GSU) in the Experimental Economics Center (ExCEN) over the period from November 2013 to March 2016. See Section 2.4 for a full description of the data.

The results presented in Chapter 3 build on Chapter 2 and assess the effect that access to the Internet has on an individual's financial literacy. There were 5 Internet sessions conducted in total: 2 sessions using the initial labels and 3 sessions using the new labels. Again we focus our analysis on the 3 Internet sessions using the new labeling scheme to be able to compare the results directly with our control treatment of individuals that did not have access to the Internet presented in Chapter 2.

The sessions involving Internet access were administered in the same manner as the sessions without Internet access. The one variation in protocol was that the participants were given a sheet of paper that listed the 11 financial questions they could use the Internet to research. Once all the participants had the list of questions, an Internet browser was remotely opened on each computer, and they were allowed access to the Internet for 15 minutes. After the 15 minutes, the browsers on all the computers closed automatically. The participants were instructed that during the 15 minutes of Internet access they should research the section's questions as efficiently as possible, and make notes as needed on the paper provided. Immediately after the Internet access ended the participants recorded their responses using the computer interface. The pertinent text that was updated in the experimental instructions for participants allowed access to the Internet is as follows:

> Using the methodology described above we will ask you 11 questions today. Immediately before each section starting, you will be given that section's questions on a piece of paper. You can write on these papers as needed. After everyone in the experiment has the appropriate question sheet in hand we will then open an Internet browser on your computer remotely, and allow access to the Internet for 15 minutes. It is a timed 15 minutes, after which the browsers on your computer will close automatically. It is during this time that you should research the section's questions as efficiently as possible, and make notes as needed on the paper provided. Immediately after the

Internet access ends, you will be asked to respond to the same questions using the computer interface.

It is up to you to use your given time on the Internet as efficiently as possible to research the questions that will be handed out in advance. You should balance the information you find on the Internet with the strength of your personal beliefs and the possibility of them being wrong. There are three important points for you to keep in mind when making your decisions:

- Your belief about the correct answer to each question is a personal judgment after being given access to the Internet, and that depends on the information you have about the topic of the question.
- Depending on your choices and the correct answer you can earn up to $50.
- 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.

The complete experimental instruction sets for the Internet treatments are available in Appendix E.

There are two Internet treatments that differ only by the inclusion of a cautionary statement in the instruction set text that states "You are not required to use any information found on the Internet." The caveat was constructed to test the significance of a reminder to participants that they are not obliged to use the information found online. Of the three Internet sessions using the new labeling scheme, Session 17 did not have the caveat, and Sessions 20 and 25 had the caveat. The total number of participants for Sessions 17, 20, and 25 are 33, 32, and 39, respectively, thus totaling 104 participants who had access to the Internet.

3.4 Results

The results from these experiments are straightforward. In general, we find that access to the Internet increases the literacy of individuals and has a large effect by boosting confidence in beliefs, which leads to reduction in the standard deviation of responses. Access to the Internet formally improves literacy by our measures in eight of the eleven questions and does so in a dramatic way for several: the literacy measures of the other three questions stay within a percentage point or two compared to the literacy measures without Internet access. It appears that the Internet is being utilized successfully as an informational complement, as one might have expected from the metaphor of a cognitive production function: the internet is substituting for other intrinsic inputs. Of course, substitution can be a good thing if the expanded production input enhances output at no (major) extra cost, which is the case as we see here since there was no cost for the subjects to access the Internet.

As with previous results we begin the initial analysis going through a detailed example to begin. Consider responses to the question about the savings account with 2% interest left to grow over five years shown in Figure 3.1 and Table 3.1, Table 3.2, and Table 3.3. To anticipate a pattern in evaluating these responses, initially we examine the distribution of responses in Figure 3.1, then observe the *marginal* effect of being assigned to an Internet treatment (Table 3.1), then examine the *marginal* effect of being in an Internet treatment that was additionally given a caveat (Table 3.2), and finally examine the *total* effect of each covariate with respect to bias away from the true answer and imprecision of beliefs.

Beginning with Figure 3.1 we see that the overall average belief for individuals was $110.70 and for participants with access to the Internet was $109.80, and a reported p-value = 0.060 on a hypothesis test if there is a difference in average responses. Here we *cannot* reject the hypothesis that the average response of individuals and those with access to the Internet are statistically different from one another at the 5% significance level. We further observe from Figure 3.1 that the modal responses for both individuals and those with Internet access are in the bin that contains the true answer and that the underlying distributions are shaped similarly. Participants with Internet access appear to allocate more tokens on average to the bin containing the true answer than individuals. We can read off the average token allocations directly from the figure, or see a numeric representation of them using the L literacy index shown at the top of Table 3.3. Thus the reported L literacy indices of 0.53 for individuals and 0.68 for subjects given access to the Internet correspond to an average token allocation of 53 and 68, respectively, in the middle bin containing the correct answer of $110.41 in Figure 3.1.

Turning to specific statistical findings, we will focus on the effect of the treatment, and not so much the demographics. In the top panel of Table 3.1 we see that the overall average belief was $111.51; as usual, the constant refers to the "omitted category," which is a White male underclassman at GSU. We observe that those having Internet access have an average reduction of \$0.53 in their belief when compared to the omitted category, but that this is not statistically significantly different from zero (p-value = 0.06) at the 5% significance level. In the bottom panel of Table 3.1, the interpretation of the coefficients for dispersion is immediate, because the dependent variable is in natural log units. Thus the coefficient estimate β implies a $\beta \times 100$ percent change in the standard deviation of beliefs, σ, for a one unit change in that covariate. So the statistically insignificant effect (p-value = 0.22) of having access to the Internet is interpreted as a 15% reduction in the standard deviation of beliefs compared to the omitted category. The information just below the title in Table 3.1 is generally self-explanatory. The p-value for Internet refers to a test of the hypothesis that the effect on the average *and* the standard deviation are jointly zero, and complements the corresponding hypothesis tests of the average and the standard deviation in the body of the table. Here the overall effect of having access to the Internet is not statistically significant and different from zero (p-value < 0.110). Figure 3.1 shows the distribution of reports for each bin, side-by-side for individuals and those with Internet, allowing a visual interpretation of differences in reports.

Table 3.2 is analyzed in the exact same manner as Table 3.1, thereby teasing apart results for whether a caveat was announced from those in an Internet treatment. Recall that this caveat was simply telling subjects that they didn't have to answer any question with the information they found online. Here we see that the marginal effect of giving subjects access to the Internet without a caveat (internetNC) leads to a statistically significant reduction of \$0.75 from the average report (p-value = 0.03) and a 38% reduction in the standard deviation of beliefs (p-value = 0.05) compared to the omitted category. There is *no* statistically significant effect for subjects given access to the Internet with a caveat (internetC) compared to the omitted category. Finally, the text at the top of the table shows that the overall effect of being in an Internet treatment with and without a caveat is statistically significant and different from zero (p-value < 0.015): thus there is evidence that the underlying distribution of reports is different for those given a caveat versus no caveat. Of course, these are estimates of the effects of treatments on averages and standard deviations, and form the basis for evaluating bias and precision.

Table 3.3 provides an evaluation of the bias and imprecision of beliefs that are associated with demographics and having access to the Internet (with and without caveat), and is different from the estimates underlying Table 3.2. In the top panel of Table 3.3 we estimate the "total effect" of the covariate, and then compare the estimated average belief to the correct value. The bottom panel of Table 3.3 compares the estimated standard deviation for each covariate to the overall pooled standard deviation. Thus we measure bias for the average, and relative imprecision for the standard deviation. These are not marginal effects: for each covariate of interest it is as if we use the average value of all other covariates for that covariate of interest. For students with reported high grade point averages, for instance, the top panel of Table 3.3 shows the difference between the average belief of students with high grades is \$0.39 less than the correct answer of \$110.41, and this difference is statistically significantly different from zero (*p*-value = 0.04). This is different from Table 3.2, where the coefficient shows the difference compared to the omitted category.

With this as background, the top panel of Table 3.3 shows that there is statistically significant evidence of bias from the correct answer for those with Internet access not given the caveat (*p*-value = 0.02) and no evidence of bias for those with Internet access and given the caveat (*p*-value = 0.12). The bottom panel of Table 3.3 shows us that neither Internet treatment exhibits less imprecision of beliefs compared to the average imprecision. For this question we see that "most GSU students get it and those with Internet access exhibit slightly higher literacy." This is also shown by the higher levels of the literacy indexes L and W with access to the Internet compared to individuals (Table 3.3).

Responses to the question about the starting age of Social Security are evaluated in Figure 3.2 and Table 3.4, Table 3.5, and Table 3.6. Again it is instructive to first examine the displays in Figure 3.2. The correct answer is 62 and located in the first bin in Figure 3.2. The beliefs of individuals assigned to the control treatment exhibit low bias and low confidence represented by a diffuse token allocation with a modal response that is, just barely, at the bin containing the correct answer. Individuals in the control treatment allocated on average only 26 tokens to the correct bin. By striking contrast, the subjects in the Internet treatment exhibit low bias and high confidence, allocating on average 87 tokens to the correct bin. We see that the effect of having access to the Internet is statistically significant (p-value < 0.001) and leads to a 92% average reduction in the standard deviation of beliefs when compared to the omitted category (Table 3.4). The effect of having access to the Internet persists even when controlling for the caveat: the average reduction in the standard deviation of beliefs strengthens to 139% for those not given the caveat compared to the omitted category (Table 3.5). In Table 3.6 we see evidence of bias from the correct answer from the Internet treatment with the caveat (p-value < 0.001), but not from the Internet treatment with no caveat (p-value < 0.010). It is likely that the participants who heard the cautionary message when given access to the Internet reserved a small amount of their tokens and hedged that allocation to the bins immediately adjacent to the correct bin as "insurance" for some payout, which would explain the bias observed. Overall we see a sharp improvement in literacy with access to the Internet, as measured by the two indices (Table 3.6).

Responses to the question about the age for Medicare eligibility, another important signpost for retirement planning, are evaluated in Figure 3.3 and Table 3.7, Table 3.8, and Table 3.9. We see the same pattern here as observed for the previous question: diffuse beliefs from the individual subjects that had no access to the internet, and a strikingly unimodal response in the bin containing the correct answer from the subjects that did have access to the Internet (Figure 3.3). The correct answer is 65, and we see that the effect of having access to the Internet is a decrease in bias and increase in confidence when evaluating the two panels of Figure 3.3. The results show a 93% average reduction in the standard deviation of beliefs when compared to the omitted category (Table 3.7) and a greater than 500% improvement in literacy with access to the Internet, as measured by either of our two indices (Table 3.9).

Responses to the financial literacy question about the real interest rate are reported in Figure 3.4 and Table 3.10, Table 3.11, and Table 3.12. In this case we observe in Figure 3.4 that the underlying distribution of beliefs looks similar when comparing participants with and without Internet access. The correct answer, \$98.98, is the modal response in both cases, and confidence is low in each treatment, as shown by the diffuse allocation of beliefs. This is the first response that having access to the Internet (Table 3.10), with or without a caveat (Table 3.11), which does *not* statistically significantly alter beliefs compared to the omitted category. In this instance literacy, using our two measures (Table 3.12), remained at similar levels regardless of access to the Internet. Subjects exposed to the Internet are biased from the correct answer in a statistically significant manner by around +\$1.59 (Table 3.12) in total. There is no significant difference in confidence with access to the Internet.

Responses to the savings horizons are reported in Figure 3.5 and Table 3.13, Table 3.14, and Table 3.15. From Figure 3.5 we observe generally high literacy for individuals with and without Internet access and an underlying belief distribution shaped similarly for each treatment. In each case there is a modal response in the bin containing the correct response of 4 months. Overall responses from participants having access to the Internet are not statistically significantly different than the average response from White male underclassmen (Table 3.13). However, there is a statistically significant difference when comparing the no caveat Internet treatment with the omitted category (Table 3.14). Literacy is modestly improved, and we observe the L index improving by 7 percentage points to a level of 0.84 with access to the Internet (Table 3.15).

Responses to the liability for a stolen credit card are reported in Figure 3.6 and Table 3.16, Table 3.17, and Table 3.18. In the left panel of Figure 3.6 we see low confidence in the distribution of beliefs for individuals without access to the Internet and an incorrect modal response of \$0, which is biased from the correct answer of \$50. In the right panel of Figure 3.6 we see a more confident distribution of beliefs for participants given access to the Internet, with some weight being placed at \$0, but a large spike on the bin that contained the correct answer. The overall effect of having access to the Internet was statistically significant (p-value < 0.001) and led to a 73% average reduction in the standard deviation of beliefs when compared to the omitted category (Table 3.16). Literacy with access to the Internet shows a sharp improvement over no access, increasing dramatically whether measured by either of the two indices (Table 3.18).

Responses to the liability for a stolen debit card are reported in Figure 3.7, Table 3.19, Table 3.20, and Table 3.21, that convey results very similar to the previous question. We again see less confidence for participants without access to the Internet and an incorrect modal response of $0 compared to those with Internet access having more confidence and the correct modal response of $500 (Figure 3.7). The overall effect of having access to the Internet is statistically significant (p-value < 0.001) and led to a 102% average reduction in the standard deviation of beliefs when compared to the omitted category (Table 3.19). There is no statistical evidence of bias from the correct answer for either Internet treatment and we again see a sharp improvement in literacy as measured by either of the two indices (Table 3.21).

Responses to the question about the nominal interest rate are documented in Figure 3.8, Table 3.22, Table 3.23, and Table 3.24. In this case we observe that the underlying distribution of beliefs looks strikingly similar between participants with and without Internet access (Figure 3.8). The correct answer, 5%, is the modal response in both cases, and confidence is high and bias appears low in each treatment. As the literacy is high to begin with we find that having access to the Internet (Table 3.22), with or without a caveat (Table 3.23), does not statistically significantly alter beliefs compared to the omitted category. In this instance literacy, using the two measures (Table 3.24), remained at similar levels regardless of access to the Internet. This is a question that "GSU students get right" regardless of having access to the Internet.

Responses to the question about the simple interest rate are documented in Figure 3.9, Table 3.25, Table 3.26, and Table 3.27, and are similar to the previous question. Again the distribution of beliefs is similar for participants with and without access to the Internet, with subjects in each treatment holding a high degree of confidence around the correct answer of $102 as their modal response (Figure 3.9). This is also a question that "GSU students get right," and literacy remained at similar levels regardless of access to the Internet (Table 3.27).

The final two questions were about beliefs concerning longevity risk for men and women. Figure 3.10 and Figure 3.11 show that the effect of Internet access, compared to individuals without access, is a sharp increase in confidence and an increase in modal responses that are in the bin containing the correct answer. The overall effect of access to the Internet is statistically significant for the questions regarding the remaining life of men (p-value = 0.007, Table 3.28) and remaining life of women (p-value < 0.001, Table 3.31) when compared to White male underclassmen. Responses for both questions in either Internet treatment were negatively biased from the correct answer, and these effects are statistically significant at a 1% significance level (Table 3.30 and Table 3.33). However, literacy, as measured by the two indices (Table 3.30 and Table 3.33), increased around two-fold with access to the Internet for both questions since there was greater bias without access to the Internet.

3.5 Conclusions

Table 3.34 is an extension of Table 2.39 which displays a summary view of the financial literacy questions and the associated pooled L and W indices for each question for both the Individual control subjects introduced in Chapter 2 and the subjects that were given access to the Internet as a scaffold in this chapter. Table 3.35 is like Table 3.34, but focuses only on the L index and the changes to it when compared to the Individual literacy measures. In the fourth column of Table 3.35 are the pooled measures of the L index for individuals responding with only their private literacy. In the sixth column of Table 3.35 are the L index measures for the Internet. Column seven is the "+/-" difference between the L index for those with Internet access compared to individuals without. We see "+/-" values for the Internet treatment that range from -0.02 for fin14, "the interest rate" question, all the way to +0.66 for fin3, "the Medicare eligibility" question.

In Table 3.35 we observe enhanced literacy directly in eight of our eleven financial questions when participants are given access to the Internet. We see that the largest enhancements to literacy from the Internet scaffold with procedural and longevity risk questions, thus suggesting the importance of access to, and comprehension of, sources of information found online. The three questions (fin7, fin13, and fin14) for which we observed either no change, or a very small change in literacy, given access to the Internet are numeracy questions dealing with interest. It could be that the participants chose to calculate these directly "in their head" instead of searching the Internet for an answer.

Overall there is clear evidence that access to the Internet served as an effective and constructive scaffold with respect to the revealed literacy of individuals.

Figure 3.1: Elicited Beliefs about
The Savings Account with 2% Interest Question,
Comparing Internet and Individual Responses, New Labels

Internet average: \$109.8 (*N*=104)   Individual average: \$110.7 (*N*=106)
p-value on test of difference in averages = 0.060
True response was \$110.41

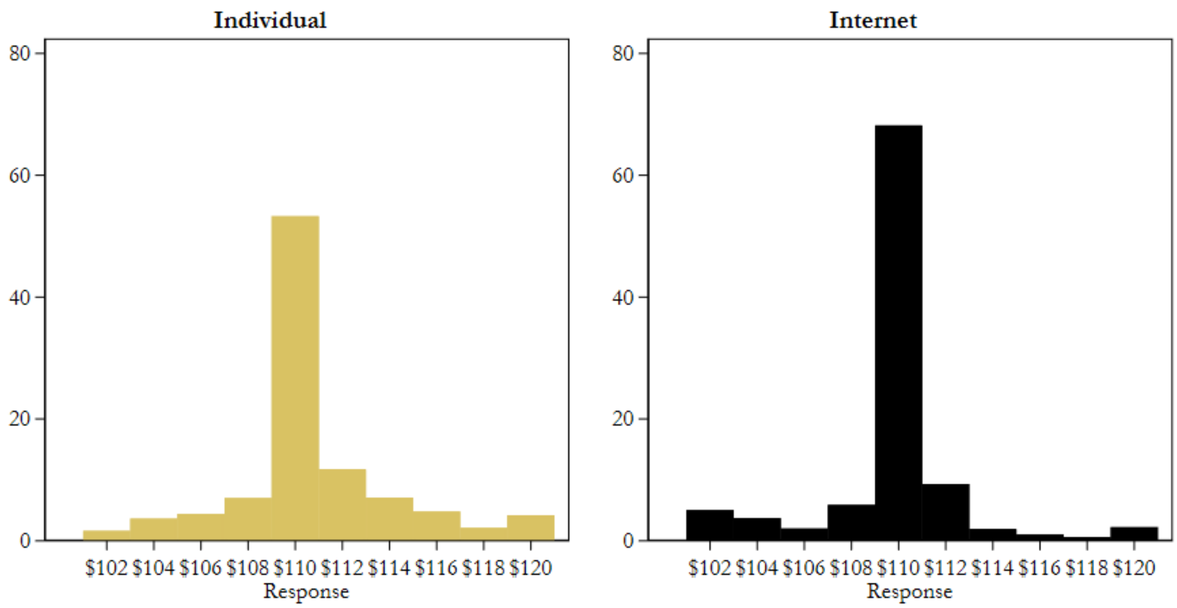


Table 3.1: Effective Literacy of Access to the Internet for the Question about the Savings Account with 2% Interest, New Labels

Correct answer: $110.41 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.110 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-0.53	0.06	-1.08	0.03
female	-0.31	0.22	-0.80	0.18
asian	-0.12	0.75	-0.88	0.63
black	-0.39	0.15	-0.94	0.15
christian	-0.16	0.51	-0.64	0.31
gpaHI	-0.46	0.11	-1.03	0.11
junior	-0.23	0.42	-0.79	0.33
senior	-0.65	0.05	-1.30	0.00
constant	111.51	<0.001	110.69	112.33
		LnSigma		
Internet	-0.15	0.22	-0.38	0.09
female	-0.07	0.60	-0.33	0.19
asian	0.75	<0.001	0.34	1.16
black	0.58	<0.001	0.25	0.91
christian	0.25	0.09	-0.04	0.54
gpaHI	-0.38	<0.001	-0.61	-0.14
junior	-0.33	0.03	-0.62	-0.04
senior	-0.04	0.79	-0.35	0.27
constant	0.75	<0.001	0.36	1.15

Table 3.2: Effective Literacy of Access to the Internet
With and Without Caveat, for
the Question about the Savings Account with 2% Interest, New Labels

Correct answer: $110.41 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.092 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.015 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-0.75	0.03	-1.42	-0.07
internetC	-0.46	0.11	-1.01	0.10
female	-0.26	0.28	-0.74	0.22
asian	-0.22	0.57	-1.01	0.56
black	-0.43	0.11	-0.96	0.09
christian	-0.14	0.56	-0.62	0.34
gpaHI	-0.37	0.15	-0.88	0.13
junior	-0.12	0.66	-0.66	0.42
senior	-0.69	0.04	-1.33	-0.05
constant	111.43	<0.001	110.60	112.26
		LnSigma		
internetNC	-0.38	0.05	-0.75	-0.00
internetC	-0.11	0.39	-0.36	0.14
female	-0.04	0.79	-0.30	0.23
asian	0.82	<0.001	0.41	1.23
black	0.57	<0.001	0.25	0.89
christian	0.26	0.08	-0.03	0.56
gpaHI	-0.41	<0.001	-0.65	-0.16
junior	-0.27	0.08	-0.58	0.04
senior	-0.03	0.84	-0.35	0.28
constant	0.74	<0.001	0.36	1.11

Table 3.3: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Account with 2% Interest, New Labels

Bias is relative to the correct answer: $110.41
Additional imprecision is relative to the average imprecision: 3.18
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.53 (Individuals) and 0.68 (Internet)
Literacy index W = 0.50 (Individuals) and 0.67 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
Bias from the True Response				
internetNC	-1.07	0.02	-1.90	-0.24
internetC	-0.43	0.12	-0.99	0.12
female	-0.40	0.08	-0.84	0.03
asian	-0.15	0.37	-0.95	0.64
black	-0.23	0.24	-0.69	0.22
christian	-0.19	0.27	-0.62	0.24
gpaHI	-0.39	0.04	-0.74	-0.04
junior	-0.35	0.18	-0.89	0.19
senior	-0.66	0.11	-1.48	0.15
Additional Imprecision of Beliefs				
internetNC	-0.69	0.13	-1.59	0.20
internetC	-0.13	0.38	-0.87	0.62
female	0.10	0.37	-0.37	0.58
asian	-0.06	0.39	-0.84	0.72
black	0.28	0.21	-0.20	0.76
christian	0.23	0.25	-0.23	0.70
gpaHI	-0.59	0.03	-1.09	-0.10
junior	-0.98	<0.001	-1.49	-0.47
senior	0.17	0.38	-0.79	1.13

Figure 3.2: Elicited Beliefs about The Social Security Start Question, Comparing Internet and Individual Responses, New Labels

Internet average: 62.5 years (N=104) Individual average: 65.8 years (N=106)
p-value on test of difference in averages < 0.001
True response was 62

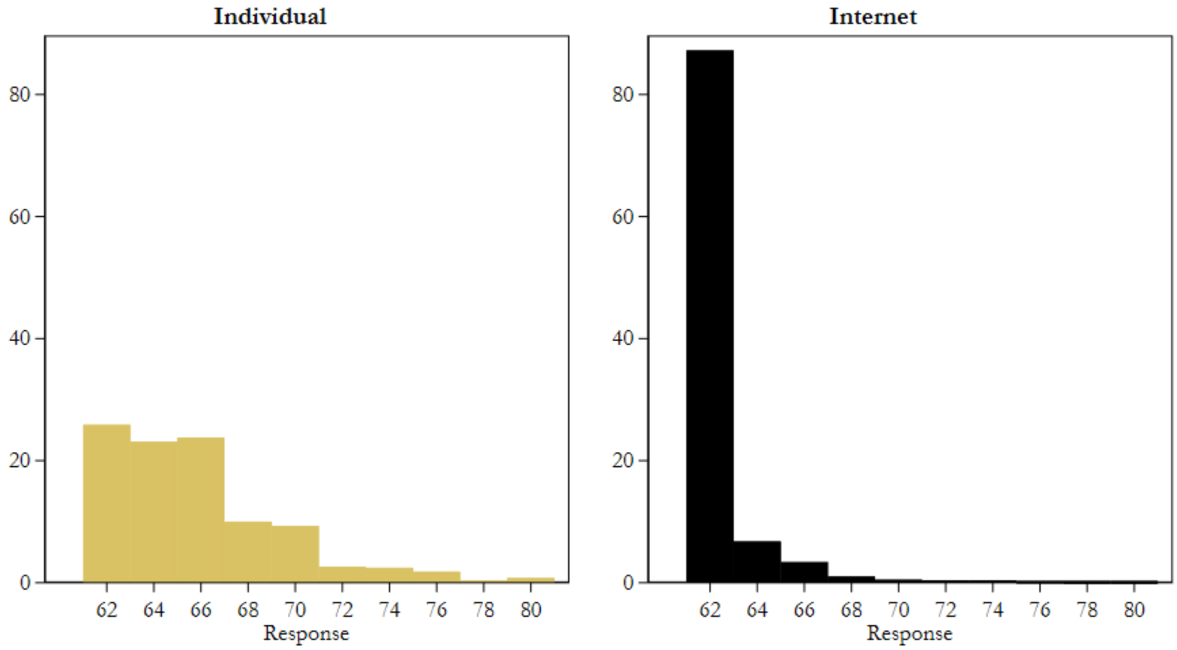


Table 3.4: Effective Literacy of Access to the Internet for the Question about the Social Security Start Age, New Labels

Correct answer: 62 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-3.38	<0.001	-3.91	-2.85
female	-0.16	0.38	-0.53	0.20
asian	0.41	0.17	-0.17	1.00
black	0.13	0.44	-0.20	0.45
christian	0.26	0.10	-0.05	0.56
gpaHI	-0.20	0.26	-0.55	0.15
junior	-0.01	0.97	-0.39	0.38
senior	-0.23	0.25	-0.62	0.16
constant	65.84	<0.001	65.17	66.50
		LnSigma		
Internet	-0.92	<0.001	-1.18	-0.66
female	-0.19	0.23	-0.49	0.12
asian	0.75	<0.001	0.23	1.27
black	0.34	0.02	0.06	0.62
christian	0.17	0.19	-0.08	0.42
gpaHI	-0.24	0.07	-0.50	0.02
junior	0.01	0.98	-0.43	0.45
senior	-0.18	0.27	-0.49	0.14
constant	1.12	<0.001	0.74	1.50

Table 3.5: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Social Security Start Age, New Labels

Correct answer: 62 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-3.73	<0.001	-4.27	-3.19
internetC	-3.30	<0.001	-3.85	-2.74
female	-0.05	0.75	-0.36	0.26
asian	0.45	0.09	-0.08	0.97
black	0.08	0.60	-0.21	0.36
christian	0.22	0.08	-0.03	0.47
gpaHI	-0.21	0.17	-0.50	0.09
junior	0.11	0.54	-0.24	0.45
senior	-0.26	0.10	-0.56	0.05
constant	65.84	<0.001	65.22	66.46
		LnSigma		
internetNC	-1.39	<0.001	-1.86	-0.92
internetC	-0.87	<0.001	-1.15	-0.59
female	-0.09	0.53	-0.36	0.19
asian	0.86	<0.001	0.34	1.38
black	0.30	0.03	0.03	0.57
christian	0.15	0.26	-0.11	0.40
gpaHI	-0.28	0.02	-0.52	-0.04
junior	0.18	0.38	-0.22	0.57
senior	-0.17	0.25	-0.46	0.12
constant	1.10	<0.001	0.74	1.46

Table 3.6: Literacy Bias and Imprecision, by Demographics, for the Question about the Social Security Start Age, New Labels

Bias is relative to the correct answer: 62
Additional imprecision is relative to the average imprecision: 3.17
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.26 (Individuals) and 0.87 (Internet)
Literacy index W = 0.21 (Individuals) and 0.84 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	0.30	0.10	-0.06	0.67
internetC	0.63	<0.001	0.35	0.92
female	1.99	<0.001	1.56	2.43
asian	1.79	<0.001	1.07	2.51
black	2.22	<0.001	1.76	2.67
christian	2.24	<0.001	1.79	2.68
gpaHI	2.02	<0.001	1.51	2.52
junior	1.58	<0.001	0.95	2.21
senior	1.75	<0.001	1.17	2.33
		Additional Imprecision of Beliefs		
internetNC	-1.51	0.04	-2.90	-0.12
internetC	-1.43	<0.001	-2.00	-0.85
female	-0.01	0.40	-0.47	0.44
asian	-0.11	0.39	-0.99	0.78
black	0.12	0.34	-0.32	0.57
christian	0.08	0.37	-0.34	0.50
gpaHI	0.11	0.37	-0.41	0.63
junior	-0.46	0.18	-1.16	0.25
senior	-0.48	0.03	-0.90	-0.07

Figure 3.3: Elicited Beliefs about The Medicare Eligibility Question, Comparing Internet and Individual Responses, New Labels

Internet average: 64.0 years (N=104)   Individual average: 57.3 years (N=106)

p-value on test of difference in averages < 0.001

True response was 65

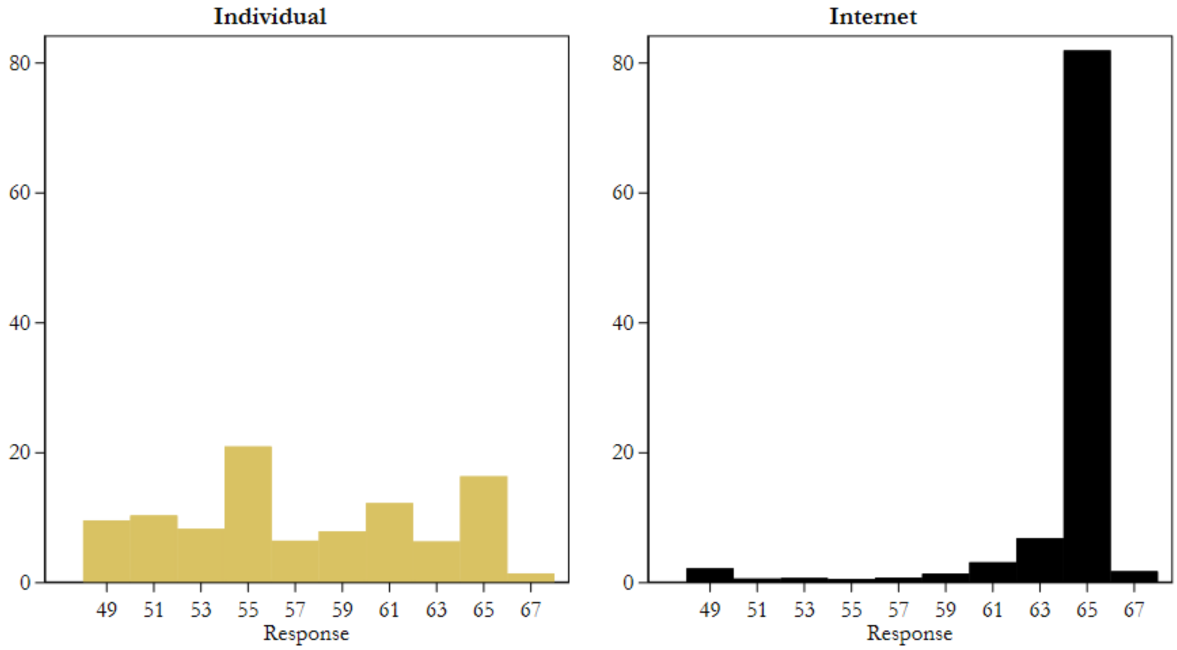


Table 3.7: Effective Literacy of Access to the Internet for the Question about Medicare Eligibility, New Labels

Correct answer: 65 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	6.75	<0.001	5.72	7.79
female	0.25	0.37	-0.30	0.80
asian	-0.71	0.05	-1.44	0.01
black	-0.74	0.02	-1.38	-0.10
christian	-0.17	0.57	-0.75	0.41
gpaHI	0.11	0.70	-0.45	0.66
junior	-0.52	0.35	-1.62	0.58
senior	0.09	0.71	-0.39	0.58
constant	58.00	<0.001	56.70	59.30
		LnSigma		
Internet	-0.93	<0.001	-1.22	-0.65
female	0.04	0.76	-0.23	0.32
asian	0.67	<0.001	0.22	1.11
black	0.58	<0.001	0.28	0.87
christian	-0.13	0.44	-0.48	0.21
gpaHI	-0.25	0.10	-0.54	0.05
junior	0.47	0.07	-0.03	0.98
senior	-0.11	0.36	-0.36	0.13
constant	1.34	<0.001	0.90	1.78

Table 3.8: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about Medicare Eligibility, New Labels

Correct answer: 65 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	6.73	<0.001	5.60	7.85
internetC	6.76	<0.001	5.71	7.80
female	0.25	0.38	-0.31	0.81
asian	-0.69	0.05	-1.39	0.01
black	-0.74	0.02	-1.38	-0.10
christian	-0.18	0.56	-0.76	0.41
gpaHI	0.11	0.72	-0.47	0.68
junior	-0.52	0.36	-1.62	0.58
senior	0.08	0.75	-0.44	0.61
constant	58.01	<0.001	56.71	59.32
		LnSigma		
internetNC	-0.90	<0.001	-1.36	-0.43
internetC	-0.94	<0.001	-1.27	-0.62
female	0.04	0.78	-0.24	0.32
asian	0.65	<0.001	0.22	1.09
black	0.58	<0.001	0.29	0.87
christian	-0.13	0.46	-0.48	0.22
gpaHI	-0.24	0.11	-0.54	0.06
junior	0.46	0.08	-0.05	0.96
senior	-0.12	0.36	-0.36	0.13
constant	1.34	<0.001	0.90	1.78

Table 3.9: Literacy Bias and Imprecision, by Demographics, for the Question about Medicare Eligibility, New Labels

Bias is relative to the correct answer: 65
Additional imprecision is relative to the average imprecision: 5.38
Sample sizes: 106 Individuals and 104 Internet
Literacy index $L = 0.16$ (Individuals) and 0.82 (Internet)
Literacy index $W = 0.16$ (Individuals) and 0.80 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-1.47	0.02	-2.68	-0.26
internetC	-0.71	<0.001	-1.18	-0.25
female	-3.78	<0.001	-4.57	-2.99
asian	-3.98	<0.001	-5.43	-2.53
black	-4.61	<0.001	-5.52	-3.71
christian	-4.48	<0.001	-5.34	-3.62
gpaHI	-3.98	<0.001	-4.91	-3.05
junior	-3.93	<0.001	-5.54	-2.32
senior	-3.10	<0.001	-4.30	-1.90
		Additional Imprecision of Beliefs		
internetNC	-1.39	0.14	-3.29	0.50
internetC	-2.98	<0.001	-4.00	-1.96
female	-0.16	0.30	-0.58	0.25
asian	0.04	0.40	-0.77	0.85
black	0.24	0.19	-0.15	0.63
christian	0.08	0.36	-0.28	0.45
gpaHI	-0.04	0.39	-0.51	0.43
junior	0.39	0.29	-0.56	1.33
senior	-0.60	0.11	-1.34	0.13

Figure 3.4: Elicited Beliefs about The Real Interest Rate Question, Comparing Internet and Individual Responses, New Labels

Internet average: \$100.6 (N=104)   Individual average: \$101.0 (N=106)
p-value on test of difference in averages = 0.407
True response was \$98.98

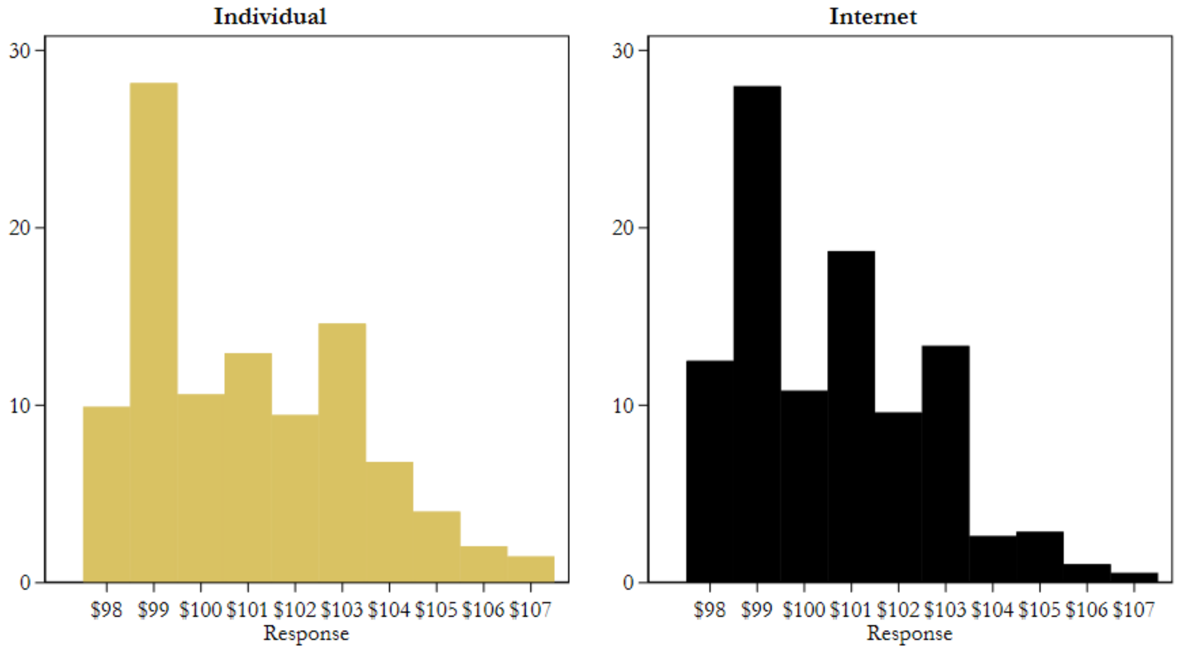


Table 3.10: Effective Literacy of Access to the Internet for the Question about the Real Interest Rate, New Labels

Correct answer: $98.98 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.182 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-0.22	0.39	-0.73	0.28
female	0.59	0.01	0.12	1.07
asian	0.76	0.04	0.04	1.49
black	0.57	0.08	-0.07	1.21
christian	0.71	0.01	0.15	1.28
gpaHI	-0.66	<0.001	-1.13	-0.18
junior	-0.17	0.58	-0.76	0.43
senior	-0.43	0.17	-1.04	0.19
constant	100.05	<0.001	99.23	100.88
		LnSigma		
Internet	-0.15	0.07	-0.30	0.01
female	0.05	0.49	-0.10	0.20
asian	0.21	0.14	-0.07	0.50
black	0.13	0.19	-0.07	0.34
christian	0.15	0.10	-0.03	0.33
gpaHI	-0.17	0.02	-0.31	-0.03
junior	0.02	0.88	-0.19	0.22
senior	0.07	0.45	-0.11	0.25
constant	0.56	<0.001	0.29	0.84

Table 3.11: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Real Interest Rate, New Labels

Correct answer: $98.98 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.536 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.351 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-0.42	0.27	-1.16	0.32
internetC	-0.14	0.61	-0.68	0.40
female	0.63	<0.001	0.16	1.10
asian	0.77	0.03	0.06	1.48
black	0.57	0.08	-0.07	1.20
christian	0.68	0.02	0.10	1.26
gpaHI	-0.72	<0.001	-1.23	-0.20
junior	-0.16	0.60	-0.75	0.43
senior	-0.43	0.17	-1.04	0.18
constant	100.09	<0.001	99.26	100.92
		LnSigma		
internetNC	-0.20	0.06	-0.41	0.01
internetC	-0.12	0.18	-0.30	0.06
female	0.06	0.42	-0.09	0.21
asian	0.22	0.12	-0.06	0.50
black	0.15	0.14	-0.05	0.34
christian	0.13	0.16	-0.05	0.32
gpaHI	-0.19	0.02	-0.35	-0.03
junior	0.00	0.96	-0.20	0.21
senior	0.07	0.47	-0.11	0.25
constant	0.57	<0.001	0.30	0.84

Table 3.12: Literacy Bias and Imprecision, by Demographics, for the Question about the Real Interest Rate, New Labels

Bias is relative to the correct answer: \$98.98
Additional imprecision is relative to the average imprecision: 2.12
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.28 (Individuals) and 0.28 (Internet)
Literacy index W = 0.29 (Individuals) and 0.28 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	1.59	<0.001	0.94	2.24
internetC	1.64	<0.001	1.24	2.04
female	2.08	<0.001	1.77	2.40
asian	1.55	<0.001	1.02	2.08
black	2.10	<0.001	1.78	2.42
christian	2.14	<0.001	1.83	2.45
gpaHI	1.55	<0.001	1.23	1.87
junior	1.61	<0.001	1.11	2.12
senior	1.47	<0.001	0.92	2.03
		Additional Imprecision of Beliefs		
internetNC	-0.14	0.28	-0.48	0.19
internetC	-0.15	0.11	-0.33	0.04
female	-0.01	0.40	-0.15	0.14
asian	-0.10	0.33	-0.42	0.22
black	0.05	0.32	-0.09	0.19
christian	0.04	0.34	-0.10	0.19
gpaHI	-0.17	0.08	-0.36	0.02
junior	-0.16	0.15	-0.39	0.07
senior	0.04	0.38	-0.26	0.35

Figure 3.5: Elicited Beliefs about The Savings Horizon Question, Comparing Internet and Individual Responses, New Labels

Internet average: 4.4 months (N=104) Individual average: 5.0 months (N=106)
p-value on test of difference in averages = 0.242
True response was 4 months (4m)

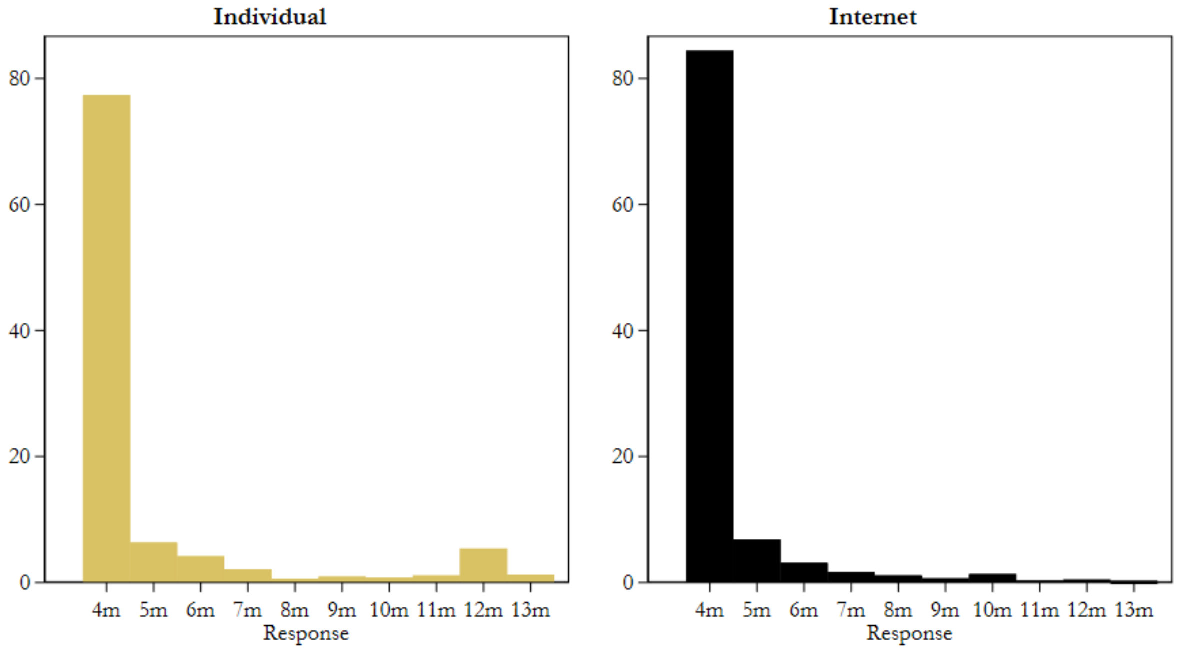


Table 3.13: Effective Literacy of Access to the Internet for the Question about the Savings Horizon, New Labels

Correct answer: 4 months (4m) N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.341 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-0.18	0.21	-0.46	0.10
female	-0.10	0.45	-0.37	0.16
asian	0.31	0.05	-0.00	0.61
black	0.33	0.04	0.02	0.63
christian	0.37	<0.001	0.09	0.65
gpaHI	-0.04	0.75	-0.27	0.19
junior	0.36	0.16	-0.14	0.86
senior	-0.03	0.77	-0.27	0.20
constant	4.21	<0.001	3.77	4.65
		LnSigma		
Internet	-0.31	0.14	-0.73	0.11
female	-0.39	0.06	-0.79	0.01
asian	0.71	0.05	0.01	1.40
black	0.76	<0.001	0.23	1.28
christian	0.63	<0.001	0.17	1.10
gpaHI	-0.02	0.90	-0.40	0.35
junior	0.66	0.01	0.15	1.17
senior	-0.03	0.89	-0.49	0.43
constant	-0.37	0.34	-1.13	0.39

Table 3.14: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Savings Horizon, New Labels

Correct answer: 4 months (4m) N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.129 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.042 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-0.36	0.05	-0.72	0.00
internetC	-0.11	0.45	-0.40	0.18
female	0.06	0.60	-0.17	0.29
asian	0.32	0.08	-0.03	0.67
black	0.13	0.41	-0.19	0.45
christian	0.25	0.09	-0.04	0.54
gpaHI	-0.17	0.15	-0.40	0.06
junior	0.32	0.17	-0.14	0.77
senior	-0.16	0.20	-0.40	0.09
constant	4.35	<0.001	3.87	4.84
		LnSigma		
internetNC	-1.07	<0.001	-1.75	-0.38
internetC	-0.21	0.36	-0.65	0.23
female	-0.22	0.27	-0.60	0.17
asian	0.93	<0.001	0.29	1.56
black	0.72	<0.001	0.22	1.22
christian	0.65	<0.001	0.18	1.11
gpaHI	-0.13	0.48	-0.51	0.24
junior	0.85	<0.001	0.36	1.35
senior	0.01	0.96	-0.45	0.47
constant	-0.46	0.20	-1.17	0.25

Table 3.15: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Horizon, New Labels

Bias is relative to the correct answer: 4 months (4m)
Additional imprecision is relative to the average imprecision: 1.82
Sample sizes: 106 Individuals and 104 Internet
Literacy index $L = 0.77$ (Individuals) and 0.84 (Internet)
Literacy index $W = 0.74$ (Individuals) and 0.83 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	0.28	0.02	0.05	0.50
internetC	0.47	<0.001	0.23	0.72
female	0.63	<0.001	0.37	0.90
asian	0.52	0.03	0.07	0.96
black	0.86	<0.001	0.53	1.18
christian	0.87	<0.001	0.55	1.19
gpaHI	0.66	<0.001	0.34	0.97
junior	0.96	<0.001	0.31	1.62
senior	0.64	<0.001	0.22	1.06
		Additional Imprecision of Beliefs		
internetNC	-0.91	<0.001	-1.50	-0.32
internetC	-0.51	0.03	-0.94	-0.07
female	-0.11	0.36	-0.58	0.35
asian	-0.18	0.37	-1.12	0.76
black	0.21	0.27	-0.25	0.68
christian	0.21	0.27	-0.25	0.68
gpaHI	-0.03	0.40	-0.58	0.52
junior	0.54	0.18	-0.31	1.39
senior	-0.06	0.39	-0.80	0.69

Figure 3.6: Elicited Beliefs about The Stolen Credit Card Question, Comparing Internet and Individual Responses, New Labels

Internet average: $80.6 ($N$=104) Individual average: $171.0 ($N$=106)
p-value on test of difference in averages = 0.086
True response was $50

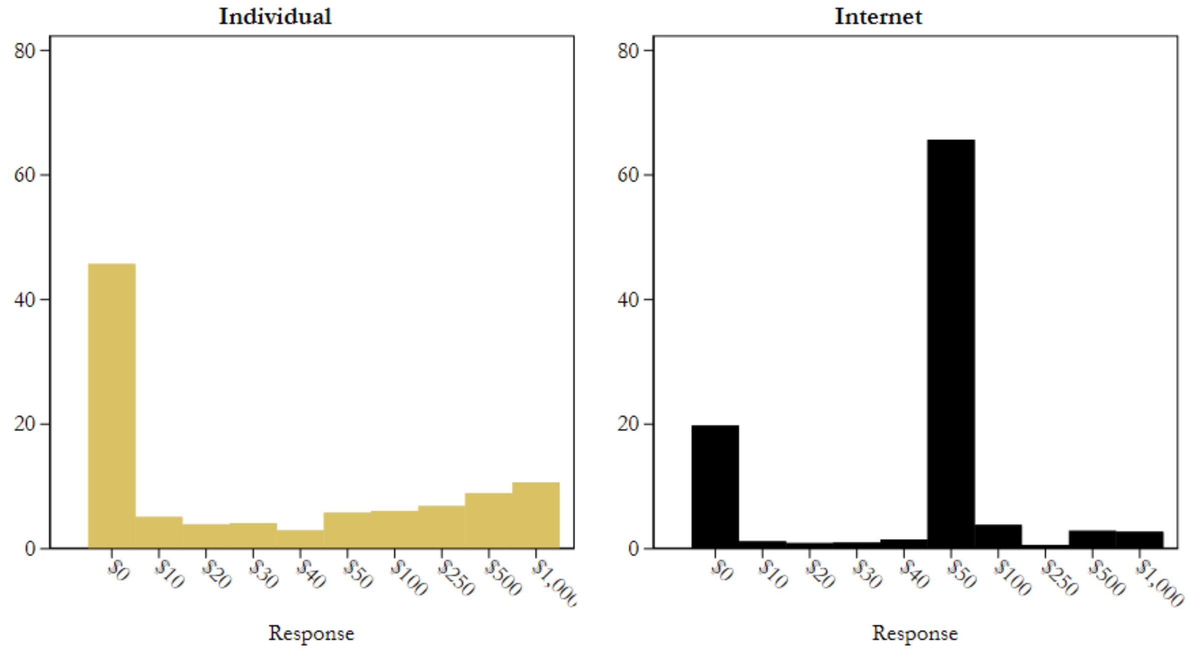


Table 3.16: Effective Literacy of Access to the Internet for the Question about the Stolen Credit Card, New Labels

Correct answer: \$50 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-69.65	0.02	-129.01	-10.29
female	3.15	0.86	-31.25	37.56
asian	-13.86	0.56	-60.83	33.11
black	2.74	0.91	-44.02	49.50
christian	3.28	0.89	-44.84	51.39
gpaHI	10.37	0.56	-24.60	45.34
junior	-34.75	0.14	-80.92	11.43
senior	-26.89	0.33	-81.30	27.52
constant	161.18	<0.001	90.54	231.82
		LnSigma		
Internet	-0.73	<0.001	-1.11	-0.35
female	0.11	0.55	-0.26	0.49
asian	-0.07	0.86	-0.79	0.65
black	0.15	0.58	-0.37	0.67
christian	0.07	0.75	-0.38	0.52
gpaHI	0.17	0.38	-0.21	0.54
junior	-0.41	0.06	-0.84	0.02
senior	-0.24	0.37	-0.77	0.29
constant	5.46	<0.001	4.75	6.18

Table 3.17: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Stolen Credit Card, New Labels

Correct answer: \$50 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.026 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-47.39	0.18	-115.90	21.12
internetC	-74.66	0.02	-135.63	-13.70
female	1.49	0.93	-33.80	36.78
asian	-21.02	0.37	-67.33	25.29
black	9.29	0.70	-38.33	56.91
christian	-4.35	0.86	-53.55	44.85
gpaHI	9.80	0.57	-23.98	43.57
junior	-36.65	0.07	-76.44	3.13
senior	-17.97	0.53	-73.71	37.76
constant	160.55	<0.001	90.76	230.33
		LnSigma		
internetNC	-0.45	0.12	-1.02	0.12
internetC	-0.89	<0.001	-1.29	-0.50
female	0.13	0.46	-0.22	0.49
asian	-0.12	0.72	-0.79	0.54
black	0.25	0.33	-0.25	0.76
christian	0.04	0.87	-0.41	0.49
gpaHI	0.12	0.52	-0.24	0.48
junior	-0.49	0.02	-0.90	-0.08
senior	-0.16	0.57	-0.71	0.39
constant	5.44	<0.001	4.75	6.13

Table 3.18: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Credit Card, New Labels

Bias is relative to the correct answer: $50
Additional imprecision is relative to the average imprecision: 218.67
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.06 (Individuals) and 0.66 (Internet)
Literacy index W = 0.06 (Individuals) and 0.64 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	41.20	0.05	1.67	80.74
internetC	25.74	0.02	5.09	46.38
female	76.52	<0.001	41.06	111.97
asian	46.88	0.06	0.47	93.29
black	89.27	<0.001	50.30	128.23
christian	92.77	<0.001	56.41	129.12
gpaHI	82.14	<0.001	41.50	122.79
junior	30.92	0.08	-3.04	64.88
senior	53.71	0.04	4.88	102.54
		Additional Imprecision of Beliefs		
internetNC	-64.55	0.11	-144.12	15.02
internetC	-91.62	<0.001	-138.70	-44.54
female	9.34	0.37	-37.14	55.82
asian	-39.01	0.23	-112.27	34.24
black	24.08	0.24	-23.03	71.19
christian	21.87	0.24	-20.82	64.57
gpaHI	11.26	0.36	-40.89	63.41
junior	-62.84	0.06	-125.56	-0.12
senior	-17.79	0.36	-94.53	58.95

Figure 3.7: Elicited Beliefs about The Stolen Debit Card Question, Comparing Internet and Individual Responses, New Labels

Internet average: \$472.0 (N=104) Individual average: \$941.8 (N=106)
p-value on test of difference in averages = 0.108
True response was \$500

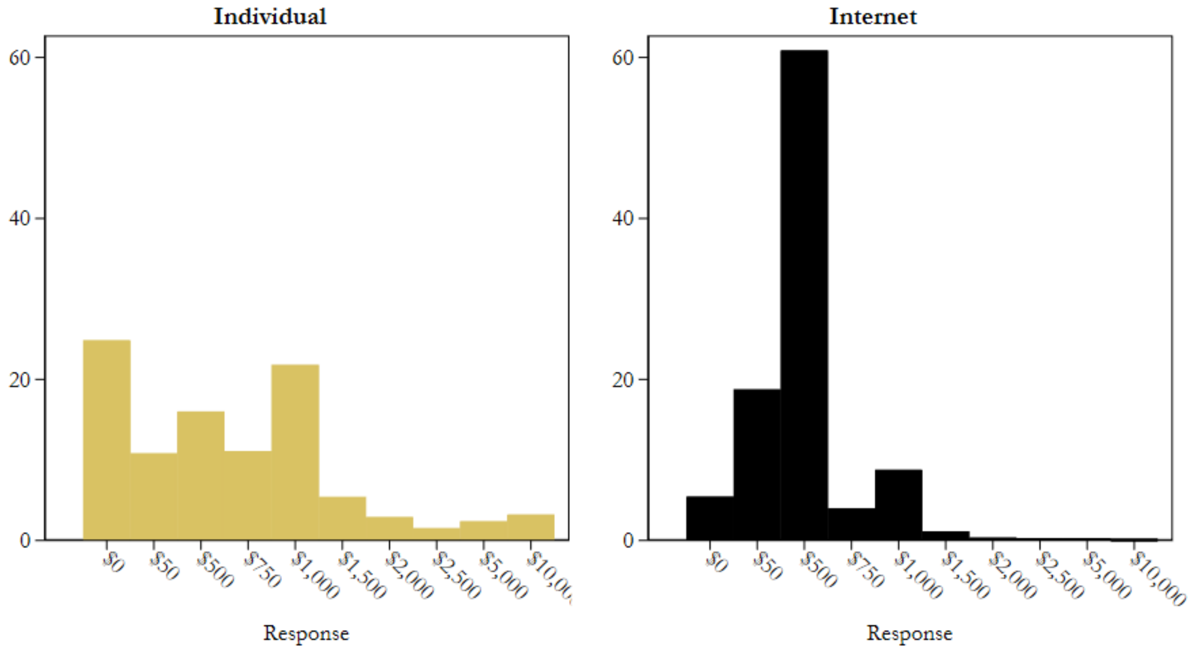


Table 3.19: Effective Literacy of Access to the Internet for the Question about the Stolen Debit Card, New Labels

Correct answer: \$500 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-312.64	<0.001	-517.78	-107.49
female	-46.35	0.43	-160.86	68.17
asian	46.00	0.51	-91.37	183.37
black	119.13	0.14	-40.33	278.59
christian	35.95	0.63	-109.73	181.62
gpaHI	-56.32	0.25	-151.98	39.34
junior	7.76	0.93	-159.75	175.27
senior	26.61	0.68	-98.74	151.96
constant	756.43	<0.001	487.57	1025.30
		LnSigma		
Internet	-1.02	<0.001	-1.39	-0.64
female	-0.02	0.91	-0.45	0.40
asian	-0.10	0.74	-0.71	0.51
black	0.11	0.68	-0.43	0.66
christian	0.23	0.31	-0.22	0.68
gpaHI	0.07	0.72	-0.30	0.43
junior	0.24	0.41	-0.34	0.83
senior	0.43	0.09	-0.06	0.93
constant	6.76	<0.001	6.08	7.43

Table 3.20: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Stolen Debit Card, New Labels

Correct answer: \$500   N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.004 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-222.31	0.09	-482.02	37.40
internetC	-312.52	<0.001	-507.14	-117.91
female	-68.65	0.28	-192.26	54.97
asian	45.60	0.56	-107.64	198.84
black	153.34	0.11	-33.01	339.68
christian	30.30	0.68	-112.69	173.28
gpaHI	-29.32	0.57	-129.99	71.34
junior	-15.37	0.84	-163.16	132.42
senior	25.80	0.73	-121.01	172.62
constant	725.17	<0.001	454.18	996.16
		LnSigma		
internetNC	-0.72	0.05	-1.44	-0.00
internetC	-1.09	<0.001	-1.48	-0.70
female	-0.08	0.70	-0.50	0.34
asian	-0.18	0.55	-0.77	0.41
black	0.18	0.51	-0.36	0.72
christian	0.22	0.33	-0.23	0.67
gpaHI	0.12	0.51	-0.25	0.50
junior	0.12	0.67	-0.42	0.65
senior	0.41	0.10	-0.08	0.90
constant	6.75	<0.001	6.09	7.42

Table 3.21: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Debit Card, New Labels

Bias is relative to the correct answer: $500
Additional imprecision is relative to the average imprecision: 1073.09
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.16 (Individuals) and 0.61 (Internet)
Literacy index W = 0.12 (Individuals) and 0.59 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-22.45	0.36	-117.66	72.75
internetC	-30.57	0.23	-87.15	26.01
female	177.29	0.02	38.04	316.53
asian	-8.44	0.39	-121.57	104.68
black	248.62	<0.001	96.91	400.33
christian	282.50	<0.001	94.19	470.82
gpaHI	224.01	0.06	2.50	445.52
junior	31.71	0.35	-94.62	158.03
senior	393.23	0.10	-70.45	856.91
		Additional Imprecision of Beliefs		
internetNC	-667.21	<0.001	-899.68	-434.73
internetC	-653.20	<0.001	-786.72	-519.67
female	-32.15	0.39	-392.54	328.23
asian	-565.52	<0.001	-743.56	-387.48
black	42.15	0.39	-320.42	404.72
christian	180.75	0.29	-266.07	627.58
gpaHI	147.12	0.35	-410.74	704.99
junior	-395.22	0.03	-733.11	-57.33
senior	639.51	0.15	-255.98	1535.00

Figure 3.8: Elicited Beliefs about The Nominal Interest Question, Comparing Internet and Individual Responses, New Labels

Internet average: 4.4% (N=104) Individual average: 4.5% (N=106)
p-value on test of difference in averages = 0.181
True response was 5%

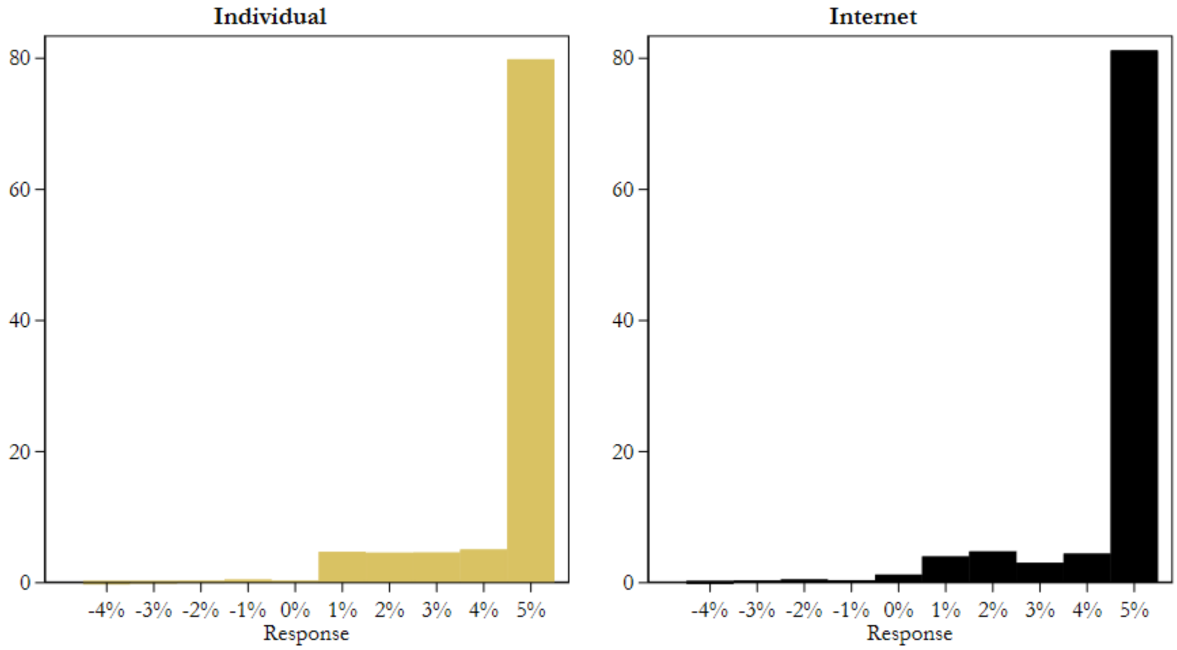


Table 3.22: Effective Literacy of Access to the Internet for the Question about Nominal Interest, New Labels

Correct answer: 5% N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.364 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	0.15	0.20	-0.08	0.38
female	-0.17	0.15	-0.41	0.06
asian	-0.10	0.53	-0.40	0.20
black	0.03	0.85	-0.29	0.34
christian	-0.12	0.37	-0.39	0.15
gpaHI	0.50	<0.001	0.20	0.80
junior	0.11	0.47	-0.20	0.42
senior	0.28	0.04	0.02	0.55
constant	4.20	<0.001	3.70	4.71
		LnSigma		
Internet	-0.12	0.43	-0.43	0.18
female	0.04	0.83	-0.33	0.41
asian	0.34	0.15	-0.12	0.79
black	0.18	0.40	-0.24	0.59
christian	0.22	0.27	-0.18	0.62
gpaHI	-0.73	<0.001	-1.04	-0.43
junior	0.03	0.91	-0.45	0.50
senior	-0.40	0.04	-0.78	-0.01
constant	0.30	0.27	-0.23	0.83

Table 3.23: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about Nominal Interest, New Labels

Correct answer: 5% N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.296 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.554 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	0.25	0.13	-0.07	0.57
internetC	0.12	0.30	-0.11	0.36
female	-0.18	0.14	-0.42	0.06
asian	-0.10	0.53	-0.40	0.20
black	0.06	0.71	-0.26	0.38
christian	-0.12	0.36	-0.39	0.14
gpaHI	0.53	<0.001	0.22	0.84
junior	0.09	0.57	-0.23	0.42
senior	0.30	0.02	0.04	0.55
constant	4.16	<0.001	3.65	4.67
		LnSigma		
internetNC	-0.26	0.29	-0.73	0.22
internetC	-0.09	0.57	-0.41	0.23
female	0.05	0.78	-0.32	0.43
asian	0.34	0.15	-0.12	0.80
black	0.13	0.54	-0.29	0.55
christian	0.22	0.27	-0.17	0.62
gpaHI	-0.77	<0.001	-1.10	-0.44
junior	0.03	0.91	-0.44	0.49
senior	-0.41	0.03	-0.79	-0.04
constant	0.35	0.19	-0.18	0.88

Table 3.24: Literacy Bias and Imprecision, by Demographics, for the Question about Nominal Interest, New Labels

Bias is relative to the correct answer: 5%
Additional imprecision is relative to the average imprecision: 1.30
Sample sizes: 106 Individuals and 104 Internet
Literacy index $L = 0.80$ (Individuals) and 0.81 (Internet)
Literacy index $W = 0.76$ (Individuals) and 0.83 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-0.57	0.01	-1.00	-0.14
internetC	-0.56	<0.001	-0.85	-0.28
female	-0.58	<0.001	-0.75	-0.40
asian	-0.46	<0.001	-0.75	-0.17
black	-0.63	<0.001	-0.84	-0.42
christian	-0.59	<0.001	-0.78	-0.40
gpaHI	-0.34	<0.001	-0.47	-0.20
junior	-0.60	<0.001	-1.01	-0.19
senior	-0.38	<0.001	-0.59	-0.17
		Additional Imprecision of Beliefs		
internetNC	0.06	0.39	-0.47	0.58
internetC	0.06	0.39	-0.44	0.56
female	-0.08	0.30	-0.29	0.13
asian	-0.11	0.35	-0.53	0.31
black	0.07	0.36	-0.23	0.38
christian	0.02	0.40	-0.28	0.32
gpaHI	-0.41	<0.001	-0.61	-0.20
junior	0.25	0.31	-0.43	0.93
senior	-0.25	0.14	-0.60	0.09

Figure 3.9: Elicited Beliefs about The Interest Rate Question, Comparing Internet and Individual Responses, New Labels

Internet average: $101.7 (*N*=104) Individual average: $101.7 (*N*=106)
p-value on test of difference in averages = 0.386
True response was $102.00

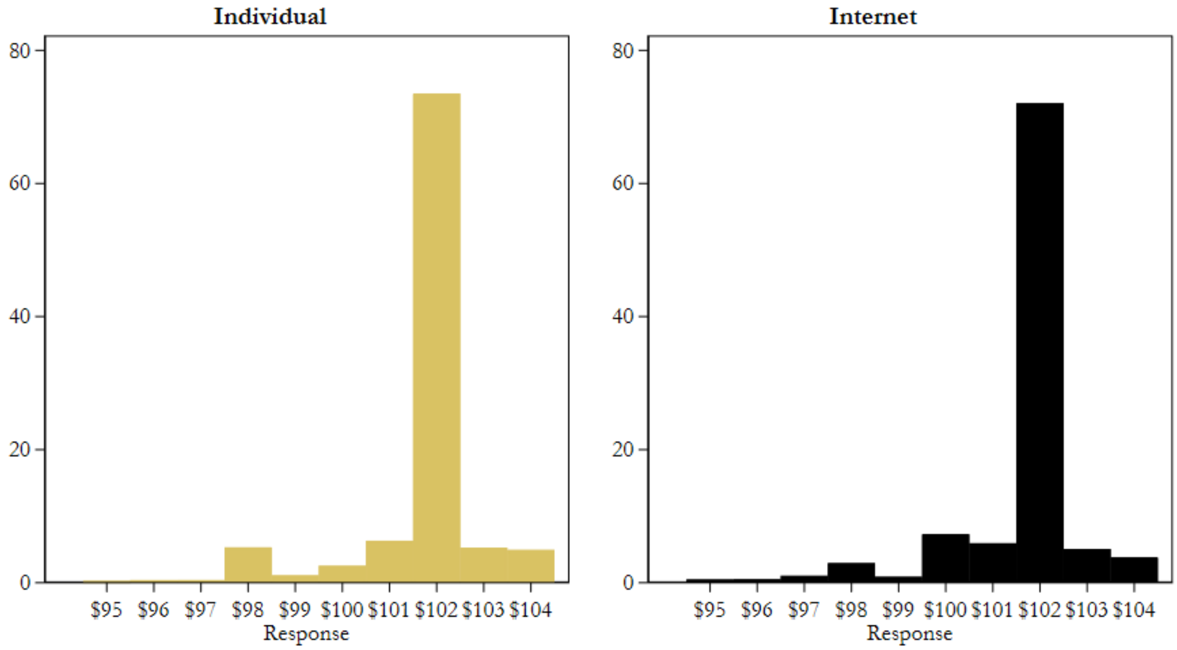


Table 3.25: Effective Literacy of Access to the Internet for the Question about the Interest Rate, New Labels

Correct answer: \$102.00 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.659 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	-0.09	0.40	-0.29	0.12
female	0.12	0.25	-0.08	0.32
asian	-0.08	0.59	-0.35	0.20
black	-0.06	0.62	-0.30	0.17
christian	-0.13	0.28	-0.36	0.10
gpaHI	0.28	0.01	0.07	0.50
junior	0.17	0.18	-0.08	0.42
senior	0.22	0.06	-0.01	0.45
constant	101.58	<0.001	101.24	101.93
		LnSigma		
Internet	0.05	0.77	-0.25	0.35
female	-0.07	0.65	-0.39	0.25
asian	0.54	0.03	0.04	1.03
black	0.19	0.39	-0.25	0.63
christian	0.52	<0.001	0.17	0.87
gpaHI	-0.40	<0.001	-0.69	-0.11
junior	-0.21	0.30	-0.62	0.19
senior	-0.34	0.04	-0.67	-0.01
constant	-0.08	0.76	-0.58	0.42

Table 3.26: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Interest Rate, New Labels

Correct answer: $102.00 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.384 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.698 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	-0.26	0.17	-0.62	0.11
internetC	-0.02	0.83	-0.22	0.18
female	0.13	0.20	-0.07	0.33
asian	-0.06	0.64	-0.33	0.20
black	-0.06	0.57	-0.29	0.16
christian	-0.15	0.19	-0.37	0.07
gpaHI	0.26	0.01	0.05	0.47
junior	0.21	0.10	-0.04	0.46
senior	0.23	0.05	-0.00	0.46
constant	101.59	<0.001	101.26	101.92
		LnSigma		
internetNC	0.16	0.45	-0.26	0.57
internetC	-0.01	0.96	-0.33	0.31
female	-0.07	0.65	-0.40	0.25
asian	0.54	0.04	0.03	1.04
black	0.18	0.42	-0.26	0.63
christian	0.57	<0.001	0.23	0.91
gpaHI	-0.37	0.01	-0.67	-0.07
junior	-0.23	0.29	-0.64	0.19
senior	-0.37	0.02	-0.69	-0.05
constant	-0.11	0.67	-0.61	0.39

Table 3.27: Literacy Bias and Imprecision, by Demographics, for the Question about the Interest Rate, New Labels

Bias is relative to the correct answer: $102.00
Additional imprecision is relative to the average imprecision: 1.26
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.74 (Individuals) and 0.72 (Internet)
Literacy index W = 0.68 (Individuals) and 0.71 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-0.39	0.06	-0.79	-0.00
internetC	-0.31	0.02	-0.55	-0.06
female	-0.25	<0.001	-0.41	-0.09
asian	-0.26	0.06	-0.53	-0.00
black	-0.33	<0.001	-0.52	-0.14
christian	-0.35	<0.001	-0.54	-0.16
gpaHI	-0.14	0.07	-0.28	0.01
junior	-0.25	0.12	-0.57	0.06
senior	-0.13	0.21	-0.35	0.10
		Additional Imprecision of Beliefs		
internetNC	0.03	0.40	-0.41	0.47
internetC	-0.01	0.40	-0.41	0.39
female	-0.03	0.38	-0.23	0.17
asian	-0.11	0.32	-0.44	0.22
black	0.09	0.31	-0.16	0.34
christian	0.14	0.23	-0.11	0.38
gpaHI	-0.22	0.10	-0.47	0.03
junior	-0.05	0.39	-0.63	0.53
senior	-0.22	0.21	-0.60	0.16

Figure 3.10: Elicited Beliefs about
The Remaining Life for Men Question,
Comparing Internet and Individual Responses, New Labels

Internet average: 50.2 years (N=104) Individual average: 47.6 years (N=106)
p-value on test of difference in averages = 0.008
True response was 57.1

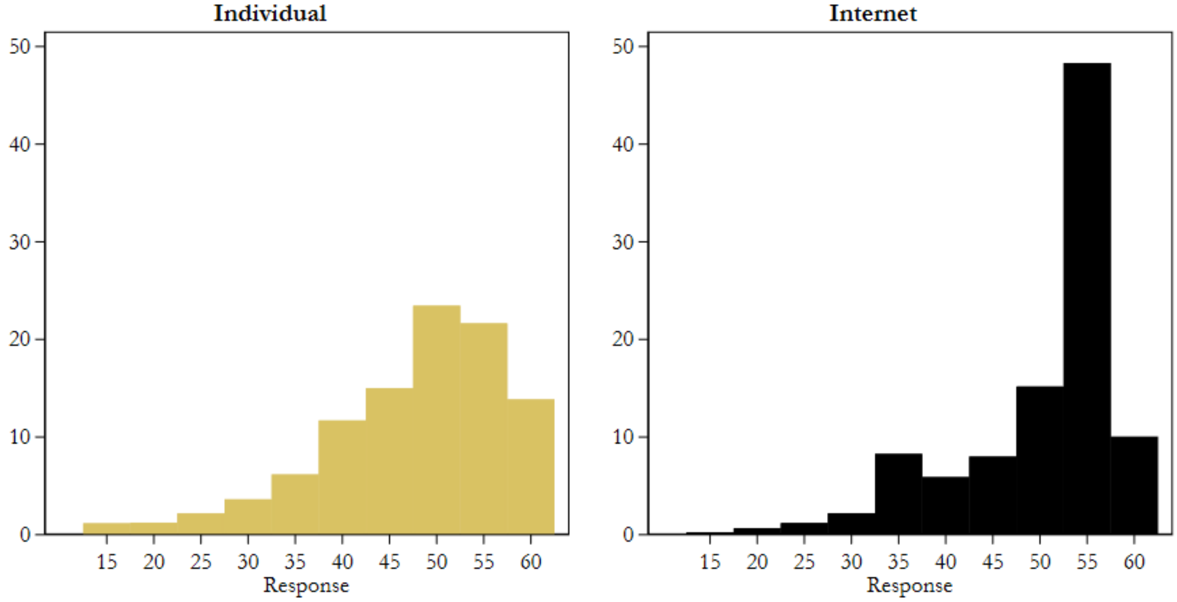


Table 3.28: Effective Literacy of Access to the Internet for the Question about the Remaining Life for Men, New Labels

Correct answer: 57.1 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is 0.007 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	3.01	<0.001	0.91	5.11
female	0.93	0.30	-0.85	2.71
asian	-4.23	0.01	-7.53	-0.93
black	-3.99	<0.001	-6.05	-1.94
christian	-0.55	0.61	-2.67	1.58
gpaHI	1.31	0.16	-0.52	3.14
junior	-0.67	0.65	-3.52	2.18
senior	1.18	0.18	-0.57	2.93
constant	49.55	<0.001	46.00	53.10
		LnSigma		
Internet	-0.28	<0.001	-0.46	-0.09
female	-0.12	0.25	-0.31	0.08
asian	0.85	<0.001	0.52	1.19
black	0.70	<0.001	0.47	0.93
christian	0.06	0.66	-0.22	0.35
gpaHI	-0.26	<0.001	-0.43	-0.09
junior	0.08	0.55	-0.17	0.32
senior	-0.13	0.24	-0.36	0.09
constant	1.83	<0.001	1.46	2.21

Table 3.29: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Remaining Life for Men, New Labels

Correct answer: 57.1 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.009 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	2.16	0.13	-0.67	4.99
internetC	3.29	<0.001	1.17	5.41
female	1.10	0.25	-0.76	2.96
asian	-4.20	0.01	-7.46	-0.94
black	-4.08	<0.001	-6.17	-2.00
christian	-0.63	0.56	-2.73	1.47
gpaHI	1.20	0.20	-0.65	3.04
junior	-0.54	0.71	-3.42	2.33
senior	1.05	0.26	-0.79	2.89
constant	49.66	<0.001	46.13	53.18
		LnSigma		
internetNC	-0.33	<0.001	-0.56	-0.10
internetC	-0.26	0.02	-0.47	-0.05
female	-0.11	0.27	-0.31	0.09
asian	0.85	<0.001	0.52	1.18
black	0.70	<0.001	0.46	0.93
christian	0.06	0.67	-0.23	0.36
gpaHI	-0.27	<0.001	-0.44	-0.10
junior	0.07	0.58	-0.18	0.33
senior	-0.13	0.26	-0.35	0.10
constant	1.84	<0.001	1.46	2.21

Table 3.30: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Men, New Labels

Bias is relative to the correct answer: 57.1
Additional imprecision is relative to the average imprecision: 9.17
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.22 (Individuals) and 0.48 (Internet)
Literacy index W = 0.18 (Individuals) and 0.42 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-7.81	<0.001	-10.34	-5.27
internetC	-6.52	<0.001	-8.41	-4.64
female	-8.25	<0.001	-9.64	-6.85
asian	-8.42	<0.001	-10.98	-5.86
black	-9.19	<0.001	-10.72	-7.66
christian	-8.43	<0.001	-9.84	-7.02
gpaHI	-7.22	<0.001	-8.57	-5.86
junior	-7.49	<0.001	-9.81	-5.18
senior	-7.70	<0.001	-10.09	-5.31
		Additional Imprecision of Beliefs		
internetNC	-1.18	0.10	-2.60	0.23
internetC	-0.52	0.34	-2.40	1.36
female	0.00	0.40	-1.08	1.09
asian	0.28	0.38	-1.36	1.92
black	0.52	0.26	-0.59	1.64
christian	0.08	0.39	-0.90	1.07
gpaHI	-0.97	0.10	-2.10	0.15
junior	-0.39	0.38	-2.74	1.95
senior	0.16	0.40	-2.21	2.54

Figure 3.11: Elicited Beliefs about
The Remaining Life for Women Question,
Comparing Internet and Individual Responses, New Labels

Internet average: 53.5 years (N=104)   Individual average: 50.3 years (N=106)
p-value on test of difference in averages < 0.001
True response was 61.7

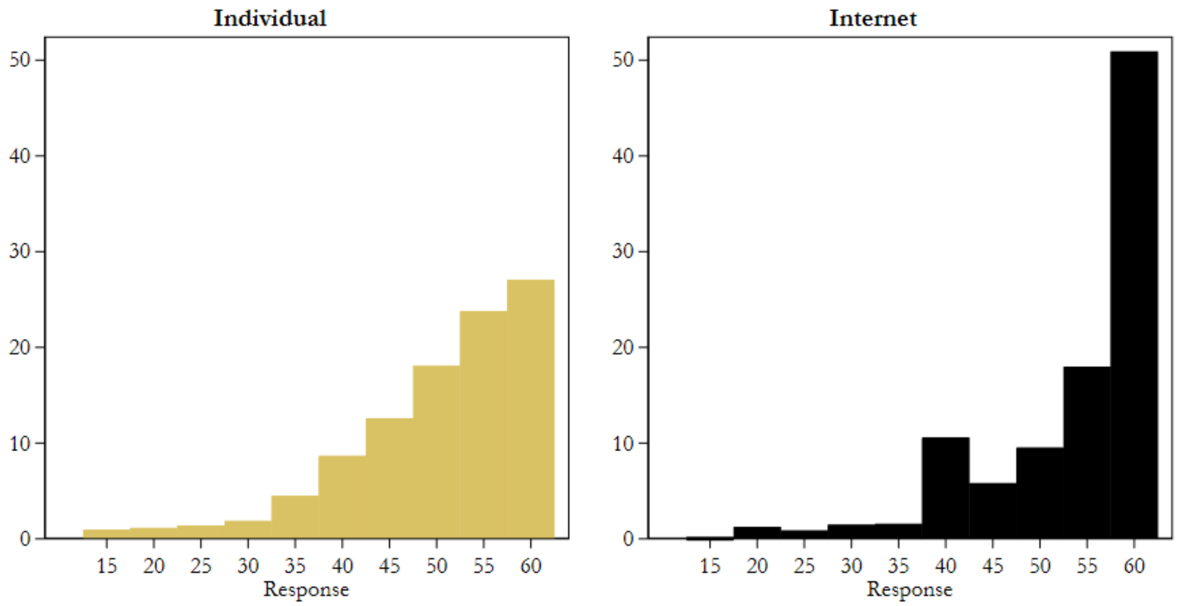


Table 3.31: Effective Literacy of Access to the Internet for the Question about the Remaining Life for Women, New Labels

Correct answer: 61.7 N=210 GSU undergraduates
p-value for test of hypothesis that Internet is zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
Internet	3.55	<0.001	1.78	5.32
female	-0.02	0.98	-1.74	1.70
asian	-7.17	<0.001	-11.07	-3.28
black	-3.45	<0.001	-5.30	-1.61
christian	-0.68	0.48	-2.58	1.22
gpaHI	2.60	<0.001	0.77	4.43
junior	0.27	0.85	-2.47	3.01
senior	1.32	0.14	-0.44	3.08
constant	52.41	<0.001	49.42	55.40
		LnSigma		
Internet	-0.19	0.05	-0.38	-0.00
female	-0.04	0.69	-0.24	0.16
asian	1.05	<0.001	0.66	1.44
black	0.56	<0.001	0.28	0.84
christian	0.16	0.25	-0.11	0.44
gpaHI	-0.30	<0.001	-0.48	-0.12
junior	0.01	0.94	-0.27	0.29
senior	-0.13	0.22	-0.34	0.08
constant	1.75	<0.001	1.40	2.10

Table 3.32: Effective Literacy of Access to the Internet With and Without Caveat, for the Question about the Remaining Life for Women, New Labels

Correct answer: 61.7 N=210 GSU undergraduates
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Average)
p-value for test of hypothesis that internetNC and internetC are jointly zero is < 0.001 (Overall)

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
internetNC	2.56	0.11	-0.61	5.74
internetC	3.89	<0.001	2.11	5.67
female	0.15	0.87	-1.63	1.94
asian	-7.39	<0.001	-11.33	-3.45
black	-3.48	<0.001	-5.29	-1.67
christian	-0.91	0.30	-2.63	0.81
gpaHI	2.51	<0.001	0.69	4.32
junior	0.54	0.70	-2.24	3.33
senior	1.17	0.19	-0.58	2.92
constant	52.55	<0.001	49.70	55.39
		LnSigma		
internetNC	-0.10	0.49	-0.39	0.19
internetC	-0.24	0.02	-0.45	-0.03
female	-0.05	0.62	-0.26	0.15
asian	1.09	<0.001	0.71	1.46
black	0.55	<0.001	0.28	0.83
christian	0.22	0.08	-0.03	0.46
gpaHI	-0.30	<0.001	-0.48	-0.11
junior	-0.00	1.00	-0.29	0.29
senior	-0.12	0.23	-0.32	0.08
constant	1.72	<0.001	1.40	2.03

Table 3.33: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Women, New Labels

Bias is relative to the correct answer: 61.7
Additional imprecision is relative to the average imprecision: 9.35
Sample sizes: 106 Individuals and 104 Internet
Literacy index L = 0.27 (Individuals) and 0.51 (Internet)
Literacy index W = 0.22 (Individuals) and 0.53 (Internet)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
internetNC	-8.97	<0.001	-11.92	-6.02
internetC	-7.84	<0.001	-9.82	-5.86
female	-9.94	<0.001	-11.40	-8.47
asian	-11.90	<0.001	-15.11	-8.69
black	-10.23	<0.001	-11.62	-8.84
christian	-9.81	<0.001	-11.22	-8.40
gpaHI	-8.64	<0.001	-10.17	-7.11
junior	-8.39	<0.001	-10.50	-6.28
senior	-9.10	<0.001	-11.62	-6.59
		Additional Imprecision of Beliefs		
internetNC	-0.43	0.37	-2.57	1.71
internetC	-0.46	0.37	-2.72	1.80
female	0.23	0.38	-1.12	1.59
asian	1.82	0.15	-0.76	4.39
black	-0.19	0.38	-1.28	0.91
christian	-0.04	0.40	-1.30	1.22
gpaHI	-0.46	0.35	-2.19	1.26
junior	-1.05	0.27	-3.35	1.26
senior	0.49	0.37	-1.95	2.94

Table 3.34: Pooled Measures of *L* and *W* Indices, Individual "Control" and Internet Scaffold

			Individual Literacy Measures		Internet Literacy Measures	
Question	Type	Correct Answer	*L*	*W*	*L*	*W*
fin1 - Savings Account 2%	Numeracy	$110.41	0.53	0.50	0.68	0.67
fin2 - Social Security Start Age	Procedural	62	0.26	0.21	0.87	0.84
fin5 - Medicare Eligibility	Procedural	65	0.16	0.16	0.82	0.80
fin7 - Real Interest Rate	Numeracy	$98.98	0.28	0.29	0.28	0.28
fin9 - Savings Horizon	Numeracy	4 months	0.77	0.74	0.84	0.83
fin10 - Stolen Credit Card	Procedural	$50	0.06	0.06	0.66	0.64
fin11 - Stolen Debit Card	Procedural	$500	0.16	0.12	0.61	0.59
fin13 - Nominal Interest	Numeracy	5%	0.80	0.76	0.81	0.83
fin14 - Interest Rate	Numeracy	$102	0.74	0.68	0.72	0.71
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22	0.18	0.48	0.42
fin16 - Remaining Life for Women	Longevity Risk	61.7 years	0.27	0.22	0.51	0.53

*Pooled literacy measures of *L* and *W* are initally reported for each question in their respective section. Chapter 2 for indvidual measures and Chapter 3 for Internet.

Table 3.35: Pooled Measures of *L* Index, Comparing the Internet Scaffold to Individual Measures

			Individual Literacy Measures		Internet Literacy Measures	
Question	Type	Correct Answer	*L*		*L*	+/-
fin1 - Savings Account 2%	Numeracy	$110.41	0.53		0.68	0.15
fin2 - Social Security Start Age	Procedural	62	0.26		0.87	0.61
fin5 - Medicare Eligibility	Procedural	65	0.16		0.82	0.66
fin7 - Real Interest Rate	Numeracy	$98.98	0.28		0.28	0.00
fin9 - Savings Horizon	Numeracy	4 months	0.77		0.84	0.07
fin10 - Stolen Credit Card	Procedural	$50	0.06		0.66	0.60
fin11 - Stolen Debit Card	Procedural	$500	0.16		0.61	0.45
fin13 - Nominal Interest	Numeracy	5%	0.80		0.81	0.01
fin14 - Interest Rate	Numeracy	$102	0.74		0.72	-0.02
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22		0.48	0.26
fin16 - Remaining Life for Women	Longevity Risk	61.7 years	0.27		0.51	0.24

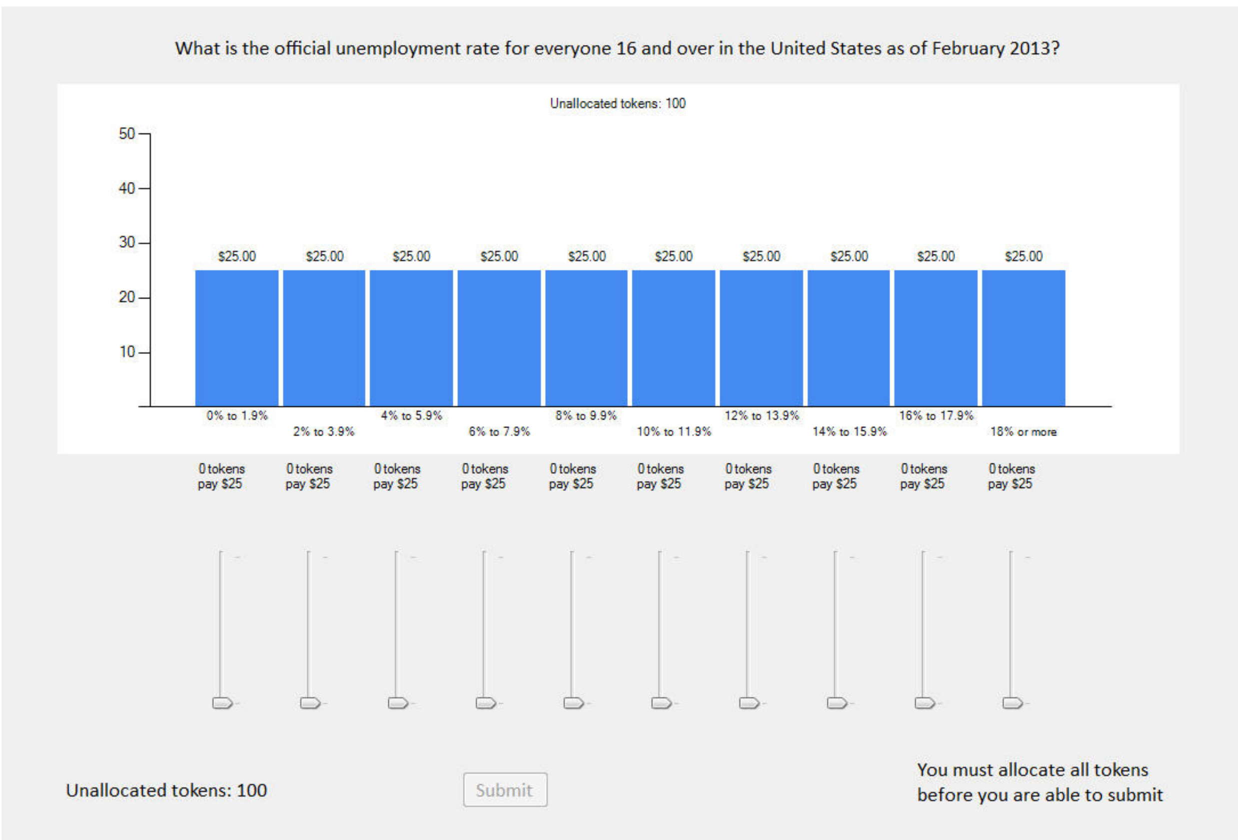
*Pooled literacy measures of L are initally reported for each question in their respective section. Chapter 2 for indvidual measures and Chapter 3 for Internet measures. The "+/-" column is the difference of the treatment compared to the Individual "control" subjects.

Appendix E: Experimental Instructions

1. Instructions for Belief Elicitation, Lab Sessions Internet with No Caveat

This is a task where you will be paid according to how accurate your beliefs are about certain things. You will be presented with 11 questions and asked to place some bets on your beliefs about the answers to each question. You will actually 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 everyone has made all decisions, and that process is explained below.

Here is an example of what the computer display of a question might look like.

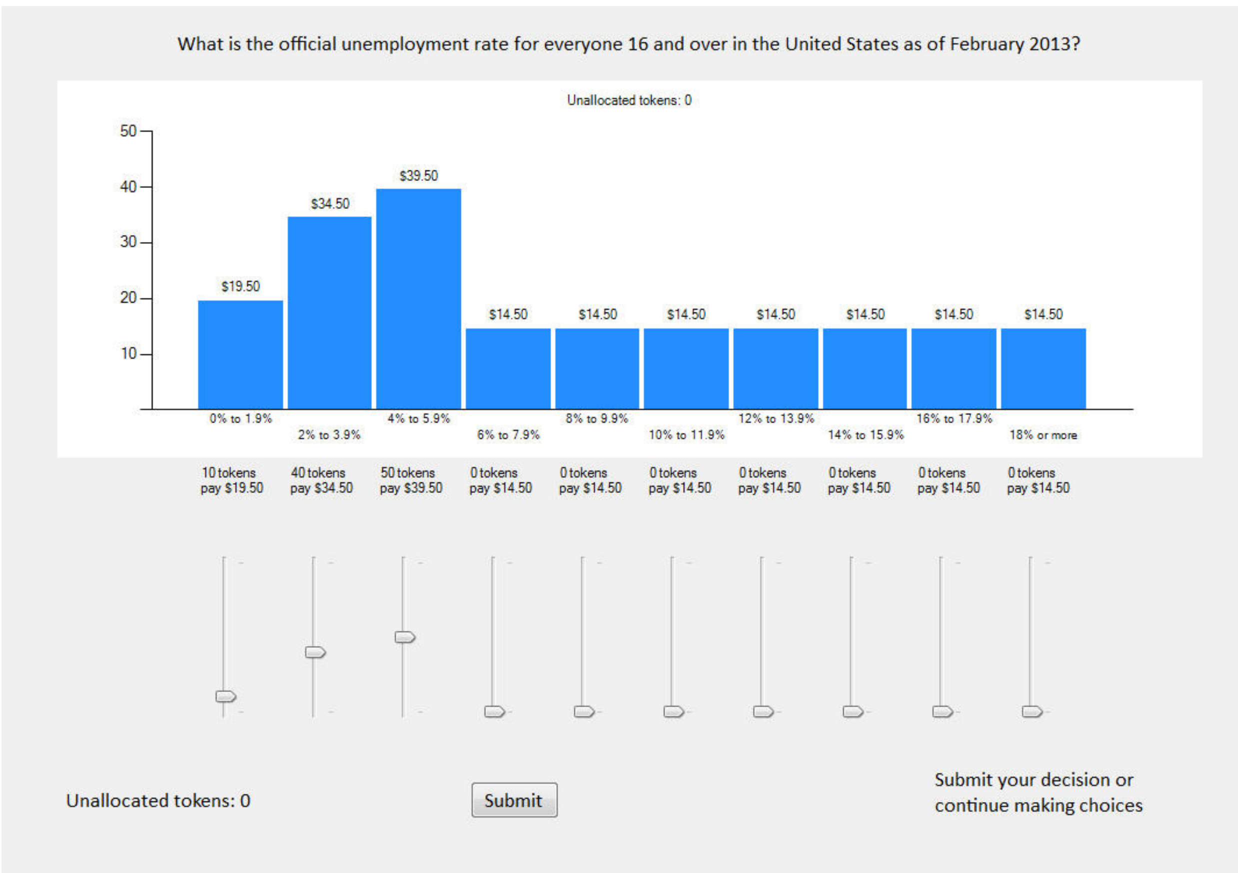


The display on your computer will be larger and easier to read. You have 10 sliders to adjust, shown at the bottom of the screen, and you have 100 tokens to allocate across the sliders. Each slider allows you to allocate tokens to reflect your belief about the answer to this question. You must allocate 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. Please note that the bars above each slider correspond to that particular slider. In the above example, the tokens you allocate to each bar will naturally reflect your beliefs about the official unemployment rate for everyone 16 and over in February 2013. The first bar corresponds to your belief that the unemployment rate is between 0% and 1.9%. The second bar corresponds to your belief that the unemployment rate is between 2% and 3.9%, and so on. Each bar shows the amount of money you could earn if the official unemployment rate is in the interval shown under the bar.

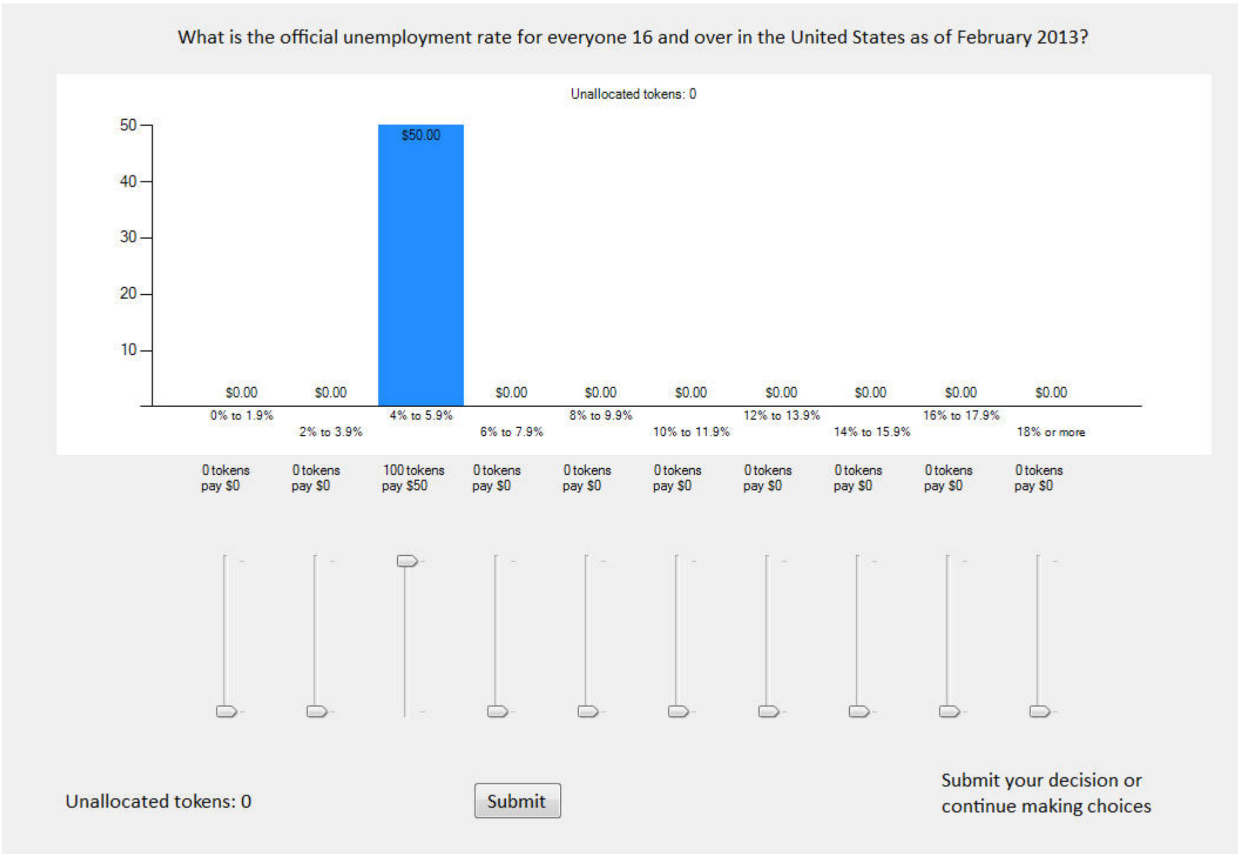
To illustrate how you use these sliders, suppose you think there is a fair chance the unemployment rate is just under 5%. Then you might allocate the 100 tokens in the following way: 50 tokens to the interval 4% to 5.9%, 40 tokens to the interval 2% to 3.9%, and 10 tokens to the interval 0% to 1.9%. So you can see in the picture below that if indeed the unemployment rate is between 4% and 5.9% you would earn $39.50. You would earn less than $39.50 for any other outcome. You would earn $34.50 if the unemployment rate is between 2% and 3.9%, $19.50 if it is between 0% and 1.9%, and for any other unemployment rate you would earn $14.50.



You can adjust the allocation as much as you want to best reflect your personal beliefs about the unemployment rate.

Your earnings depend on your reported beliefs and, of course, the true answer. For instance, suppose you allocated your tokens as in the figure shown above. The true unemployment rate is actually 7.7%, according to the *Bureau of Labor Statistics*. So if you had reported the beliefs shown above, 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 unemployment rates between 4% and 5.9%. Then you would have faced the earnings outcomes shown below.



Note the "good news" and "bad news" here. If the unemployment rate is indeed between 4% and 5.9%, you earn the maximum payoff, shown here as $50. But the true unemployment rate is 7.7%, so you would have earned nothing in this task.

Using the methodology described above we will ask you 11 questions today. Immediately before each section starting, you will be given that section's questions on a piece of paper. You can write on these papers as needed. After everyone in the experiment has the appropriate question sheet in hand we will then open an Internet browser on your computer remotely, and allow access to the Internet for 15 minutes. It is a timed 15 minutes, after which the browsers on your computer will close automatically. It is during this time that you should research the section's questions as efficiently as possible, and make notes as needed on the paper provided. Immediately after the Internet access ends, you will be asked to respond to the same questions using the computer interface.

It is up to you to use your given time on the Internet as efficiently as possible to research the questions that will be handed out in advance. You should balance the information you find on the Internet with the strength of your personal beliefs and the possibility of them being wrong. There are three important points for you to keep in mind when making your decisions:

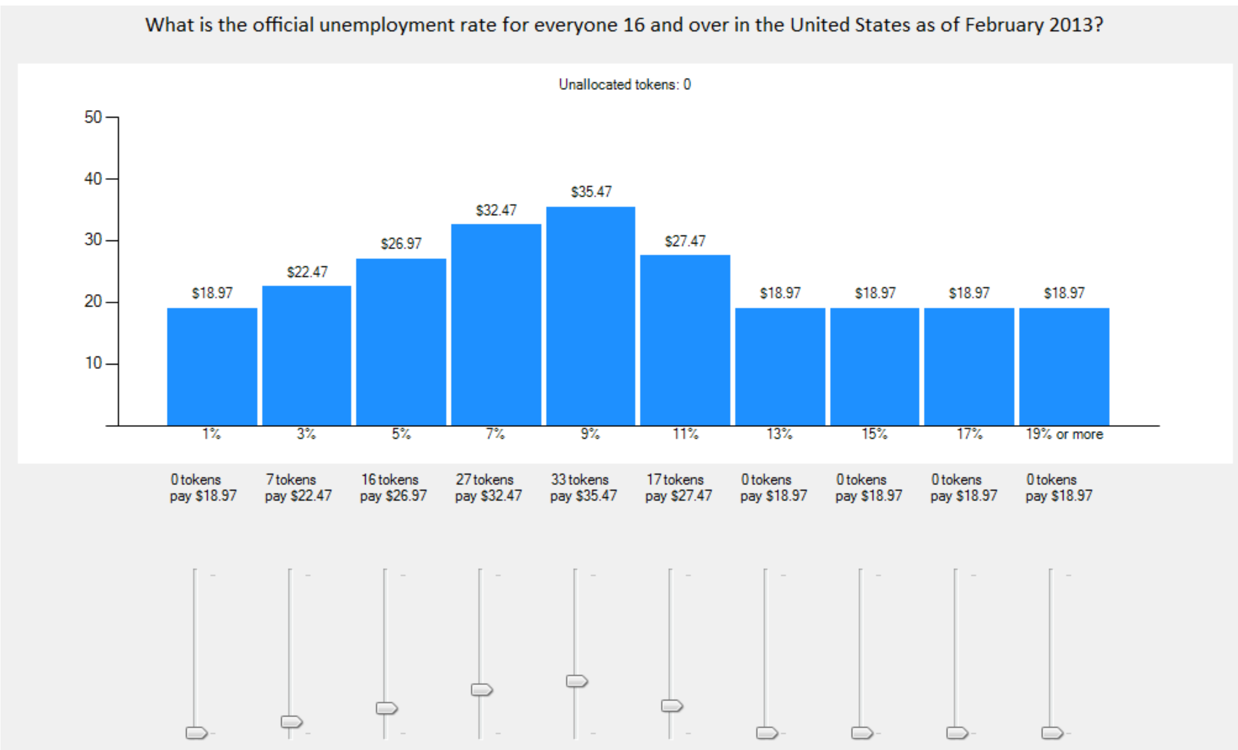
- **Your belief about the correct answer to each question is a personal judgment after being given access to the Internet, and that depends on the information you have about the topic of the question.**

- **Depending on your choices and the correct answer you can earn up to $50.**

- **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.

For some of the questions we will round the correct answer to the nearest amount shown under each bar. For example, the decision screen for the unemployment question might have shown unemployment rates of 1%, 3%, 5%, 7%, 9%, 11%, 13%, 15%, 17% and 19% or more, as shown below.



In this case, the correct answer of 7.7% would have been rounded to 7% rather than rounded to 9%, and the payment would have been $32.47.

When you are satisfied with your decisions, you should click on the **Submit** button and confirm your choices. When everyone is finished we will roll a 20-sided die to determine which question number 1 to 11 will be played out. A re-roll will occur for numbers rolled of 12 through 20. 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.

Are there any questions?

2. Instructions for Belief Elicitation, Lab Sessions Internet with Caveat

[This set of instructions was exactly the same as those listed above in this Appendix E, with the following exception shown in *italics* below starting on the third page of the instruction set. This addition was meant to remind the participants that they did not have to use the information they found on the Internet.]

Using the methodology described above we will ask you 11 questions today. Immediately before starting this task you will be given the questions we're asking you about on a piece of paper. You can write on the paper as needed. After everyone in the experiment has the question sheet in hand we will then open an Internet browser on your computer remotely, and allow access to the Internet for 15 minutes. It is a timed 15 minutes, after which the browsers on your computer will close automatically. *You are not required to use any information found on the Internet,* but it is during this time that you should research the section's questions as efficiently as possible, and make notes as needed on the paper provided. Immediately after the Internet access ends, you will be asked to respond to the same questions using the computer interface.

It is up to you to use your given time on the Internet as efficiently as possible to research the questions that will be handed out in advance. *Again, you are not required to answer the questions using the information that you find on the Internet.* You should balance the information you find on the Internet with the strength of your personal beliefs and the possibility of them being wrong. There are three important points for you to keep in mind when making your decisions…

Chapter 4 - Extended and Effective Literacy of Exogenously Formed Groups

4.1 Introduction

The concept of effective literacy is formally defined in the literature in terms of the literacy that applies to an individual in the context of and supplemented by the resources of their household. The concept was introduced by Basu and Foster (1998), as discussed in Section 1.2.1, in the context of national surveys of literacy. It introduced a new approach to evaluate the aggregate literacy level in a country or region. In the broader economics of the household literature this is related to the bargaining problem within the household; an extensive survey of this literature can be found in Browning, Chiappori, and Weiss (2014). Effective literacy is defined in terms of a Leontief production technology: the effective literacy of a group of N individuals is the literacy of the maximally literate member of that group. If this measure is meant to define an upper bound, or a setting in which the maximally literate member of the group has dictatorial bargaining power, this would be an appropriate assumption. However, if we are interested in understanding the literacy of households, committees, or villages, then we must clearly worry about the "literacy production function" that takes as inputs the literacy of each of the members of those groups and determines what the joint decision is. This idea immediately resonates with emphasis in the household economics literature as well as the development literature on intra-household bargaining power. It is not at all obvious that the most literate person in the group will be able to have an effect on the group decision that is commensurate with their literacy, rather than reflect their bargaining power, which may be limited.

One could go directly into the field and examine the effective financial literacy of endogenously formed groups, also estimating models of bargaining power, but this course would bring with it a large number of potential confounds. We instead begin in the laboratory to examine effective financial literacy of *exogenously* formed groups of 2 who are asked to answer the battery of financial literacy questions jointly. Those results will be compared to the control treatment, evaluated in Chapter 2. Since the groups are constructed by randomly pairing students in the laboratory, we assume them to be strangers, so the results presented here can be viewed as looking at the pure group effect on effective literacy.

4.2 Literature Review

The concept of scaffolding was introduced in Chapter 1 to span a wide range of ways in which an individual agent might "exploit the external structure." Language and social interaction between people has been a prominent example of that structure from the earliest days of the scaffolding literature. Similar ideas have appeared in different guises in various disciplines: the literature on the "wisdom of the crowds" and "groupthink," reviewed in Section 1.2.1, are prime examples. The wisdom of the crowd results when social scaffolding produces a net gain in effective literacy, and groupthink arises when social scaffolding produces a net loss in effective literacy.

4.3 Experimental Design

The data presented in Chapters 2, 3, and 4 are obtained in a common manner from experiments conducted with students at Georgia State University (GSU) in the Experimental Economics Center (ExCEN) over the period from November 2013 to March 2016. See Section 2.4 for a full description of the data.

The results presented in Chapter 4 build on Chapter 2 and assess the effect that access to another person, as part of an exogenously formed group of two individuals, has on the financial literacy of the group. There were 7 sessions run in total involving groups; 2 sessions using the initial labels and 5 sessions using new labels. Again we focus our analysis on the 5 group sessions using the new labeling scheme to then compare the results with our control treatment of individuals presented in Chapter 2. The experiments involving groups presented with new labels were run in Sessions 21, 22, 23, 26, and 27 and had the following number of participants, respectively: 16, 22, 22, 28, and 34. Thus there were 122 individual participants in the group sessions in total, consisting of 61 exogenously formed groups with two members in each.

The sessions involving groups were run logistically in the same manner as the sessions involving individuals without Internet access. The one variation in protocol was that subjects in the group treatment were randomly assigned to exogenously formed groups comprised of two individuals and required to make a joint decision when answering the literacy questions. After the groups were assigned the experimenters allowed 5 minutes of introduction and casual talk to be exchanged between group members before continuing through the experimental instruction set with them. The instructions did not provide any guidance as to how the group was to arrive at the allocation of tokens to the belief questions. The experimental instruction text generally referred to the group as an agent, leaving it to the individuals to decide how to make the group decisions. The complete set of instructions for groups can be found in Appendix F.

In the event that a correct answer was realized in the belief task, each member in the group would receive the full payment amount that was listed on the joint decision screen. This was done to keep incentive levels the same throughout all sessions.

The two other tasks in the session were completed individually by every participant and not as members of a group. One of these tasks is to complete a standard socio-demographic survey. The other task is used to estimate a measure of an individual's atemporal risk aversion using a battery of binary choices over risky lotteries. Following Harrison and Ulm (2016) we then use each individual subject's risk preferences to recover latent beliefs using the scoring rules employed in the belief elicitation task for the various financial questions.

In Chapters 2 and 3 the application to employ this method is straightforward as those treatments deal solely with individual participants making their own decisions in both the belief task and the risk task. However, the method is not as direct when recovering latent beliefs belonging to a group making a joint decision and having access to only individual risk measures. The method that is adopted here assigns the raw token allocation given by the group to each individual in that group, and then employs the recovery of latent beliefs for each participant using their individual measure of risk. This acknowledges that a joint decision was made by group members and that we then recover the latent beliefs of the group based on each member's individual risk aversion. Hence we are recovering the latent beliefs that each subject agreed to when they agreed to a joint allocation of tokens. If the individuals within a group have different risk preferences, which is the usual case, then the recovered beliefs from the agreed group allocation of tokens will differ from each other.

4.4 Results

The results from these experiments echo the experiments on the effect of Internet access. We find that having access to interacting with another person generally increases literacy.

Consider responses to the question about the savings account with 2% interest left to grow for five years shown in Figure 4.1, Table 4.1, and Table 4.2. From Figure 4.1 we see that the overall average beliefs for individuals was \$110.7 and for groups was \$110.5, with a p-value = 0.630 on a hypothesis testing if there is a difference in average responses. Here we *cannot* reject the hypothesis that the average response of individuals and groups are statistically different from one another. We further observe from Figure 4.1 that the modal responses for both individuals and groups are in the bin that contains the true answer and that the underlying distributions are shaped similarly. However, groups appear to allocate more tokens on average to the bin containing the true answer than individuals. We can read off the average token allocations directly from the figure, or see a numeric representation of them using the L literacy index shown at the top of Table 4.2. Recall that the L literacy index is calculated by summing the average tokens in the bin containing the correct answer and dividing by 100. Thus the reported L literacy indices of 0.53 for individuals and 0.67 for groups in Table 4.2, correspond to an average token allocation of 53 and 67, respectively, in the bin containing the correct answer in Figure 4.1. Thus we see that bias is no different between individuals and groups, but confidence is, leading to an increase in literacy for groups.

The body of Table 4.1 displays the results from the interval regression with additional covariates. In the top panel of Table 4.1 we see that the overall average belief was \$110.94; as usual, the constant refers to the "omitted category," which is a White male underclassmen at GSU. We observe that students with a reported high grade point average hold beliefs that are on average \$0.91 lower than the omitted category, and that this difference is statistically significantly different from zero (p-value < 0.001).

In the bottom panel of Table 4.1, the interpretation of the coefficients for dispersion is again immediate, because the dependent variable is in natural log units. Thus the coefficient estimate β implies a $\beta \times 100$ percent change in the standard deviation of beliefs, σ, for a one unit change in that covariate. So the statistically significant effect (p-value < 0.001) of being in a group is interpreted as a 51% reduction in the standard deviation of beliefs compared to the omitted category.

The information in Table 4.1 below the title is generally self-explanatory. The p-value for group refers to a test of the hypothesis that the effect on the average and the standard deviation are jointly zero, and complements the corresponding hypothesis tests of the average or the standard deviation in the body of the table. Here the overall effect of being in a group is statistically significant and different from zero (p-value < 0.001). Figure 4.1 shows the distribution of reports for each bin, side-by-side for individuals and those in groups, allowing a visual interpretation of differences in reports.

Table 4.2 provides an evaluation of the bias and imprecision of beliefs that is associated with demographics and being in a group, and is different from the estimates underlying Table 4.1. In the top panel of Table 4.2 we estimate the "total effect" of the covariate, and then compare the estimated average belief to the true value. The bottom panel of Table 4.2 compares the estimated standard deviation for each covariate to the overall pooled standard deviation. Thus we measure bias for the average, and relative imprecision for the standard deviation. These are not marginal effects: for each covariate of interest it is as if we use the average value of all other covariates for that covariate of interest. For females, for instance, the top panel of Table 4.2 shows that the difference between the average belief of a female is \$0.16 greater than the correct answer of $110.41, but this difference is not statistically significantly different from zero (*p*-value = 0.31). This is different from Table 4.1, where the coefficient shows the difference compared to the omitted category. Thus the *p*-values in Table 4.2 tell us what we want to know about bias for females, and the *p*-values in Table 4.1 tell us what we want to know about marginal differences in beliefs for females compared to the omitted category.

With this background, Table 4.2 shows us that there is no statistically significant evidence of bias from the correct answer at the 5% significance level for any demographic covariates, on whether or not in a group. In the bottom panel of Table 4.2 we see at the 5% significance level that students with junior class standing and those in a group exhibit less imprecision of beliefs compared to the average imprecision. For this question we see that "most GSU students get it and those in groups get it more because of increased precision." This is also shown by the higher levels of the literacy indexes L and W of groups compared to individuals.

Responses to the question about the age at which one can start taking out Social Security benefits are displayed in Figure 4.2, Table 4.3, and Table 4.4. Both the individual subjects and the group subjects exhibit a wide range of beliefs about this question, for which the true answer is 62. This correct answer is (just) the modal belief of individuals, but even in that case we observe significant weight being attached to the next two bins corresponding to 64 and 66. The subjects responding in groups report similarly by placing the most weight in the first three bins, but the weight placed on 62 drops compared to individual subjects. The net effect is a decline in literacy for groups, using our two measures (Table 4.4). We see in Table 4.3 that the overall effect of being in a group is statistically significantly different from zero (*p*-value < 0.001) even though the results in the table text are mixed. Here the marginal effect of being in a group on the average response is not statistically significantly different from the omitted category (top panel Table 4.3). However, being in a group implies a 38% reduction in the standard deviation of beliefs compared to the omitted category, which is statistically significant at the 5% level (bottom panel Table 4.3). The top panel of

Table 4.4 shows us that all of the covariates are statistically significantly biased from the correct answer (*p*-value < 0.001). The bottom panel of Table 4.4 shows that groups and upperclassmen hold more precise beliefs than the average, although with the significant bias that both exhibit this is again not a welcome effect in terms of the quality of likely decisions contingent on these beliefs.

Responses to the comparable question about the age at which eligibility for Medicare starts are shown in Figure 4.3, Table 4.5, and Table 4.6. The correct response is 65, and that is the modal response of groups but not of individuals (Figure 4.3). We observe that individuals instead place the greatest weight on the age being 55, with other spikes at 61 and 65. Hence there is a noticeable increase in literacy according to the two measures when making decisions in a group (Table 4.6): when considering the L index, literacy increases 16 percentage points to 0.31 over the individuals measured. We find that the overall effect of being in a group is not statistically significantly different (*p*-value = 0.111) from the omitted category, even though the marginal effect of being in a group on the average belief is +1.76 was statistically significant at the 5% level (Table 4.5). The results from Table 4.6 are similar to the previous questions and again shows that all of the covariates are statistically significantly biased from the correct answer (*p*-value < 0.001). However, no covariates exhibit differing levels of imprecision relative to the average imprecision (bottom panel of Table 4.6).

Responses to the question about the real interest rate are collated in Figure 4.4, Table 4.7, and Table 4.8. They reveal that both individuals and groups have the correct answer, $98.98, as their modal response, and that groups place much more weight on the correct value than individuals (Figure 4.4). The top panel of Table 4.7 shows that groups, females, Christians, and students with a reported higher grade point average hold different beliefs than the omitted category, and that these differences are statistically significant at the 5% level. Further, the text below the title of Table 4.7 shows the overall effect of being in a group is statistically significant and different from zero (*p*-value < 0.001). Groups place more weight on the true answer as shown by an L index which is 17 percentage points higher than individuals (Table 4.8). Again, all covariates are statistically significantly biased from the correct answer (*p*-value < 0.001) with no covariates exhibiting differing levels of imprecision relative to the average imprecision at the 5% significance level (Table 4.8).

Responses to the savings horizon question have some of the highest literacy measures and are shown in Figure 4.5, Table 4.9, and Table 4.10. Both individuals and groups report modal responses in the correct belief bin of 4 months (Figure 4.5). The groups place more weight on the correct bin, and that is about the only difference between the two histograms. Table 4.9 shows that the overall result and marginal effect of being in a group are not statistically significant. Table 4.10 shows that only those in groups and Asians were unbiased from the correct answer at the 5% significance level (top panel) and that no covariate exhibited differing levels of imprecision (bottom panel). The implication is an increase in literacy for groups according to both of our measures (Table 4.10).

Responses to the question about liability for debt incurred on a stolen credit card are shown in Figure 4.6, Table 4.11, and Table 4.12. The correct answer is $50, but both individuals and groups place significant weight on the belief that it is zero; the only difference is that groups place greater weight on that biased belief of zero (Figure 4.6). We observe a significant bias from being in a group, and some evidence of increased confidence. But the beliefs are so poorly calibrated to start with, given the weight attached by both treatments to zero, that there cannot be much effect on literacy from being in a group since it is already close to being zero when responding as an individual (Table 4.12).

A similar pattern is exhibited in responses to the comparable question about liability from debt incurred on a debit card that is stolen, shown in Figure 4.7, Table 4.13, and Table 4.14. The difference is that individuals and groups put some weight on the correct belief of $500, and groups exhibit some evidence of increased confidence by allocating virtually zero tokens to the last 5 bins (Figure 4.7), leading to a modest increase in literacy from being in a group using either measure (Table 4.14). The additional confidence of groups is statistically significant (*p*-value < 0.001) whether measured by the marginal effect

(Table 4.13) or the total effect (Table 4.14). Given that groups are unbiased from the correct answer, the additional precision of their beliefs enhances literacy.

Responses to the question about nominal interest are shown in Figure 4.8, Table 4.15, and Table 4.16. We observe one clear mode at the correct belief of 5% for both individuals and groups, with more weight being placed in the correct bin by groups (Figure 4.8). There is a high level of corresponding literacy for both individuals and groups given by the *L* indices of 0.80 and 0.93, respectively (Table 4.16). Here is a case in which "everybody gets it on average" and the marginal effect (Table 4.15) of being in a group leads to greater confidence in beliefs.

Responses to the question about the value of a $100 principal in one year with an interest rate of 2% are shown in Figure 4.9, Table 4.17, and Table 4.18. The findings are similar to those applying to the previous question. Both individuals and groups exhibit a modal belief in the correct belief bin of $102, with more weight being placed on the correct bin by groups (Figure 4.9). There is a high level of literacy for both individuals and groups shown by the *L* indices of 0.74 and 0.90, respectively (Table 4.18). Again, here is a case that "everybody gets it on average" and the marginal effect (Table 4.17) of being in a group leads to an increase in confidence in beliefs.

The last two questions refer to beliefs about longevity risk for men and women. Responses are displayed in Figure 4.10, Table 4.19, and Table 4.20 for men, and Figure 4.11, Table 4.21, and Table 4.22 for women. The patterns in Figure 4.10 and Figure 4.11 are similar for beliefs about men and women: individuals hold more diffuse beliefs. Although the beliefs of individuals assign greater weight to the responses closer to the correct response, they leave more weight on the lower, erroneous responses compared to groups. Group responses have a mode at the true belief and a test of differences in average responses between individuals and groups confirms that these two treatments hold different average beliefs for both men (p-value = 0.026, Figures 4.10) and women (p-value < 0.001, Figure 4.11) In both cases there are significant biases in beliefs, and in both cases this comes with an interaction between being in a group and demographics. The marginal effect of being in a group for the belief about male longevity is +2.40 and statistically significant (p-value = 0.03), but the total effect of being in a group for the belief about male longevity is -7.06 and also statistically significant (p-value < 0.001). The corresponding group biases for the belief about female longevity are +2.95 and -8.45, and even more statistically significant. For both questions we see that groups do better than the omitted category, but are still biased away from the correct answer. The obvious demographic interaction, *a priori*, is with gender, and this appears to be the case for the belief about the longevity of men and women: the marginal effect of being female is the only demographic that was found to be statistically significant (p-value < 0.001) for both questions and resulted in a coefficient of -3.95 when responding about male longevity and a coefficient of -4.29 when responding about female longevity. For both questions the total effect of each demographic was negatively biased away from the correct answer and each estimate is statistically significant (p-value < 0.001). Given the relatively diffuse beliefs for individuals and the increased confidence exhibited from being in a group, there is a slight positive change in literacy according to our two measures when responding in a group (Table 4.20 and Table 4.22).

4.5 Conclusions

Table 4.23 displays a summary view of the pooled L and W indices for both our control subjects and those generated under the Group scaffold for each question. Table 4.24 then focuses only on the L index and assesses the changes when comparing the Group scaffold to the Individual literacy measures of our control subjects. In the fourth column of Table 4.24 are the pooled measures of the L index for individuals responding with only their private literacy. In the sixth column of Table 4.24 are the L index measures for the Group. Column seven is the "+/-" difference between the L index for those in a Group compared to individuals. We see "+/-" values ranging from -0.05 for fin2, "the Social Security start" question, to +0.17 for fin7, "the real interest rate" question. The overall conclusion from these results is that effective literacy, at least in exogenously formed groups of two, results in generally higher literacy compared to individuals. Further, it appears that the greatest enhancements to literacy under this scaffold are related to numeracy. Hence we again find that the social scaffold of being able to pool information with another person generally leads to enhanced literacy compared to when individuals must form beliefs alone. We also therefore demonstrate that effective literacy of a group generally exceeds the literacy of "statistically equivalent" individuals.

An open question is whether larger, exogenously formed groups would do even better, and that remains an obvious extension for future research. A separate question is whether "naturally formed" groups, such as couples living together, would exhibit greater effective literacy than comparable individuals living alone. This question is addressed in chapter 5.

Figure 4.1: Elicited Beliefs about
The Savings Account with 2% Interest Question,
Comparing Group and Individual Responses, New Labels

Group average: \$110.5 (N=61)   Individual average: \$110.7 (N=106)
p-value on test of difference in averages = 0.630
True response was \$110.41

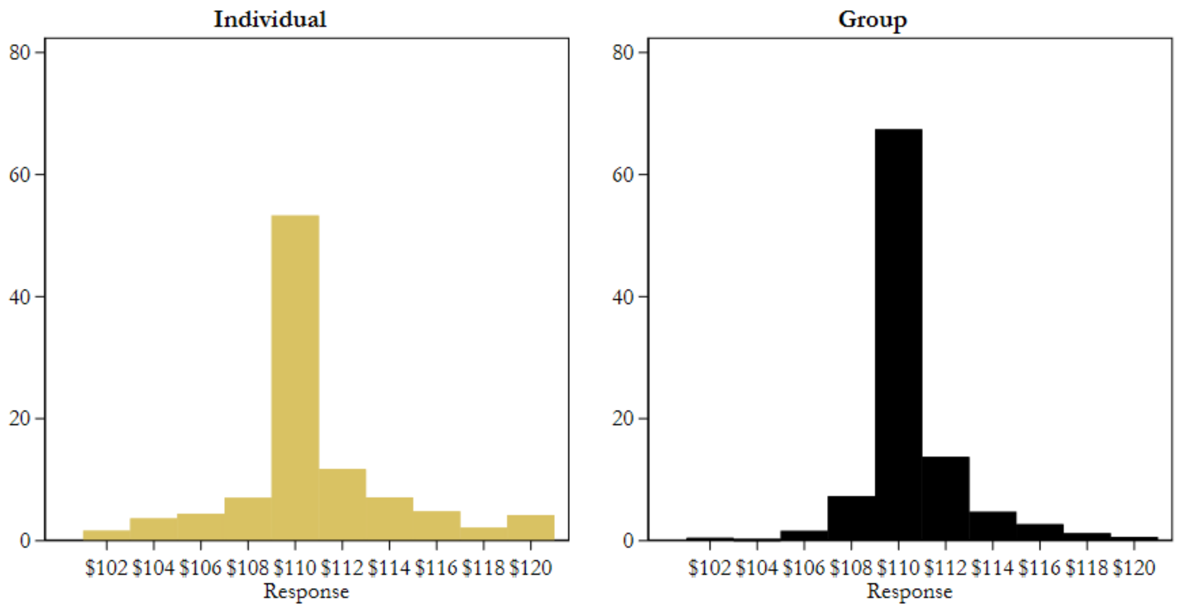


Table **4.1: Effective Literacy of Exogenous Groups**
for the Question about the Savings Account with 2% Interest, New Labels

Correct answer: \$110.41 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is < 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-0.13	0.63	-0.66	0.40
female	-0.07	0.81	-0.62	0.48
asian	0.76	0.10	-0.14	1.66
black	0.09	0.77	-0.53	0.71
christian	0.21	0.55	-0.48	0.89
gpaHI	-0.91	<0.001	-1.58	-0.23
junior	-0.32	0.40	-1.05	0.42
senior	0.01	0.97	-0.67	0.70
constant	110.94	<0.001	109.84	112.03
		LnSigma		
groups	-0.51	<0.001	-0.76	-0.26
female	0.18	0.18	-0.08	0.44
asian	0.36	0.06	-0.02	0.75
black	0.39	<0.001	0.09	0.69
christian	0.19	0.15	-0.07	0.46
gpaHI	-0.16	0.22	-0.41	0.09
junior	-0.07	0.64	-0.38	0.24
senior	-0.11	0.54	-0.46	0.24
constant	0.68	<0.001	0.33	1.04

Table 4.2: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Account with 2% Interest, New Labels

Bias is relative to the correct answer: $110.41
Additional imprecision is relative to the average imprecision: 2.95
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.53 (Individuals) and 0.67 (Groups)
Literacy index W = 0.50 (Individuals) and 0.62 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	0.12	0.31	-0.23	0.48
female	0.16	0.28	-0.22	0.55
asian	0.52	0.12	-0.14	1.17
black	0.22	0.24	-0.21	0.65
christian	0.27	0.16	-0.13	0.67
gpaHI	-0.22	0.22	-0.60	0.17
junior	-0.11	0.38	-0.77	0.56
senior	0.07	0.39	-0.46	0.60
		Additional Imprecision of Beliefs		
groups	-0.94	<0.001	-1.37	-0.51
female	0.06	0.38	-0.35	0.47
asian	-0.40	0.15	-0.96	0.15
black	0.25	0.22	-0.19	0.69
christian	0.21	0.24	-0.21	0.63
gpaHI	-0.34	0.19	-0.87	0.20
junior	-0.62	0.05	-1.21	-0.02
senior	-0.34	0.31	-1.27	0.58

Figure 4.2: Elicited Beliefs about The Social Security Start Question, Comparing Group and Individual Responses, New Labels

Group average: 65.2 years (N=61) Individual average: 65.8 years (N=106)
p-value on test of difference in averages = 0.105
True response was 62

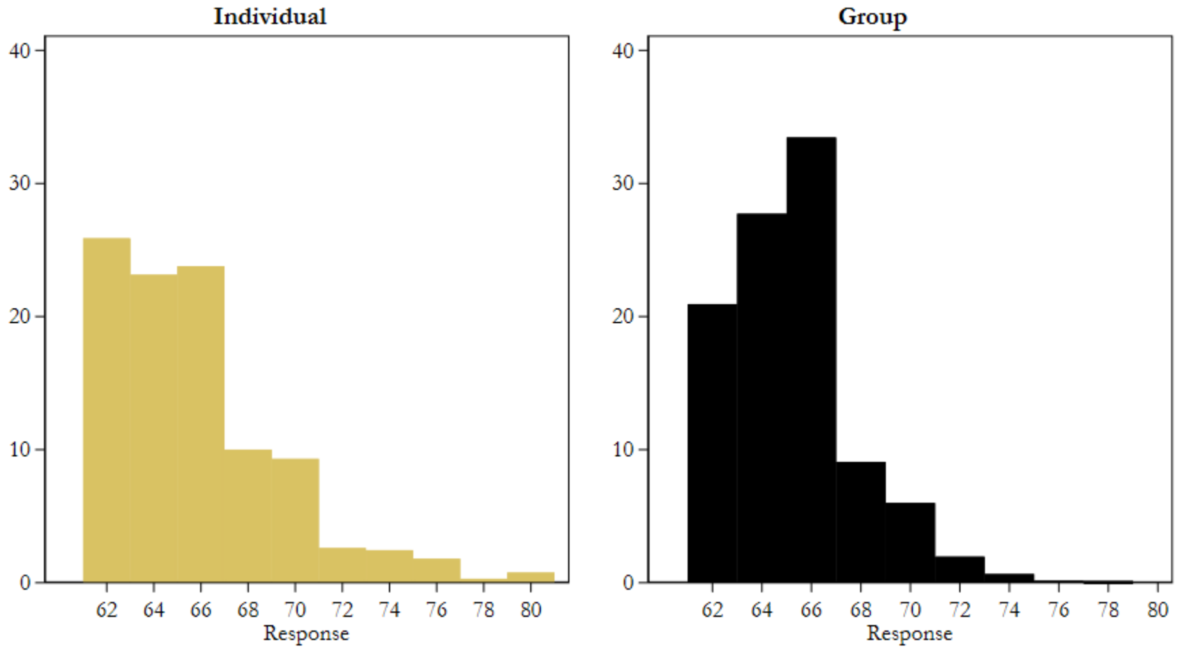


Table 4.3: Effective Literacy of Exogenous Groups for the Question about the Social Security Start Age, New Labels

Correct answer: 62 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is < 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-0.57	0.11	-1.25	0.12
female	0.24	0.55	-0.55	1.03
asian	-0.94	0.23	-2.46	0.59
black	-0.74	0.32	-2.21	0.72
christian	-0.09	0.86	-1.03	0.86
gpaHI	0.12	0.75	-0.64	0.88
junior	-0.48	0.31	-1.41	0.45
senior	-0.98	0.02	-1.77	-0.18
constant	66.56	<0.001	65.28	67.85
		LnSigma		
groups	-0.38	<0.001	-0.56	-0.20
female	0.03	0.79	-0.18	0.23
asian	-0.19	0.26	-0.51	0.14
black	-0.08	0.51	-0.32	0.16
christian	-0.05	0.64	-0.26	0.16
gpaHI	0.20	0.03	0.02	0.38
junior	-0.21	0.06	-0.42	0.01
senior	-0.27	<0.001	-0.46	-0.09
constant	1.34	<0.001	1.04	1.65

Table 4.4: Literacy Bias and Imprecision, by Demographics, for the Question about the Social Security Start Age, New Labels

Bias is relative to the correct answer: 62
Additional imprecision is relative to the average imprecision: 3.24
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.26 (Individuals) and 0.21 (Groups)
Literacy index W = 0.21 (Individuals) and 0.18 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	3.22	<0.001	2.68	3.77
female	3.63	<0.001	3.16	4.10
asian	3.44	<0.001	2.54	4.34
black	3.50	<0.001	3.03	3.96
christian	3.54	<0.001	3.10	3.99
gpaHI	3.72	<0.001	3.14	4.31
junior	3.22	<0.001	2.49	3.94
senior	2.98	<0.001	2.38	3.58
		Additional Imprecision of Beliefs		
groups	-0.75	<0.001	-1.14	-0.36
female	0.00	0.40	-0.39	0.39
asian	-0.10	0.39	-0.86	0.65
black	0.03	0.39	-0.39	0.45
christian	-0.01	0.40	-0.40	0.37
gpaHI	0.25	0.22	-0.20	0.71
junior	-0.59	0.02	-1.07	-0.11
senior	-0.52	0.01	-0.90	-0.13

Figure 4.3: Elicited Beliefs about The Medicare Eligibility Question, Comparing Group and Individual Responses, New Labels

Group average: 59.0 years (N=61)   Individual average: 57.3 years (N=106)
p-value on test of difference in averages = 0.036
True response was 65

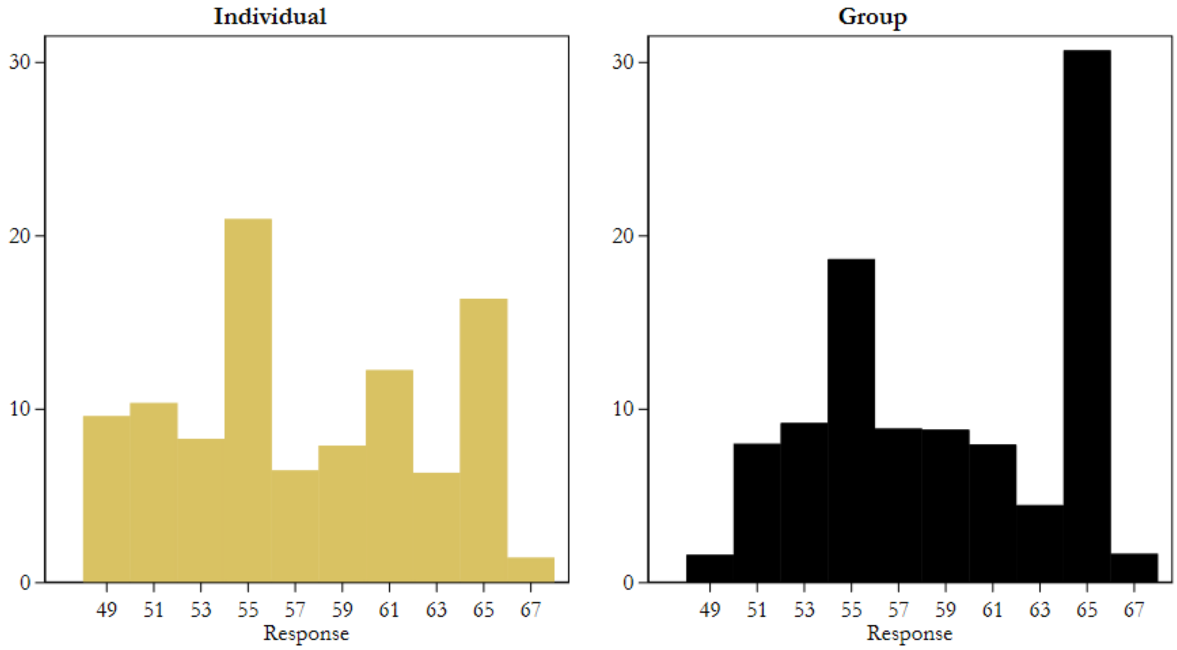


Table 4.5: Effective Literacy of Exogenous Groups for the Question about Medicare Eligibility, New Labels

Correct answer: 65 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.111 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	1.76	0.04	0.12	3.41
female	0.10	0.91	-1.61	1.81
asian	-0.01	1.00	-2.62	2.60
black	-1.04	0.35	-3.23	1.15
christian	-0.14	0.87	-1.86	1.58
gpaHI	-0.21	0.79	-1.80	1.38
junior	-0.42	0.67	-2.31	1.47
senior	1.37	0.16	-0.56	3.29
constant	57.95	<0.001	55.40	60.50
		LnSigma		
groups	0.00	0.99	-0.12	0.12
female	-0.05	0.43	-0.16	0.07
asian	0.07	0.47	-0.12	0.26
black	0.08	0.33	-0.08	0.23
christian	-0.00	0.94	-0.13	0.12
gpaHI	-0.05	0.37	-0.16	0.06
junior	-0.08	0.29	-0.24	0.07
senior	0.02	0.81	-0.13	0.16
constant	1.65	<0.001	1.44	1.86

Table 4.6: Literacy Bias and Imprecision, by Demographics, for the Question about Medicare Eligibility, New Labels

Bias is relative to the correct answer: 65
Additional imprecision is relative to the average imprecision: 5.30
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.16 (Individuals) and 0.31 (Groups)
Literacy index W = 0.16 (Individuals) and 0.37 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-5.96	<0.001	-7.16	-4.76
female	-6.97	<0.001	-7.81	-6.13
asian	-6.60	<0.001	-8.26	-4.94
black	-7.36	<0.001	-8.27	-6.45
christian	-7.22	<0.001	-8.08	-6.36
gpaHI	-6.94	<0.001	-7.95	-5.94
junior	-7.34	<0.001	-9.07	-5.62
senior	-5.86	<0.001	-7.38	-4.34
		Additional Imprecision of Beliefs		
groups	-0.18	0.24	-0.52	0.17
female	-0.12	0.29	-0.41	0.18
asian	-0.11	0.37	-0.62	0.40
black	0.12	0.30	-0.20	0.45
christian	0.04	0.39	-0.26	0.34
gpaHI	-0.04	0.39	-0.38	0.31
junior	-0.22	0.34	-0.99	0.56
senior	-0.06	0.38	-0.47	0.35

Figure 4.4: Elicited Beliefs about The Real Interest Rate Question, Comparing Group and Individual Responses, New Labels

Group average: \$100.0 (N=61) Individual average: \$101.0 (N=106)
p-value on test of difference in averages = 0.004
True response was \$98.98

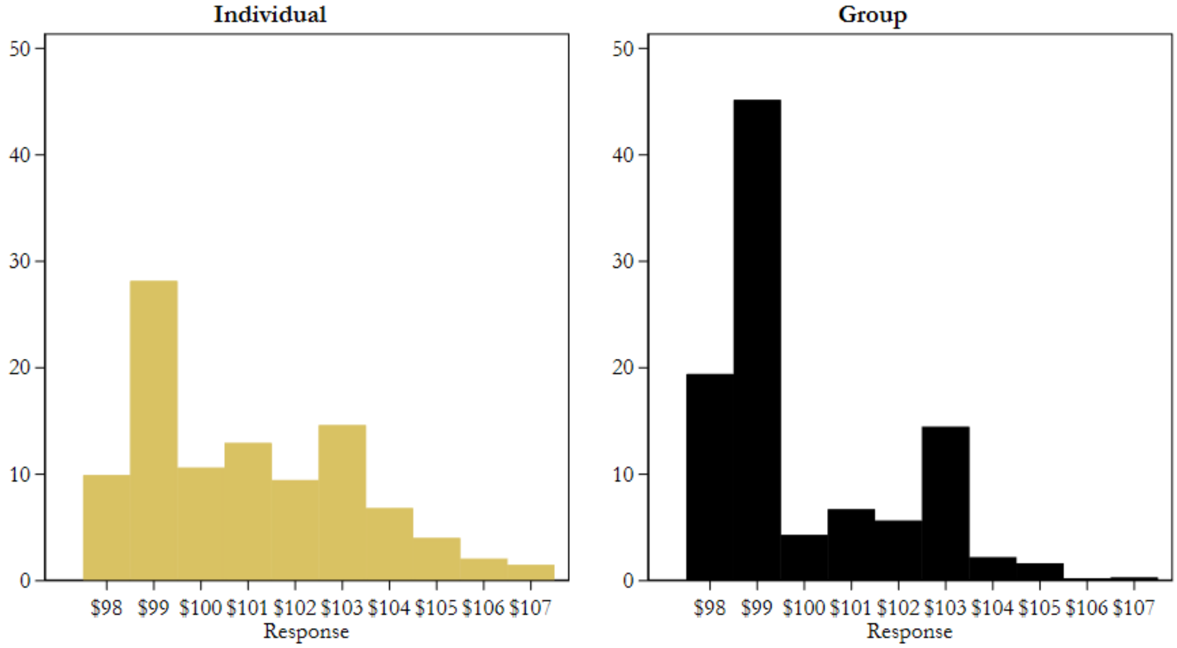


Table 4.7: Effective Literacy of Exogenous Groups for the Question about the Real Interest Rate, New Labels

Correct answer: $98.98 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is < 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-1.11	<0.001	-1.88	-0.35
female	0.65	0.04	0.02	1.28
asian	0.39	0.51	-0.77	1.55
black	-0.47	0.35	-1.46	0.52
christian	0.70	0.05	0.00	1.40
gpaHI	-0.60	0.04	-1.17	-0.04
junior	0.19	0.70	-0.76	1.13
senior	-0.10	0.77	-0.79	0.58
constant	100.66	<0.001	99.46	101.86
		LnSigma		
groups	-0.14	0.32	-0.42	0.14
female	-0.02	0.85	-0.19	0.16
asian	-0.06	0.70	-0.36	0.24
black	-0.11	0.40	-0.35	0.14
christian	0.06	0.61	-0.16	0.27
gpaHI	-0.10	0.24	-0.27	0.07
junior	-0.04	0.81	-0.34	0.26
senior	-0.11	0.32	-0.32	0.11
constant	0.91	<0.001	0.60	1.21

Table 4.8: Literacy Bias and Imprecision, by Demographics, for the Question about the Real Interest Rate, New Labels

Bias is relative to the correct answer: $98.98
Additional imprecision is relative to the average imprecision: 2.19
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index $L = 0.28$ (Individuals) and 0.45 (Groups)
Literacy index $W = 0.29$ (Individuals) and 0.38 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	1.00	<0.001	0.55	1.46
female	1.80	<0.001	1.44	2.15
asian	1.78	<0.001	1.12	2.45
black	1.63	<0.001	1.28	1.99
christian	1.81	<0.001	1.46	2.16
gpaHI	1.44	<0.001	1.06	1.82
junior	1.52	<0.001	0.92	2.13
senior	1.56	<0.001	0.89	2.22
		Additional Imprecision of Beliefs		
groups	-0.26	0.07	-0.53	0.01
female	-0.00	0.40	-0.16	0.16
asian	-0.13	0.31	-0.51	0.24
black	0.02	0.39	-0.16	0.19
christian	0.03	0.38	-0.14	0.19
gpaHI	-0.14	0.18	-0.37	0.08
junior	-0.28	0.06	-0.57	0.01
senior	-0.02	0.40	-0.41	0.36

Figure 4.5: Elicited Beliefs about The Savings Horizon Question, Comparing Group and Individual Responses, New Labels

Group average: 4.3 months (N=61) Individual average: 5.0 months (N=106)
p-value on test of difference in averages = 0.219
True response was 4 months (4m)

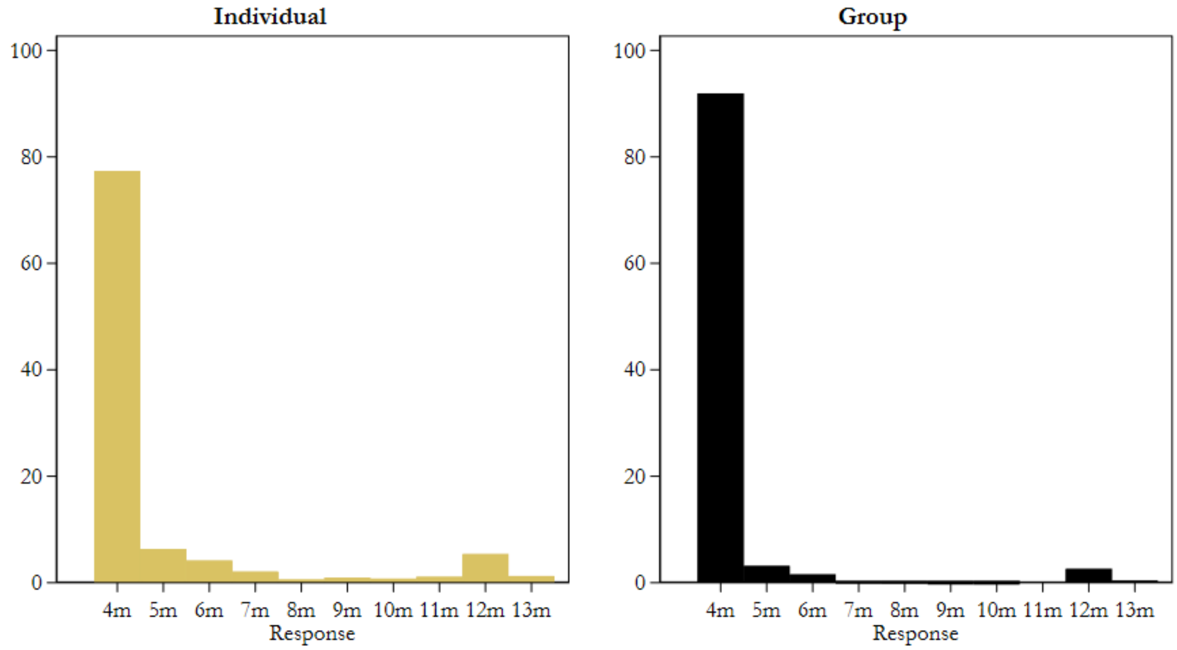


Table 4.9: Effective Literacy of Exogenous Groups for the Question about the Savings Horizon, New Labels

Correct answer: 4 months (4m)   N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.462 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-0.23	0.22	-0.60	0.14
female	0.17	0.10	-0.03	0.37
asian	0.04	0.76	-0.23	0.31
black	0.56	0.02	0.08	1.04
christian	0.17	0.26	-0.13	0.47
gpaHI	0.11	0.46	-0.18	0.40
junior	-0.03	0.88	-0.45	0.39
senior	-0.09	0.58	-0.41	0.23
constant	4.10	<0.001	3.76	4.44
		LnSigma		
groups	-0.36	0.33	-1.08	0.36
female	0.02	0.94	-0.41	0.45
asian	0.68	0.07	-0.05	1.40
black	1.43	<0.001	1.00	1.86
christian	0.27	0.37	-0.32	0.86
gpaHI	0.12	0.64	-0.38	0.61
junior	0.38	0.29	-0.33	1.08
senior	-0.06	0.81	-0.56	0.44
constant	-0.83	0.04	-1.63	-0.03

Table 4.10: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Horizon, New Labels

Bias is relative to the correct answer: 4 months (4m)
Additional imprecision is relative to the average imprecision: 2.00
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.77 (Individuals) and 0.92 (Groups)
Literacy index W = 0.74 (Individuals) and 0.93 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
Bias from the True Response				
groups	0.33	0.06	-0.01	0.66
female	0.69	<0.001	0.36	1.03
asian	0.42	0.13	-0.14	0.98
black	0.92	<0.001	0.52	1.31
christian	0.87	<0.001	0.50	1.24
gpaHI	0.70	<0.001	0.29	1.11
junior	0.89	0.04	0.06	1.72
senior	0.72	0.02	0.14	1.31
Additional Imprecision of Beliefs				
groups	-0.60	0.16	-1.47	0.27
female	-0.07	0.39	-0.63	0.49
asian	-0.41	0.33	-1.75	0.93
black	0.26	0.25	-0.27	0.79
christian	0.16	0.33	-0.37	0.69
gpaHI	-0.02	0.40	-0.69	0.66
junior	0.47	0.28	-0.60	1.55
senior	0.02	0.40	-0.87	0.91

Figure 4.6: Elicited Beliefs about The Stolen Credit Card Question, Comparing Group and Individual Responses, New Labels

Group average: \$132.8 (N=61)   Individual average: \$171.0 (N=106)
p-value on test of difference in averages = 0.102
True response was \$50

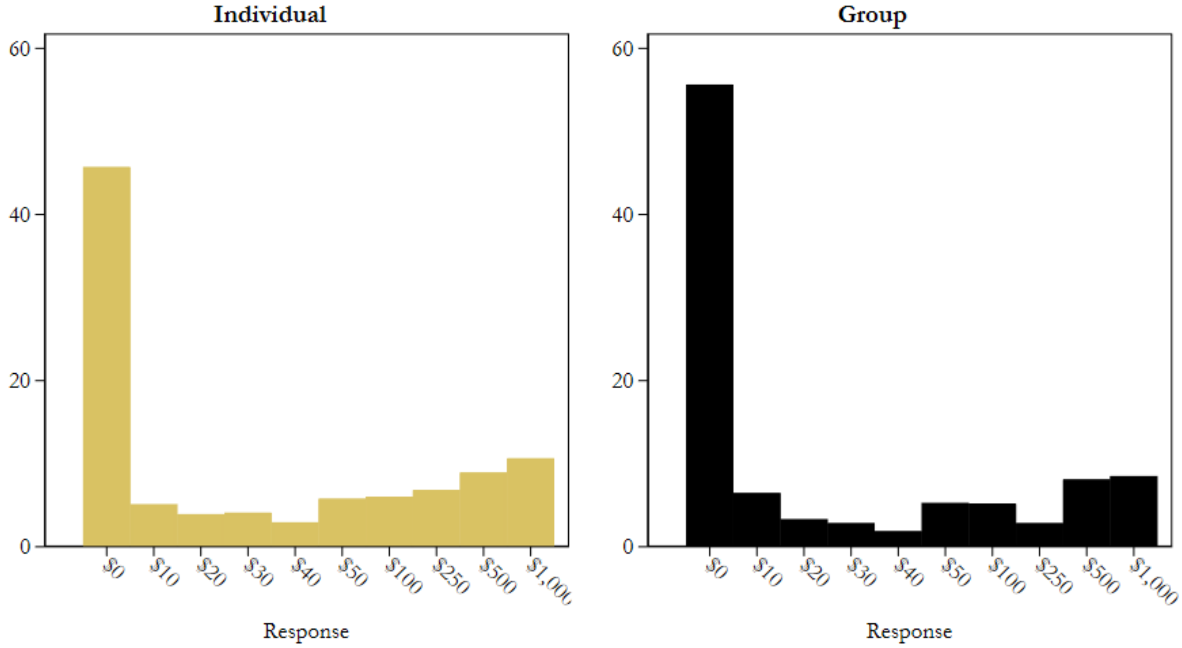


Table 4.11: Effective Literacy of Exogenous Groups for the Question about the Stolen Credit Card, New Labels

Correct answer: \$50 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.228 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-50.38	0.10	-110.85	10.08
female	37.65	0.22	-23.04	98.35
asian	3.50	0.94	-94.96	101.96
black	52.67	0.17	-22.12	127.47
christian	76.08	0.02	9.90	142.26
gpaHI	50.44	0.13	-14.02	114.91
junior	-37.05	0.33	-112.29	38.19
senior	-68.67	0.10	-151.09	13.75
constant	62.64	0.15	-22.61	147.89
		LnSigma		
groups	-0.29	0.09	-0.63	0.05
female	0.25	0.17	-0.11	0.61
asian	0.03	0.95	-0.73	0.78
black	0.24	0.36	-0.28	0.77
christian	0.40	0.06	-0.02	0.83
gpaHI	0.25	0.10	-0.04	0.54
junior	-0.05	0.79	-0.43	0.32
senior	-0.27	0.25	-0.72	0.19
constant	4.97	<0.001	4.26	5.67

Table 4.12: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Credit Card, New Labels

Bias is relative to the correct answer: $50
Additional imprecision is relative to the average imprecision: 274.87
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index *L* = 0.06 (Individuals) and 0.05 (Groups)
Literacy index *W* = 0.06 (Individuals) and 0.07 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	82.84	0.01	21.99	143.69
female	118.03	<0.001	66.80	169.25
asian	69.63	0.08	-5.59	144.85
black	127.33	<0.001	76.02	178.64
christian	138.05	<0.001	87.41	188.69
gpaHI	123.67	<0.001	60.86	186.48
junior	75.22	0.04	5.68	144.76
senior	63.07	0.08	-5.43	131.57
		Additional Imprecision of Beliefs		
groups	-17.62	0.35	-88.59	53.35
female	12.85	0.35	-38.43	64.13
asian	-43.83	0.25	-132.55	44.89
black	17.27	0.31	-31.34	65.88
christian	24.13	0.23	-21.52	69.78
gpaHI	20.94	0.32	-40.22	82.09
junior	-25.73	0.32	-104.25	52.80
senior	-39.43	0.28	-130.31	51.46

Figure 4.7: Elicited Beliefs about The Stolen Debit Card Question, Comparing Group and Individual Responses, New Labels

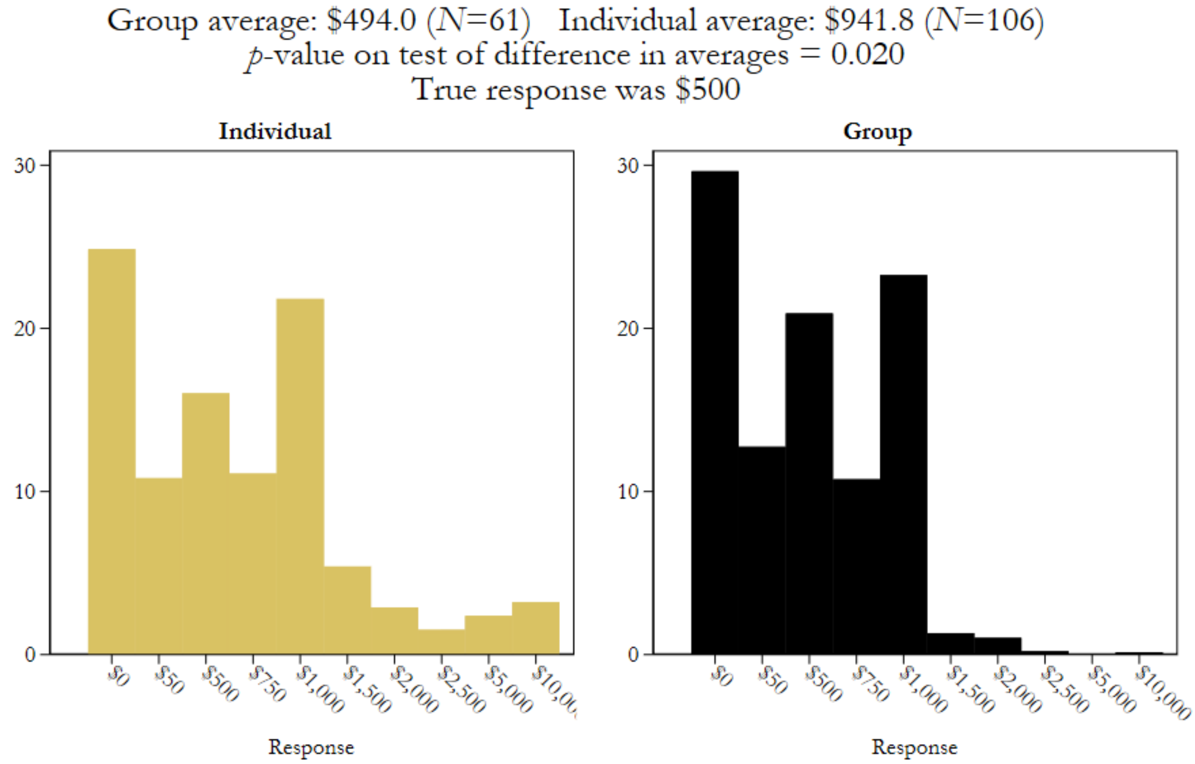


Table 4.13: Effective Literacy of Exogenous Groups for the Question about the Stolen Debit Card, New Labels

Correct answer: $500 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is < 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	-224.52	0.02	-413.97	-35.08
female	23.49	0.82	-182.23	229.21
asian	84.75	0.65	-284.26	453.76
black	126.78	0.36	-144.41	397.96
christian	25.09	0.77	-146.60	196.78
gpaHI	83.88	0.34	-88.81	256.58
junior	-223.93	0.02	-412.10	-35.76
senior	-26.06	0.88	-369.35	317.24
constant	648.15	<0.001	316.97	979.32
		LnSigma		
groups	-0.82	<0.001	-1.12	-0.53
female	-0.04	0.84	-0.44	0.36
asian	-0.54	0.12	-1.22	0.14
black	-0.30	0.25	-0.82	0.21
christian	0.25	0.23	-0.15	0.64
gpaHI	0.23	0.14	-0.08	0.55
junior	-0.11	0.57	-0.49	0.27
senior	0.60	0.01	0.12	1.08
constant	7.00	<0.001	6.35	7.64

Table 4.14: Literacy Bias and Imprecision, by Demographics, for the Question about the Stolen Debit Card, New Labels

Bias is relative to the correct answer: $500
Additional imprecision is relative to the average imprecision: 1241.33
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.16 (Individuals) and 0.21 (Groups)
Literacy index W = 0.12 (Individuals) and 0.15 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-5.95	0.40	-100.99	89.09
female	252.70	<0.001	71.93	433.46
asian	20.21	0.39	-154.58	195.00
black	321.56	<0.001	133.38	509.75
christian	371.02	<0.001	134.80	607.25
gpaHI	349.18	0.03	51.76	646.60
junior	68.56	0.31	-114.94	252.06
senior	586.43	0.11	-123.53	1296.38
		Additional Imprecision of Beliefs		
groups	-715.55	<0.001	-867.57	-563.54
female	-71.33	0.38	-486.28	343.62
asian	-660.74	<0.001	-788.46	-533.02
black	-22.14	0.40	-432.52	388.24
christian	158.59	0.33	-340.38	657.56
gpaHI	170.35	0.35	-471.71	812.40
junior	-463.01	0.04	-879.02	-46.99
senior	880.90	0.12	-222.43	1984.23

Figure 4.8: Elicited Beliefs about The Nominal Interest Question, Comparing Group and Individual Responses, New Labels

Group average: 4.8% (N=61) Individual average: 4.5% (N=106)
p-value on test of difference in averages < 0.001
True response was 5%

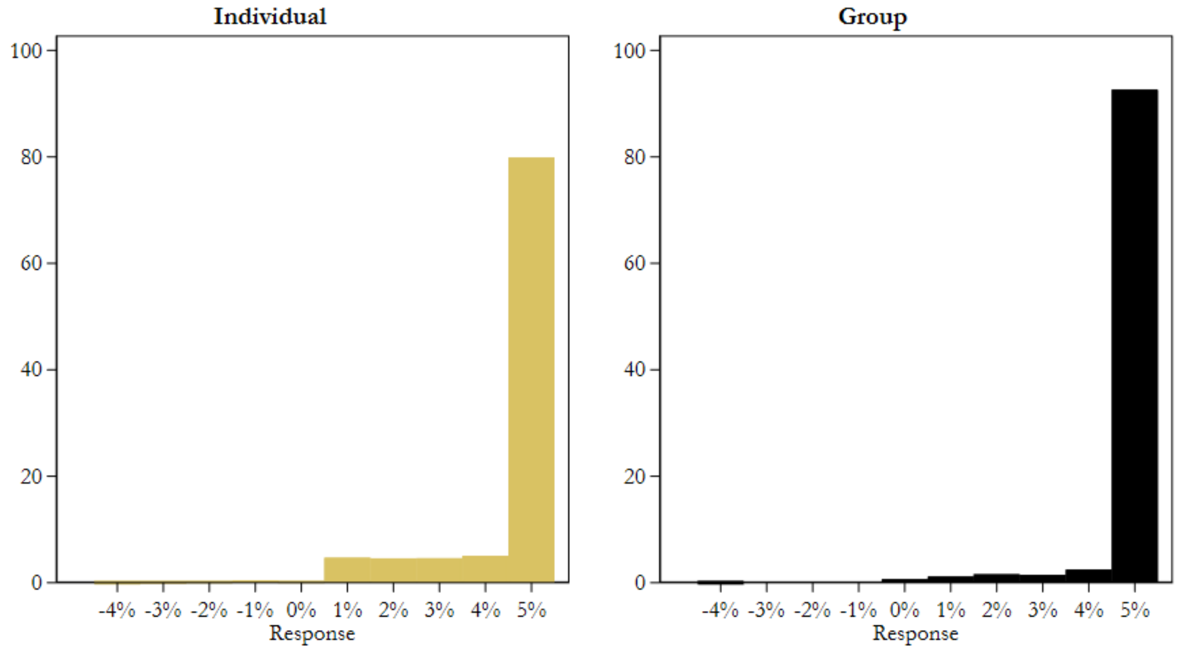


Table 4.15: Effective Literacy of Exogenous Groups for the Question about Nominal Interest, New Labels

Correct answer: 5% N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	0.31	<0.001	0.14	0.49
female	-0.12	0.17	-0.28	0.05
asian	-0.23	0.07	-0.48	0.02
black	-0.07	0.49	-0.25	0.12
christian	-0.08	0.39	-0.25	0.10
gpaHI	0.15	0.11	-0.03	0.33
junior	0.32	<0.001	0.10	0.54
senior	0.28	0.02	0.04	0.51
constant	4.50	<0.001	4.13	4.88
		LnSigma		
groups	-0.64	<0.001	-1.02	-0.26
female	0.07	0.72	-0.31	0.44
asian	0.45	0.12	-0.11	1.02
black	0.22	0.36	-0.25	0.68
christian	0.13	0.58	-0.34	0.61
gpaHI	-0.46	0.01	-0.81	-0.10
junior	-0.80	<0.001	-1.19	-0.42
senior	-0.45	0.06	-0.91	0.02
constant	0.24	0.50	-0.45	0.93

Table 4.16: Literacy Bias and Imprecision, by Demographics, for the Question about Nominal Interest, New Labels

Bias is relative to the correct answer: 5%
Additional imprecision is relative to the average imprecision: 1.09
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.80 (Individuals) and 0.93 (Groups)
Literacy index W = 0.76 (Individuals) and 0.92 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-0.19	0.03	-0.35	-0.03
female	-0.47	<0.001	-0.65	-0.30
asian	-0.36	0.01	-0.63	-0.09
black	-0.47	<0.001	-0.65	-0.28
christian	-0.44	<0.001	-0.60	-0.28
gpaHI	-0.33	<0.001	-0.50	-0.17
junior	-0.14	<0.001	-0.23	-0.05
senior	-0.25	0.01	-0.45	-0.06
		Additional Imprecision of Beliefs		
groups	-0.36	0.07	-0.73	0.02
female	0.02	0.40	-0.22	0.26
asian	-0.14	0.29	-0.50	0.21
black	0.07	0.34	-0.19	0.34
christian	0.01	0.40	-0.20	0.22
gpaHI	-0.19	0.13	-0.44	0.06
junior	-0.52	<0.001	-0.75	-0.30
senior	-0.13	0.34	-0.59	0.33

Figure 4.9: Elicited Beliefs about The Interest Rate Question, Comparing Group and Individual Responses, New Labels

Group average: \$101.9 (N=61) Individual average: \$101.7 (N=106)
p-value on test of difference in averages = 0.079
True response was \$102.00

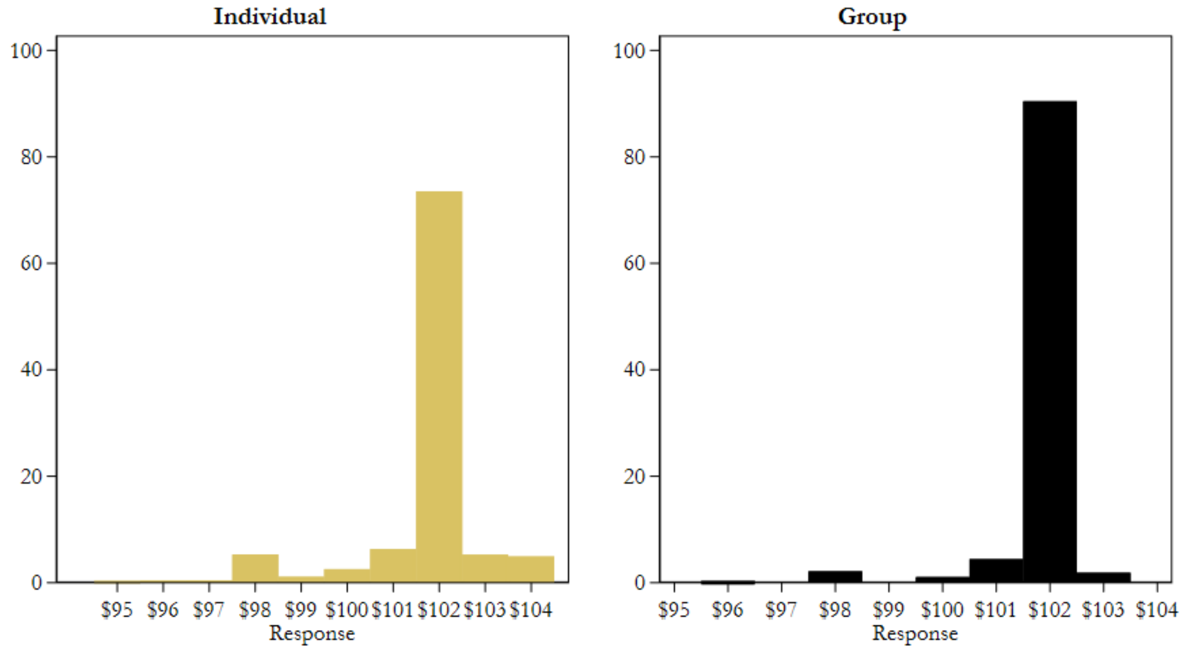


Table 4.17: Effective Literacy of Exogenous Groups for the Question about the Interest Rate, New Labels

Correct answer: $102.00 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is < 0.001 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	0.13	0.08	-0.01	0.28
female	-0.10	0.10	-0.22	0.02
asian	-0.03	0.80	-0.29	0.22
black	0.05	0.71	-0.20	0.29
christian	-0.12	0.08	-0.24	0.01
gpaHI	0.01	0.84	-0.13	0.16
junior	0.10	0.24	-0.06	0.26
senior	0.20	0.01	0.04	0.35
constant	101.78	<0.001	101.50	102.07
		LnSigma		
groups	-0.73	<0.001	-1.11	-0.35
female	0.46	0.01	0.11	0.82
asian	0.06	0.87	-0.62	0.73
black	-0.28	0.33	-0.84	0.28
christian	0.46	0.02	0.06	0.85
gpaHI	-0.07	0.73	-0.44	0.30
junior	-0.40	0.10	-0.87	0.08
senior	-0.47	0.02	-0.88	-0.06
constant	-0.07	0.84	-0.76	0.62

Table 4.18: Literacy Bias and Imprecision, by Demographics, for the Question about the Interest Rate, New Labels

Bias is relative to the correct answer: $102.00
Additional imprecision is relative to the average imprecision: 1.06
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index *L* = 0.74 (Individuals) and 0.90 (Groups)
Literacy index *W* = 0.68 (Individuals) and 0.86 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-0.14	<0.001	-0.24	-0.04
female	-0.22	<0.001	-0.36	-0.07
asian	-0.31	0.06	-0.62	0.00
black	-0.20	<0.001	-0.35	-0.06
christian	-0.26	<0.001	-0.42	-0.10
gpaHI	-0.18	0.03	-0.34	-0.02
junior	-0.23	0.06	-0.46	0.01
senior	-0.07	0.27	-0.21	0.08
		Additional Imprecision of Beliefs		
groups	-0.42	<0.001	-0.67	-0.17
female	0.06	0.35	-0.17	0.28
asian	0.07	0.38	-0.37	0.50
black	0.00	0.40	-0.23	0.24
christian	0.10	0.27	-0.13	0.33
gpaHI	-0.04	0.38	-0.30	0.22
junior	-0.19	0.29	-0.64	0.27
senior	-0.24	0.18	-0.61	0.13

Figure 4.10: Elicited Beliefs about
The Remaining Life for Men Question,
Comparing Group and Individual Responses, New Labels

Group average: 50.0 years (N=61) Individual average: 47.6 years (N=106)
p-value on test of difference in averages = 0.026
True response was 57.1

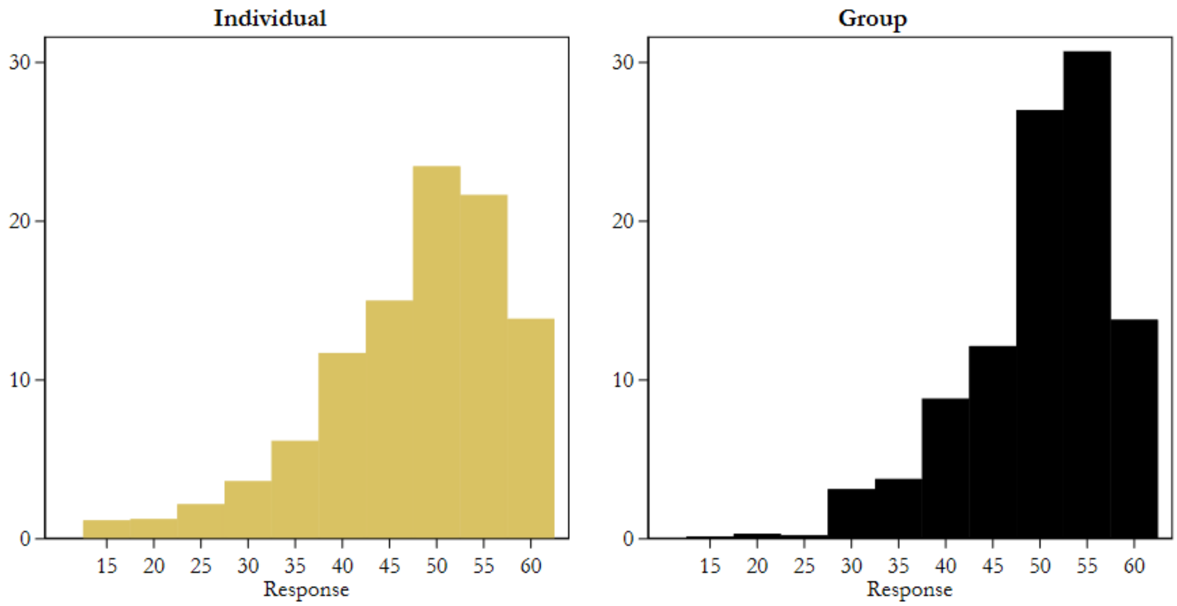


Table 4.19: Effective Literacy of Exogenous Groups for the Question about the Remaining Life for Men, New Labels

Correct answer: 57.1 N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.030 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	2.40	0.03	0.29	4.51
female	-3.95	<0.001	-6.02	-1.88
asian	-0.76	0.67	-4.32	2.79
black	-1.04	0.48	-3.89	1.82
christian	-0.05	0.97	-2.63	2.54
gpaHI	1.30	0.23	-0.84	3.43
junior	-0.70	0.63	-3.52	2.13
senior	-1.42	0.29	-4.04	1.20
constant	51.02	<0.001	47.55	54.49
		LnSigma		
groups	-0.25	0.01	-0.46	-0.05
female	0.33	<0.001	0.13	0.52
asian	0.07	0.71	-0.31	0.45
black	0.04	0.79	-0.28	0.36
christian	0.00	0.99	-0.24	0.25
gpaHI	-0.27	<0.001	-0.46	-0.09
junior	0.01	0.93	-0.25	0.28
senior	0.14	0.21	-0.08	0.36
constant	2.05	<0.001	1.65	2.45

Table 4.20: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Men, New Labels

Bias is relative to the correct answer: 57.1
Additional imprecision is relative to the average imprecision: 9.08
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.22 (Individuals) and 0.31 (Groups)
Literacy index W = 0.18 (Individuals) and 0.22 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-7.06	<0.001	-8.74	-5.37
female	-9.77	<0.001	-11.33	-8.21
asian	-7.87	<0.001	-10.47	-5.26
black	-9.31	<0.001	-10.88	-7.73
christian	-8.84	<0.001	-10.27	-7.42
gpaHI	-8.02	<0.001	-9.45	-6.59
junior	-7.64	<0.001	-9.79	-5.48
senior	-10.02	<0.001	-13.12	-6.91
		Additional Imprecision of Beliefs		
groups	-1.46	0.05	-2.83	-0.08
female	0.41	0.31	-0.72	1.54
asian	-0.48	0.36	-2.57	1.60
black	0.38	0.33	-0.80	1.56
christian	-0.03	0.40	-0.95	0.89
gpaHI	-1.10	0.07	-2.24	0.04
junior	-1.34	0.13	-3.06	0.38
senior	1.25	0.25	-1.30	3.80

Figure 4.11: Elicited Beliefs about
The Remaining Life for Women Question,
Comparing Group and Individual Responses, New Labels

Group average: 53.2 years (N=61) Individual average: 50.3 years (N=106)
p-value on test of difference in averages = 0.001
True response was 61.7

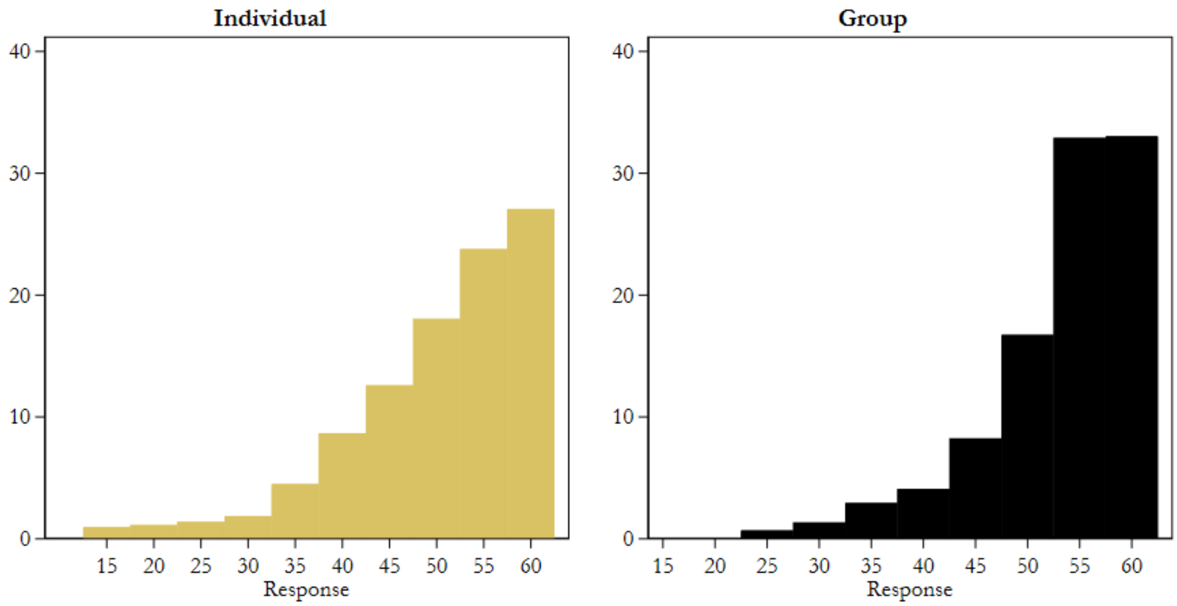


Table 4.21: Effective Literacy of Exogenous Groups for the Question about the Remaining Life for Women, New Labels

Correct answer: 61.7   N=167 GSU undergraduates
p-value for test of hypothesis that group is zero is 0.002 (Overall)

Parameter	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Average		
groups	2.95	<0.001	1.14	4.75
female	-4.29	<0.001	-6.16	-2.43
asian	-0.45	0.81	-4.05	3.15
black	0.55	0.69	-2.17	3.27
christian	0.39	0.74	-1.89	2.67
gpaHI	1.45	0.14	-0.45	3.35
junior	0.47	0.67	-1.69	2.63
senior	-0.38	0.78	-3.00	2.25
constant	52.01	<0.001	48.53	55.49
		LnSigma		
groups	-0.32	<0.001	-0.52	-0.12
female	0.40	<0.001	0.19	0.60
asian	0.01	0.97	-0.42	0.44
black	-0.11	0.50	-0.44	0.22
christian	-0.08	0.55	-0.33	0.18
gpaHI	-0.19	0.06	-0.40	0.01
junior	-0.16	0.19	-0.41	0.08
senior	0.13	0.29	-0.11	0.36
constant	2.15	<0.001	1.73	2.57

Table 4.22: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Women, New Labels

Bias is relative to the correct answer: 61.7
Additional imprecision is relative to the average imprecision: 8.86
Sample sizes: 106 Individuals and 61 Groups (hence 122 individuals)
Literacy index L = 0.27 (Individuals) and 0.33 (Groups)
Literacy index W = 0.22 (Individuals) and 0.27 (Groups)

Demographic	Estimate	*p-value*	95% CI Lower	95% CI Upper
		Bias from the True Response		
groups	-8.45	<0.001	-10.10	-6.81
female	-11.51	<0.001	-13.05	-9.98
asian	-11.26	<0.001	-14.81	-7.72
black	-10.34	<0.001	-11.67	-9.00
christian	-10.18	<0.001	-11.48	-8.88
gpaHI	-9.67	<0.001	-11.27	-8.08
junior	-8.81	<0.001	-10.48	-7.14
senior	-11.00	<0.001	-14.03	-7.97
		Additional Imprecision of Beliefs		
groups	-1.70	0.05	-3.34	-0.06
female	0.54	0.29	-0.75	1.82
asian	1.33	0.30	-2.17	4.82
black	-0.23	0.36	-1.28	0.82
christian	-0.35	0.32	-1.37	0.68
gpaHI	-0.42	0.36	-2.23	1.40
junior	-2.11	<0.001	-3.60	-0.61
senior	1.55	0.22	-1.19	4.28

Table 4.23: Pooled Measures of *L* and *W* Indices, Individual "Control" and Group Scaffold

			Individual Literacy Measures		Group Literacy Measures	
Question	Type	Correct Answer	*L*	*W*	*L*	*W*
fin1 - Savings Account 2%	Numeracy	$110.41	0.53	0.50	0.67	0.62
fin2 - Social Security Start Age	Procedural	62	0.26	0.21	0.21	0.18
fin5 - Medicare Eligibility	Procedural	65	0.16	0.16	0.31	0.37
fin7 - Real Interest Rate	Numeracy	$98.98	0.28	0.29	0.45	0.38
fin9 - Savings Horizon	Numeracy	4 months	0.77	0.74	0.92	0.93
fin10 - Stolen Credit Card	Procedural	$50	0.06	0.06	0.05	0.07
fin11 - Stolen Debit Card	Procedural	$500	0.16	0.12	0.21	0.15
fin13 - Nominal Interest	Numeracy	5%	0.80	0.76	0.93	0.92
fin14 - Interest Rate	Numeracy	$102	0.74	0.68	0.90	0.86
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22	0.18	0.31	0.22
fin16 - Remaining Life for Women	Longevity Risk	61.7 years	0.27	0.22	0.33	0.27

*Pooled literacy measures of *L* and *W* are initally reported for each question in their respective section. Chapter 2 for indvidual measures and Chapter 4 for Group measures.

Table 4.24: Pooled Measures of *L* Index, Comparing the Group Scaffold to Individual Measures

			Individual Literacy Measures		Group Literacy Measures	
Question	Type	Correct Answer	*L*		*L*	+/-
fin1 - Savings Account 2%	Numeracy	$110.41	0.53		0.67	0.14
fin2 - Social Security Start Age	Procedural	62	0.26		0.21	-0.05
fin5 - Medicare Eligibility	Procedural	65	0.16		0.31	0.15
fin7 - Real Interest Rate	Numeracy	$98.98	0.28		0.45	0.17
fin9 - Savings Horizon	Numeracy	4 months	0.77		0.92	0.15
fin10 - Stolen Credit Card	Procedural	$50	0.06		0.05	-0.01
fin11 - Stolen Debit Card	Procedural	$500	0.16		0.21	0.05
fin13 - Nominal Interest	Numeracy	5%	0.80		0.93	0.13
fin14 - Interest Rate	Numeracy	$102	0.74		0.90	0.16
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22		0.31	0.09
fin16 - Remaining Life for Women	Longevity Risk	61.7 years	0.27		0.33	0.06

*Pooled literacy measures of L are initally reported for each question in their respective section. Chapter 2 for indvidual measures and Chapter 4 for Group measures. The "+/-" column is the difference of the treatment compared to the Individual "control" subjects.

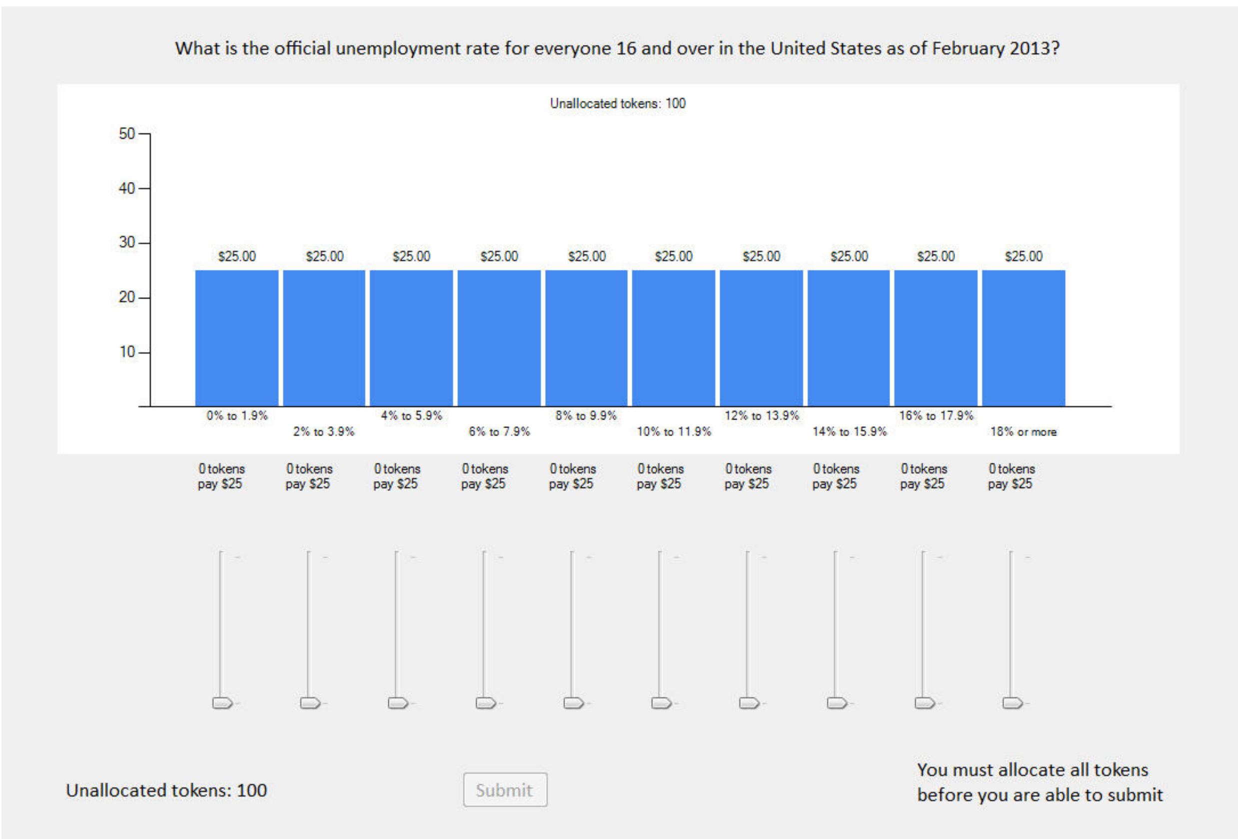
Appendix F: Experimental Instructions

1. Instructions for Belief Elicitation, Lab Sessions with Groups

Group Beliefs

This is a task where your assigned group will be paid according to how accurate your beliefs are about certain things. You will be presented with 11 questions and asked to place some bets on your group's beliefs about the answers to each question. You will actually be rewarded for your group's answer to one of these questions, so you should think carefully about answering each question. The question that is chosen for payment will be determined after everyone has made all decisions, and that process is explained below.

Here is an example of what the computer display of a question might look like.

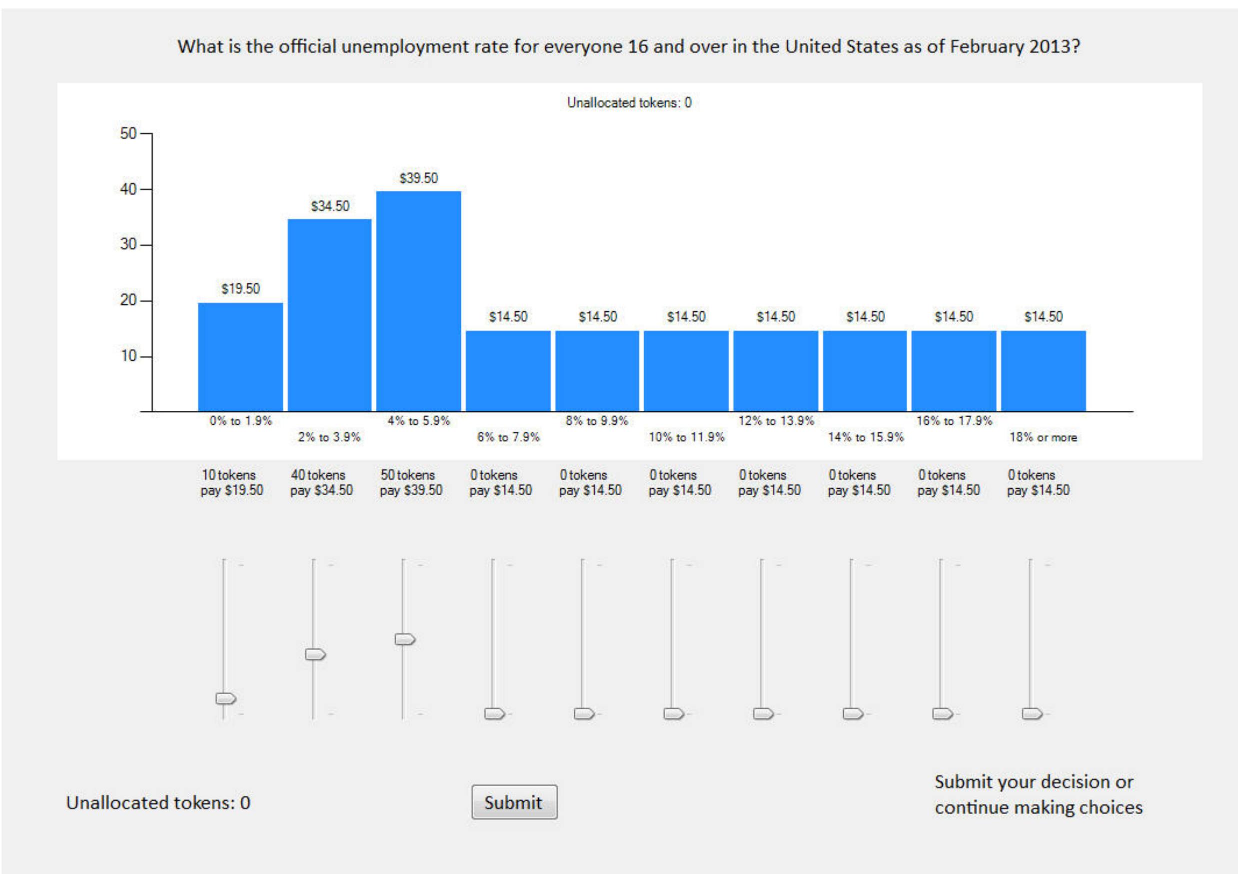


The display on the computer will be larger and easier to read. There are 10 sliders to adjust, shown at the bottom of the screen, and your group has 100 tokens to allocate across the sliders. Each slider allows the group to allocate tokens to reflect the group's belief about the answer to this question.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

While you will make all choices as a group, individually you can earn up to $50 in this task.

Where your group positions each slider depends on your group's beliefs about the correct answer to the question. Please note that the bars above each slider correspond to that particular slider. In the above example, the tokens the group allocates to each bar will naturally reflect your group's beliefs about the official unemployment rate for everyone 16 and over in February 2013. The first bar corresponds to your group's belief that the unemployment rate is between 0% and 1.9%. The second bar corresponds to your group's belief that the unemployment rate is between 2% and 3.9%, and so on. Each bar shows the amount of money you as an individual could earn if the official unemployment rate is in the interval shown under the bar.

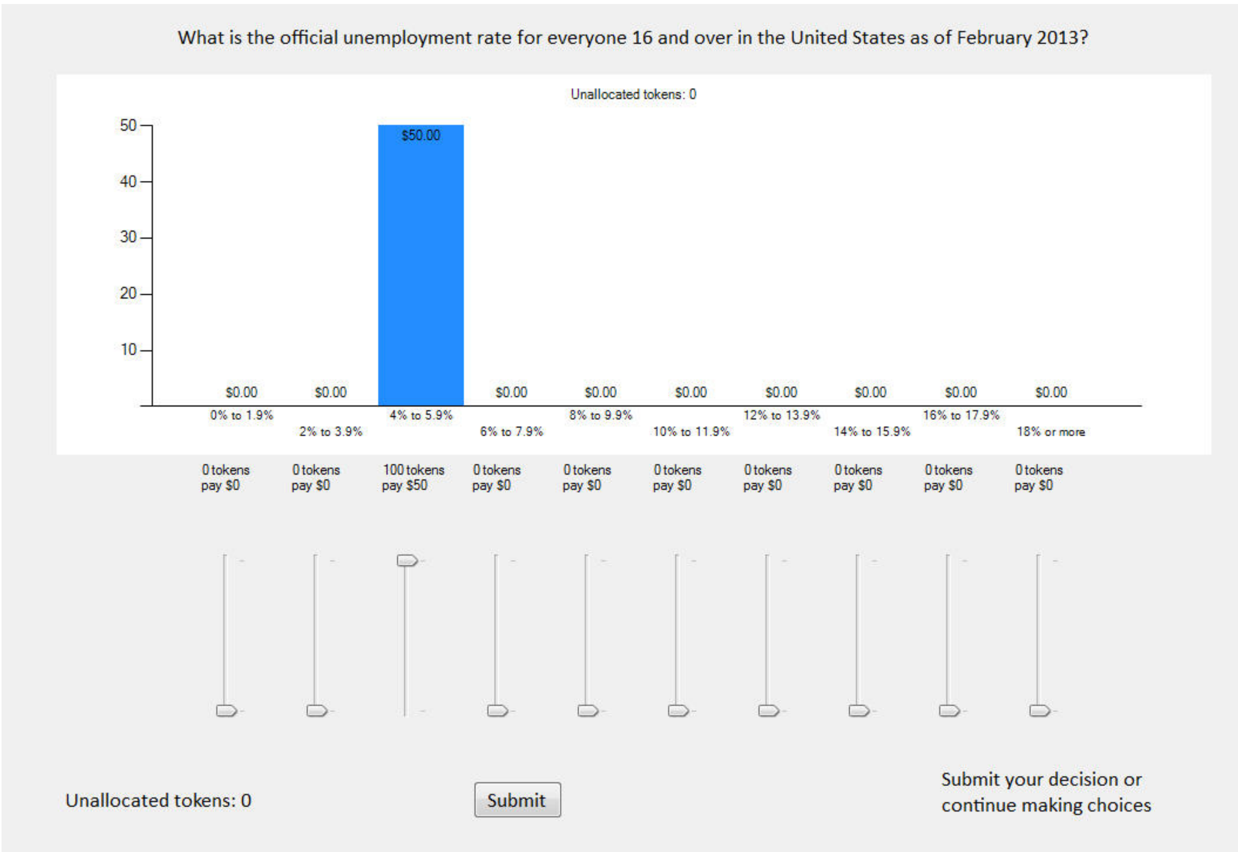
To illustrate how to use these sliders, suppose your group thinks there is a fair chance the unemployment rate is just under 5%. Then your group might allocate the 100 tokens in the following way: 50 tokens to the interval 4% to 5.9%, 40 tokens to the interval 2% to 3.9%, and 10 tokens to the interval 0% to 1.9%. So you can see in the picture below that if indeed the unemployment rate is between 4% and 5.9% you as an individual would earn $39.50. You would earn less than $39.50 for any other outcome. You would earn $34.50 if the unemployment rate is between 2% and 3.9%, $19.50 if it is between 0% and 1.9%, and for any other unemployment rate you would earn $14.50.



The allocation of tokens can be adjusted as much as your group wants to best reflect your group's beliefs about the unemployment rate.

Your earnings depend on your group's reported beliefs and, of course, the true answer. For instance, suppose your group allocated the tokens as shown in the previous figure. The true unemployment rate is actually 7.7%, according to the *Bureau of Labor Statistics*. So if your group had reported the beliefs shown in the previous figure, each individual in the group would have earned $14.50.

Suppose your group had put all your eggs in one basket, and allocated all 100 tokens to the interval corresponding to unemployment rates between 4% and 5.9%. Then you and your group members would have faced the individual earning outcomes shown below.



Note the "good news" and "bad news" here. If the unemployment rate is indeed between 4% and 5.9%, you earn the maximum payoff, shown here as $50. But the true unemployment rate is 7.7%, so you would have earned nothing in this task.

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

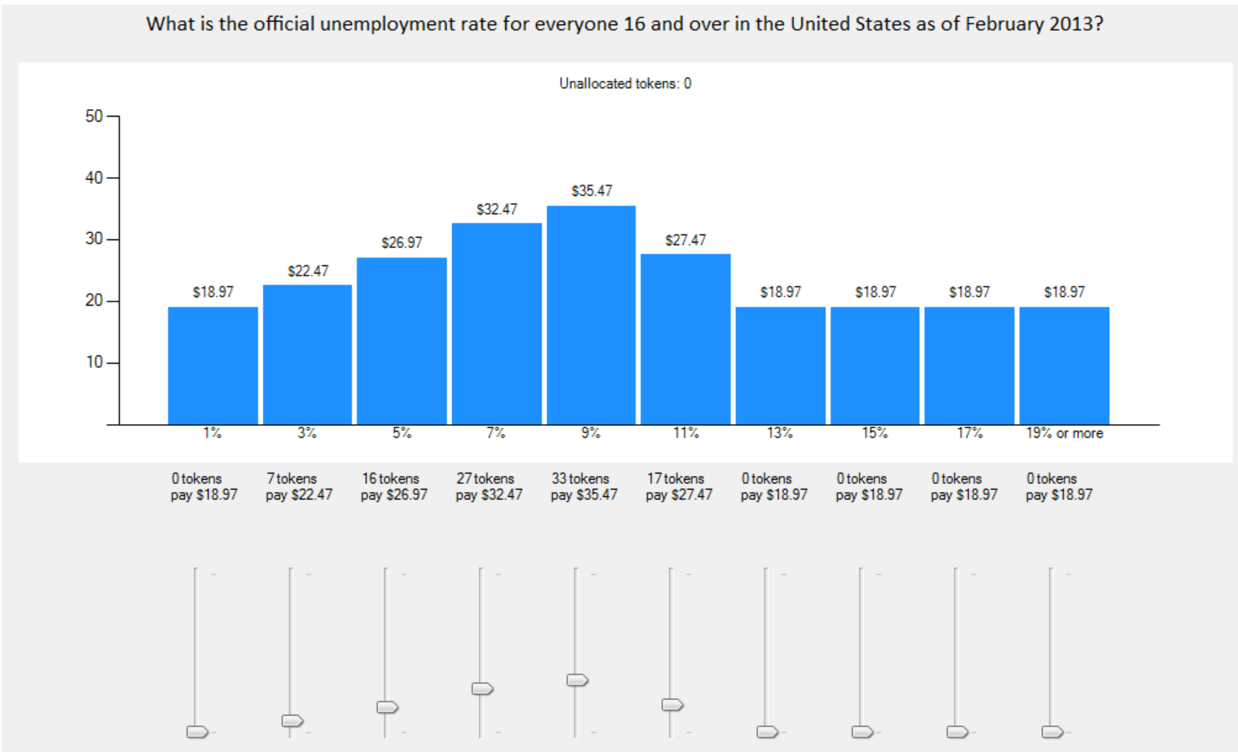
- **Your group's belief about the correct answer to each question is a group judgment, which depends on the information that individuals in the group have about the topic of the question.**

- **Depending on your group's choices and the correct answer you can individually earn up to $50. So if your group has 2 individuals in it, each individual could earn up to $50.**

- **Your group's choices might also depend on the group's willingness to take risks or to gamble.**

The decisions your group makes are a matter of your group's choice. Please work quietly with only the individual(s) that have been assigned to your group, and make your group choices by thinking carefully about the questions that are presented.

For some of the questions we will round the correct answer to the nearest amount shown under each bar. For example, the decision screen for the unemployment question might have shown unemployment rates of 1%, 3%, 5%, 7%, 9%, 11%, 13%, 15%, 17% and 19% or more, as shown below.



In this case, the correct answer of 7.7% would have been rounded to 7% rather than rounded to 9%, and the payment would have been $32.47.

When the group is satisfied with its decisions, you should click on the **Submit** button and confirm the choices. When everyone is finished we will come to each group and roll a 20-sided die until a number between 1 and 11 comes up to determine which question will be played out for each group. The members from each group should select one person from their group to roll the die. The roll is binding for all individuals in the group, thus all group members will receive the same earnings payout. The experimenter will record your individual earnings according to the correct answer and the choices your group 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.

Are there any questions?

Chapter 5 – Extended and Effective Literacy of the Household[35]

5.1 Introduction

Do households, as joint decision makers, exhibit an effective literacy that differs from individuals? We examine this question using a field sample that allows us to draw inferences about the adult population of Denmark. We compare the joint behavior of a household couple to two types of controls. One control is separate behavior of each individual in a couple, and the other is the behavior of individuals that are not members of a household couple. We undertake incentivized field experiments to estimate atemporal risk preferences and subjective beliefs about longevity and major health risks. We recruit subjects from the Danish Registry on the basis of administrative records showing their household status, among many other things.

5.2 Theoretical Framework

As explained in prior chapters, we need to elicit risk preferences and subjective beliefs of subjects. The risk preferences will be used to recover latent beliefs from observed responses to the scoring rules used to elicit beliefs. We use the same method of recovering latent beliefs for individuals *within* households as we did for recovering beliefs for individuals *within* groups in Chapter 4.[36]

Households function as institutions for the members that comprise them, thus the household itself is a scaffold that can be shared by its members. A household's scaffold is quite apart from the social

[35] This chapter is based on data collected by Steffen Andersen, Glenn Harrison, Morten Lau, and Jimmy Martínez-Correa in Denmark. They have generously agreed to allow the data to be analyzed here.

[36] In Chapters 2 and 3 the application to employ the recovery method by Harrison and Ulm (2016) is straightforward as those treatments deal solely with individual participants making their own decisions in both the belief task and the risk task. However, the method is not as direct when recovering latent beliefs belonging to a household making a joint decision and having access to only the individual risk measures for each member of the household. The method that is adopted here assigns the raw token allocation given by the household making a joint decision to each individual in that household, and then employs the recovery of latent beliefs for each household member using their individual measure of risk. This acknowledges that a joint decision was made by household members and that we then recover the latent beliefs of the household based on each member's individual risk aversion. Hence we are recovering the latent beliefs that each subject agreed to when they agreed to a joint allocation of tokens. If the individuals within the household have different risk preferences, which can be the case, then the recovered beliefs from the agreed household allocation of tokens will differ from each other.

interaction that defines the typical household, and the use of language within it; however, it adds structure to both.

We seek to understand the significance of the distinction between having household couples make joint decisions and having them make separate decisions. In the second case, again for the typical household, we are counterfactually removing a key part of the household scaffold by limiting member access. Not all of it, but arguably a key component.

Thinking of households as scaffolds leads one to immediately see the sense in which the "effective literacy" concept, introduced in Section 1.2.1, can take on many forms. One could immediately imagine other scaffolds akin to households, leading to various notions of effective literacy when different scaffolds are present. Committee deliberations, particularly if mediated by some formal mode of conduct such as Roberts Rules of Order, are one important example. Parliamentary debate is another. Trial testimony or jury deliberation is yet another.

5.3 Data

We access two general classes of data. The first is a massive collection of administrative records best described as "the Danish Registry." In fact, these data are a collection of many distinct databases kept for administrative purposes, and not one monolithic database. The second type of data we access consists of surveys and experiments conducted in the Danish population, using samples drawn from the Registry. In this case we get to ask the questions, and design the experimental tasks. Shared by both classes of data sources are the social security number of individuals, a unique identification number assigned to all legal residents of Denmark making it possible to merge experimental and administrative records.

5.3.1 The Danish Registry

We collate several administrative databases in Denmark. Our data cover the universe of all Danes in the period between 1986 and 2012, and contain demographic, economic, educational and health information. Central to each administrative register is the social security number of each individual, the CPR, making it possible for us to merge information between all registers of interest.[37]

We obtain demographic information from the Danish Civil Registration System (*CPR Registeret*). These records include the CPR of the individual, their name, residential address, gender, date of birth, marital history (number of marriages, divorces, and widowhoods). The demographic record also contain the unique household identification number, as well as CPR numbers of each individual in the household. Thus we can identify "family units" and correlate information from one member of the household with information about another member of the household. The sample contains the entire Danish population by each year, and provides a unique identifying number across individuals, households, generations and time.

We obtain income and wealth information from the records at the Danish Tax Authority (*SKAT*). These data contain total and disaggregated income and wealth information by CPR number for the entire Danish population. SKAT receives this information directly from the relevant third-party sources, because employers supply statements of wages paid to their employees, and financial institutions supply information to SKAT on their customers' deposits, interest paid or received, security investments, and dividends. Because taxation in Denmark mainly occurs at the source level, this income information is highly reliable.

We obtain individual employment and unemployment spells from the Income Registry (*Centrale Oplysningsseddelregister*) at the Danish Tax Authority. Employment and unemployment spells are identified from the statement of wages paid to employees that employers are obliged to submit to the Danish tax authorities. We use this to calculate the fraction of days of the year that someone is not employed and do so for each individual.

The status of the individual as unemployed, self-employed or salaried is formally defined by Statistics Denmark as the individual's employment status in week 48 in the year. An unemployed person is defined as

[37] There are many more data registers available in Denmark beyond those that we use.

a person without a job, who is available for the labor market and receiving unemployment benefits. A self-employed person is defined as an individual receiving the majority of their income from their own company. Salaried individuals are defined as obtaining the majority of their income from salaried work. The residual status are individuals who are not available for the formal labor market, either because they are on leave, on social welfare benefits, working illegally in informal labor market, students, home-based spouses or pensioners.

We obtain the level of education from information collected by the Danish Ministry of Education (*Undervisningsministeriet*). This register identifies the highest level of education, both formal and informal, and the resulting professional qualifications. On this basis we calculate the number of years of schooling. We also collate all grades received in the ninth grade, the final year of formal primary public schooling in Denmark, and just before some individual enters High School. These data contain grades in several different courses, and we rank all students in Denmark in each cohort from the worst to the best by a simple constructed simple geometric average. We only obtain these grades from 1994 on.

Finally, we obtain data on medical treatments and hospitalizations from the Danish National Board of Health (*Sundhedsstyelsen*). These data record individual medical treatments and discharges from hospitals. Diagnosis and treatment procedures are classified according to the ICD-10 system, which spans morbidities as well as mortalities. For mortalities, the ICD-10 codes also include detailed "event codes" that explain the manner in which certain deaths occurred. Based on hospitalization nights, we calculate the total number of days in hospital during the year.

5.3.2 Experiments

The experimental design implements choice tasks for individuals that allow us to estimate, for that individual, the theoretical structure needed to measure literacy by eliciting subjective beliefs.

5.3.3 Sampling Procedures

For logistical reasons we restrict the sample to "greater Copenhagen," to minimize travel distances to attend experimental sessions in Copenhagen. We use a family identification number to identify all family units in Denmark. In this context, a *family unit* is understood as a single or a couple[38] living with or without children.[39] In the beginning of 2014 this is a universe in Denmark of N = 2,916,677 family units that we could potentially sample from, consisting of 5,627,235 individuals.

We restrict this sample in several ways:

- We exclude family units with more than 10 members (30 family units removed).
- We exclude children from the sample since we only want to recruit adults making the economic decisions in the family unit (0 family units removed, but 1,389,749 children living at home removed).
- We exclude family units in which the oldest person is a minor (probably people living in dorms or children fleeing from their parents; 16,058 family units removed).
- We exclude family units in which a spouse is less than 18 years of age (we cannot recruit minors into our experiments; 293 family units removed).
- We exclude family units in which the youngest adult is older than 67 years (544,314 family units removed).
- We exclude family units with one adult, and where the adult is married (these family units are probably separated from their spouse, or "between spouses", and done so to ensure the recruitment of 2 adults in a household, so we have a clear separation with the controls of a single individual; 81,118 family units removed).
- From the remaining population, for logistical reasons we exclude households not living in greater Copenhagen, defined as households located with municipality codes higher than 200 (1,533,364 family units removed).

After making these exclusions we are left with 1,070,279 individuals in 741,500 of three types of family units:

1. Family units with one adult that is not married (N = 412,721);
2. Family units with two adults that are not married (N = 189,658); and
3. Family units with two adults that are married, including registered partnerships (N = 467,900).

[38] A couple is defined as two people who live together and form a couple of one of the following four types: married couples; registered partnership (introduced on 1 October 1989), which can include same-sex couples; cohabiting couples definition #1 (*samlevende par* in Danish, two people that are not married or in a registered partnership with each other, but have at least one child in common registered with a social security number); and cohabiting couples definition # 2 (*samboende par*, two adult persons of the opposite sex who have at most 15 years of age difference and that have no children registered with a social security number, and as far as social security number can say, they are not closely related to each other, in the sense of being siblings or parent-child).

[39] Children living at home are counted as part of their parents' families if they live at the same address as at least one parent; are under 25 years old; have never been married; do not have a child or children registered; and are not part of a co-habiting couple.

We define single individuals as family units of type 1, and define households as family units of types 2 and 3. From this sample we will compare the *joint* decisions of a household couple to two types of controls. One control is the *separate* behavior of each individual in a couple, and the other is the behavior of the *single* individuals that are not members of a household couple.

5.3.4 Experimental Tasks

The experimental tasks presented to our subjects are based on designs developed elsewhere, and successfully applied to the Danish population or other field subjects. Each subject faces two tasks, in addition to several other tasks[40] of no relevance here. The two task are as follows:

- An atemporal risk choice tasks consisting of 60 pairs of lotteries in the gain domain, designed to provide evidence of risk aversion as well as the tendency to make decisions consistently with EUT or RDU models. The battery based on designs from Wakker, Erev, and Weber (1994), Loomes and Sugden (1998) and Cox and Sadiraj (2008, p. 33). We explain the logic of these designs in Appendix C. The analysis of risk attitudes is given these choices follows Andersen, Harrison, Lau and Rutström (2008) and Harrison and Rutström (2008).
- A subjective belief elicitation task consists of 9 applications of the Quadratic Scoring Rule (QSR) applied to continuous distributions. We examine general economic and health literacy, but also look at knowledge of prison sentencing rates of participants found guilty of two types of crime. The economic literacy questions include two questions asked by Lusardi and Mitchell (2007, 2008) in the *Health & Retirement Survey* (HRS) of 2004 in the United States to characterize financial literacy, as well as two questions which ask about expected lifetimes of men and women. These span economic and health literacy, since they are critical inputs into retirement planning, or the lack thereof. We ask four specifically health-related mortality questions, about the prevalence of deaths from diseases of the heart,

[40] One task elicited time preferences, following Andersen, Harrison, Lau, and Rutström (2008, 2014). Another task elicited intertemporal risk preferences, also known as correlation aversion, following Andersen, Harrison, Lau, and Rutström (2018).

cancers, smoking, and vehicle crashes due to alcohol. Finally, we ask one question on the perception of prison sentencing rates of participants found guilty of property crime (e.g., burglary and theft) and one question on the perception of prison sentencing rates of participants found guilty from non-property crime (e.g., homicide, assault, violence, rape and indecent exposure). In the Danish criminal system there is a primary charge and crime that one is ever found guilty of, and it is that crime that is referred to here.

We ask 9 subjective belief elicitation questions that were held constant for each subject, since several of the questions related to each other, and this ensures maximal control for possible order effects across treatments. The correct answers to all questions were documented in the event that any subject wanted to verify the source. We reference them here as Q1 through Q9 for convenience. The maximum earnings in each decision task is 1,000 kroner for individuals and 2,000 kroner for households. Each individual and household is given a 10% chance of being paid in this part of the experiment (i.e., for their responses to the subjective belief questions). The subjective belief questions asked of all subjects were as follows:

Ql. Interest Compounding.
"Suppose you had 1,000 kroner in a savings account and the interest rate is 2% per year and you never withdraw money or interest payments. After 5 years, how much would you have in this account in total?"

The correct answer is 1,104 kroner. Responses were elicited between 1,050 kroner and 1,150 kroner in intervals of 10 kroner.

Q2. Real Interest Rate.
"Suppose you had 2,000 in a savings account. The interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, what would be the real value of the money in this account?"

The correct answer is 1,980 kroner. Responses were elicited between 1,950 kroner and 2,050 kroner in intervals of 10 kroner.

Q3. Expected Lifetime for Men.
"Based on 2012-13 statistics, if a man lived to be 20 in Denmark, how many more years would he expect to live? Note that this is not the age he would die at, but how many more years he would expect to live."

The correct answer is 58.4 years. Responses were elicited in decades. The source for the correct answer was www.statistikbanken.dk.

Q4. Expected Lifetime for Women.

"Based on 2012-13 statistics, if a woman lived to be 20 in Denmark, how many more years would she expect to live? Note that this is not he age she would die at, but how many more years she would expect to live."

The correct answer is 62.4 years. Responses were elicited in decades. The source for the correct answer was www.statistikbanken.dk.

Q5. Deaths from Heart Disease.

"What percentage of deceased people died from diseases of the heart in Denmark in 2011?"

The correct answer is 14.4%. Responses were elicited in deciles. The source for the correct answer was www.dst.dk/da/Statistik/emner/doedsfald-og-middellevetid/doedsfald.aspx.

Q6. Deaths from Cancer.

"What percentage of deceased people died from neoplasms (cancers) in Denmark in 2011?"

The correct answer is 29.0%. Responses were elicited in deciles. The source for the correct answer was www.dst.dk/da/Statistik/emner/doedsfald-og-middellevetid/doedsfald.aspx.

Q7. Deaths from Smoking.

"What percentage of deaths from all causes are attributed to smoking in Denmark in 2000?"

The correct answer is 21.5%. Responses were elicited in deciles. The source for the correct answer was http://sundhedsstyrelsen.dk/da/sundhed/tobak/tal-og-undersoegelser/~/media/4BB9830541C84E23968A206B7EDABBC4.ashx?m=.pdf.

Q8. Property Criminal Activity.

"What percentage of people found guilty of property crimes (e.g. burglary and theft) in Denmark in 2012 were sentenced to prison?"

The correct answer is 13.9%. Responses were elicited in deciles. The source for the correct answer was http://www.dst.dk/pukora/epub/upload/17959/sy.pdf.

Q9. Non-Property Criminal Activity.

"What percentage of people found guilty of violent and sexual crimes (e.g. homicide, assault, violence, rape and indecent exposure) in Denmark in 2012 were sentenced to prison?"

The correct answer is 41.8%. Responses were elicited in deciles. The source for the correct answer was http://www.dst.dk/pukora/epub/upload/17959/sy.pdf.

The first four questions are obvious adaptations of questions reviewed in Chapter 2 about financial literacy and longevity risk. The next three questions elicit beliefs about basic health risks and their correlates. One is the general risk of heart disease, another is the general risk of cancers, the two leading causes of death in the Denmark.[41] Then we turn to the role of smoking in deaths. The final two questions consider beliefs about the risk of receiving a prison sentence in Denmark if convicted of property crimes (Q8) or violent and sexual crimes (Q9). These were controls for a separate experiment in which these questions were asked of known criminals.

5.4 Results

We start by examining the effects of our three types of decision maker on preferences and beliefs of a "representative" decision maker, just to focus attention on the general pooled results. We then reconsider those results from estimates at the individual level.

5.4.1 Subjective Beliefs

Tables 5.1 through 5.9, and Figures 5.1 through 5.9, provide an aggregate, unconditional summary of the responses to each question. This allows us to assess overall levels of literacy, before the conditional, nuanced analysis in Tables 5.10 through 5.27, and Figures 5.10 through 5.18, of the effects of socio-demographic characteristics and our core treatments. In effect, the conditional analysis just tells us deviations from the overall average of the unconditional analysis, so the two analyses are complementary.

Table 5.1 and Figure 5.1 show that adult Danes have high literacy with respect to *nominal* interest earnings on savings accounts: the striking, modal response in Figure 5.1 was in the correct response interval. Literacy with respect to *real* interest rates is lower, as shown in Table 5.2 and Figure 5.2, and is generally biased. Although the modal response was again in the correct response interval, there is a much higher level of imprecision than for nominal interest rates and the average is biased. The literacy index W drops from 0.46 for the nominal interest rate question to 0.27 for the real interest rate question. The average responses for the nominal interest rate question, without adjustment for risk preferences, are not statistically

[41] As reported in ("Europe in Figures - Eurostat Yearbook - Statistics Explained") section 3.2 Causes of Death.

significantly different from the true responses, but are statistically significantly biased for the real interest rate question.

Tables 5.3 and 5.4, and Figures 5.3 and 5.4, show generally well-calibrated beliefs about longevity risk for men and women in Denmark. The average responses, without adjustment for risk preferences, are right at the true responses, and not statistically significantly different from them. The true response interval for men is the modal response, and for women it is effectively the modal response: for women, the true response of 58.8 is right at the border of the intervals 50 to 59 and 60 to 69.

Literacy rates for adult Danes plummet, however, when we consider health risks and the perception of prison sentencing rates if found guilty of crime.

For the mortality risk of heart disease, cancer and smoking, Danes are both unbiased and imprecise in their subjective perceptions. Tables 5.5, 5.6 and 5.7, and Figures 5.5, 5.6 and 5.7, show that they overstate each of these health risks on average: by 15.5 percentage points, 4.8 percentage points, and 6.8 percentage points, respectively. In all three cases this difference just falls within a 95% confidence interval. This bias is of great significance if the decision-making agents, individuals or households, violate the Reduction of Compound Lotteries (ROCL) axiom. If they behave consistently with ROCL, then their weighted average mortality risk is unbiased in these three cases. If they violate ROCL, however, then other aspects of the belief distribution than the mean become a factor in decision-making: in most models in which ROCL is violated, the lack of confidence in the mean belief is the driver of "uncertainty aversion" in decision-making.

In the case of the perception of prison sentencing rates if found guilty of a crime, Tables 5.8 and 5.9, and Figures 5.8 and 5.9, show that Danes again overestimate the risk of getting a prison sentence. For property crimes the actual rate of receiving a prison sentence if found guilty is only 13.9%, but the average perceived sentencing rate is 37%, a massive 23.1 percentage points difference. The average perceived sentencing rate for non-property crime, specifically violent and sexual crimes, is not off by as much, but this is accompanied by a strikingly diffuse belief. In general, the average Dane has "no clue" what the risk of imprisonment is if found guilty for major crimes: as noted earlier, in a separate project these beliefs are being compared with those that have received a prison sentence for a crime in Denmark, to see if they are better calibrated.

Turning now to the detailed, conditional results, we want to see what demographic characteristics account for bias and imprecision in beliefs, and of course the effect of making separate or joint household reports. The demographic measures are each generally defined as binary measures as follows: **singles** is 1 for individuals defined as family units of type 1, **separate** and **joint** are households defined as family units of types 2 and 3, with the former making decisions separately and the latter making joint decisions, **female** is a 1 for females, **young** is someone under 30, **middle** is someone between 30 and 40, **old** is someone over 50, **married** is someone that is legally married or in a registered partnership, **highschool** is someone that has completed a high-school diploma, **college** is for someone that has completed at least 3 additional years of education after high school (this can include technical school training), and **owner** is someone that pays property tax on a dwelling (house, apartments, cottages, etc.). The final demographic **inc_rank** is a non-binary variable that varies between 0 and 1 with a median value of 0.5. It is a constructed variable that indicates the income rank position in the population; those with higher values and closer to 1 are individuals that have higher incomes relative to others.

Although not the most interesting example in terms of behavior, because of the relatively high literacy, consider Table 5.10 and Figure 5.10, for the question about nominal interest rates. The body of Table 5.10 has the same structure as Table 5.1, but with additional covariates. From Table 5.1 we see that the overall average belief was 1103.17 kroner: this is different from the coefficient on the constant term in the top panel of Table 5.10, 1095.93 kroner, because the covariates in Table 5.10 do not all have averages of zero. As usual, the constant refers to the "omitted category," which is a single decision-maker who is male, unmarried, between ages 41 and 49, and so on; we refer to this omitted category as "singles" for short, with the understanding that it also involves certain values for other covariates. We observe that separated households report beliefs that are 3.40 kroner lower than singles, but that this difference is not statistically significantly different from zero (*p*-value = 0.18). There are several demographics that are statistically significant at the 5% significance level; young Danes, those with at least 3 years of college, and those with children. In each case, the coefficient of those estimates is positive and is closer to the correct answer.

In the bottom panel of Table 5.10, the interpretation of the coefficients for dispersion is immediate, because the dependent variable is in natural log units. Thus the coefficient estimate β implies a $\beta \times 100$ percent change in the standard deviation of beliefs, σ, for a one unit change in that covariate. So the statistically significant effect of being middle-aged or not (in the case the one unit is the value 1) is interpreted as a 22% reduction in the standard deviation of beliefs compared to the omitted category.

The information in Table 5.10 below the title is generally self-explanatory. The *p*-values for Separate and Joint refer to tests of the hypothesis that the effect on the average and the standard deviation are jointly zero, and complements the corresponding hypothesis tests of the average or the standard deviation in the body of the table. Figure 5.10 shows the distribution of reports for each bin, side-by-side for the Singles, Separate and Joint samples, allowing a visual interpretation of differences in reports.

Table 5.11 provides a more focused evaluation of the bias and imprecision of beliefs that is associated with demographics, and is different from the estimates underlying Table 5.10. In the top panel of Table 5.11 we estimate the "total effect" of the covariate, and then compare the estimated average belief to the true value. The bottom panel of Table 5.11 compares the estimated standard deviation to the pooled standard deviation. Thus we measure bias for the average, and relative imprecision for the standard deviation. These are not marginal effects: for each covariate of interest it is as if we use the average value of all other covariates for that covariate of interest. For females, for instance, the top panel shows that the difference between the average belief of a female is 0.34 less than the correct answer. This is different from Table 5.10, where the coefficient shows the difference compared to the omitted category; in Table 5.11 we implicitly allow females to have the age, marital status, education level, for instance, that females have, not the values of those covariates that the omitted category has. Thus the *p*-values in Table 5.11 tell us what we want to know about bias for females, and the *p*-values in Table 5.10 tell us what we want to know about marginal differences in beliefs for females.

With this background, Table 5.11 shows us that there is a statistically significant bias at the 5% significance level in the literacy of households with respect to nominal interest rates when they make separate decisions, but not when they make joint decisions. Nor is there any bias for single individuals or any other covariates. Here we see that "most Danes get it," which is also shown by the higher levels of the literacy indices *L* and *W*. In the bottom panel of Table 5.11 we see at the 5% significance level that middle-aged and old Danes have statistically significantly different precisions of beliefs compared to the average imprecision of beliefs. Here middle-age Danes exhibit less imprecision in beliefs and old Danes exhibit more imprecision of beliefs compared to the average imprecision. For the income rank estimate we see the estimated average bias away from the correct answer of 1104 kroner is 1.12 which is for a person with an average income ranking, but is not statistically significant at conventional levels. Here the 95% confidence interval is between -1.97 and 4.22 is reported over all income ranking.

When we consider literacy with respect to real interest rates, we find a general bias from the correct answer and in the same direction for all Danes, no matter what household treatment they are in, or their demographic profile. Only college graduates and those in the higher income bracket hold these biased beliefs with greater precision than others.

Turning to literacy with respect to longevity risk, we find the first significant effect of our treatments, shown most clearly in Tables 5.15 and 5.17. In this case singles and separated households have biased beliefs, underestimating the longevity of both men and women. But households making joint decisions, while they have the same bias, do not have a statistically significant bias in general (*p*-value = 0.30 in the remaining life for men and *p*-value = 0.18 in the remaining life for women). The subtlety here is that "households" as we have defined them include married households *and* other households recognized as a household in Denmark even if they are not legally married. The subtlety arises because married decision-makers have a statistically significant bias (p-value = 0.04 for both questions); thus the non-married households making joint decisions are the ones that are clearly unbiased. For the longevity risk for men, the exceptions to this tendency are young and college-educated Danes who do not exhibit a statistically significant bias at the 5% significance level. Only college-educated Danes exhibit a statistically significantly greater precision in their beliefs, which as just noted are unbiased. For the longevity risk for women, we see a similar pattern in terms of demographics, except that it is young and middle-aged Danes, and college-educated that exhibit no statistically significant bias at the 5% significance level. Additionally, Danes in higher income brackets exhibit no statistically significant bias. However, young Danes and those in higher income brackets would not be statistically significant at the 10% significance level.

Literacy with respect to the risks of death from heart diseases is poor, as noted earlier. From Table 5.19 we see that every demographic and household treatment exhibits the same qualitative bias, and it is statistically significant in all cases. Given this bias away from the correct answer for all covariates, the illiteracy is exacerbated by a greater precision of beliefs, which we see for households making joint decisions, married Danes, and middle-age. Young Danes, on the other hand, have a statistically significant *increase* in the *im*precision of their beliefs, which is not a bad thing for their overall literacy given that they are biased. This signals that they "know they don't know the correct answer" thus report more diffused beliefs.

The pattern of illiteracy for the risks of death from cancer is similar to heart disease, but nuanced in some respects. At the 5% significance level, the top panel of Table 5.21 shows that households making a separate decision have a statistically significant bias of 6.83 percentage points away from the correct answer. Singles and households making a joint decision are unbiased from the correct answer, although, singles would be deemed to have biased beliefs if evaluated at the 10% level. Additionally, middle-aged Danes and those with a high income rank do not exhibit significant bias from the correct answer. Only middle-aged

Danes exhibit *no* statistically significant bias (*p*-value = 0.12), although they also have more imprecision of beliefs (*p*-value = 0.04). Returning to households, we see the importance of considering both the bias and precision of beliefs. Households making joint decisions have a sizeable bias, as noted, but they have much greater precision about those beliefs (*p*-value < 0.001). Thus the net effect of the bias may be offset by "excess" confidence in that belief. The greater precision is barely significant for those that are married (*p*-value = 0.09), so this suggests that the greater precision (of biased beliefs) is concentrated in the households making joint decisions that are *not* legally married.

We see a similarly nuanced pattern with respect to the risks of death from smoking, shown in Table 5.23. Again there is a general pattern of bias, overestimation of the risks. But it is less significant for young or middle-aged Danes, with *p*-values of 0.13 and 0.24, respectively.

Finally, with respect to the perception of prison sentencing rates if found guilty for crimes, we see again a general bias across almost all demographics and treatments. For property crimes (Table 5.25), no covariates show a statistically significantly level of precision of their biased beliefs compared to the average imprecision. Figure 5.17 shows this result as a diffuse belief. The results are very similar for non-property crimes (Table 5.27). However, younger Danes have statistically insignificant bias from the correct answer (*p*-value = 0.06), and are the only demographic to be unbiased at the 5% significance level.

5.4.2 Atemporal Risk Preferences

Figures 5.19 and 5.20 display estimates of atemporal risk preferences using a RDU specification, with the Prelec probability weighting function as a benchmark. These estimates assume homogeneity and risk preferences across the three types of decision makers. We conclude that the *representative* single individual who is not a member of a household does not behave consistently with EUT (*p*-value = 0.0002 on the hypothesis that $\omega(p)=p$), although the quantitative magnitude of probability weighting is not great. We also find that *representative* couples that make separate decisions do not behave consistently with EUT (*p*-

value = 0.029). These two types of *representative* decision makers also have roughly the same preferences, even though a formal hypothesis test of equality has a *p*-value of only 0.038, driven by the φ parameter.

However, the striking finding is the differences when *representative* couples make a single, joint household decision. For these decisions we *cannot* reject the hypothesis that atemporal risk preferences are *consistent with EUT* (*p*-value = 0.28). Moreover, their preferences are significantly different from those of comparable households making individual, separate decisions (*p*-value < 0.001). Figure 5.19 shows that the joint household decision generates significantly less risk aversion from utility curvature than the other two types of decision maker. Figure 5.20 shows that the extent of probability weighting is not great, and is roughly the same as for singles. But the precision of these estimates of probability weighting does not justify a rejection of EUT.

We undertake comparable estimation of risk preferences at the level of the *individual* decision maker. These estimates were then used to recover the latent subjective beliefs reported in Tables 5.1 through 5.27 and Figures 5.1 through 5.18.

5.5 Conclusions

This chapter builds directly on the previous chapters and transitions from conventional laboratory experiments at Georgia State University's ExCEN laboratory to an artefactual field experiment in the greater Copenhagen metropolitan area with a subject pool of adult Danes recruited from the Registry. In Chapter 4 we examined literacy when students are paired together in an exogenously formed groups of 2 individuals. This chapter naturally extends that design to include endogenously formed households and asks if households as joint decision makers exhibit an effective literacy that differs from individuals. We compare the joint behavior of a Danish household couple to two types of controls; one control is the separate behavior of each individual in a couple, and the other is the behavior of individuals that are not in a household.

As these data were collected for a different project, a total of 9 subjective belief elicitation tasks were asked, of which 4 questions overlap with previous chapters and include topics of rates of interest, inflation, and longevity risk. The remaining 5 questions include topics of basic health risks and beliefs about the risk of receiving a prison sentence in Denmark if convicted of property crimes or violent or sexual "non-property" crimes. We initially focus on the 4 overlapping questions in detail.

When examining the question about the savings account with 2% interest we see some of the highest reported levels of literacy for adult Danes. Singles and households making joint decisions exhibited no statistically significant bias from the correct answer. However, the total effect of households making separate decisions did show bias and they were the least confident in their response, having a standard deviation of 42 kroner. Further, the percentage of raw token allocations to the bin containing the correct answer, the *L* literacy index, is 0.51 for singles, 0.49 for households making separate decisions, and 0.62 for households making joint decisions. Adult Danes score similarly in this index compared with the population of undergraduate students at GSU on a comparable question (fin1), with GSU students scoring *L* literacy index values of 0.53 for single individuals and 0.67 for exogenously formed groups of two individuals.

Next we look at the question about the real interest rate for the Danes. We observe that singles and households making decisions separately or jointly all exhibited statistically significant bias from the correct answer. Households that made joint decisions appear more confident when compared to singles or households making separate decisions. This is confirmed in the bottom panel of Table 5.12: households making joint decisions exhibit a 31% reduction in the standard deviation of beliefs compared to singles; however, this is *not* statistically significant at the 5% level. The Danish population score *L* literacy indices of 0.30 for singles, 0.35 for households making separate decisions, and 0.46 for households making joint decisions. Again we see that households making joint decisions are better off than those making individual decisions. As a further comparison, using the sample of undergraduate students at GSU responding to a comparable question (fin7) we found *L* literacy index values of 0.28 for single individuals and 0.45 for exogenously formed groups consisting of two individuals. The naturally formed household groups in Denmark making joint decisions are comparably literate under this measure as anonymously paired group of GSU students in the laboratory.

Moving on to the last two overlapping questions between the two studies, we turn to literacy with respect to longevity risk for men and women. In Denmark we find that singles and households making separate decisions have biased beliefs, underestimating the longevity of both men and women. However, households making joint decisions do not exhibit a statistically significant bias in general for either question. Danes score *L* literacy indices of 0.41 for singles, 0.45 for households making separate decisions, and 0.46 for households making joint decisions for the question about the remaining life for men. For the question about remaining life for women, Danes score higher *L* literacy indices of 0.45 for singles, 0.56 for households making separate decisions, and 0.51 for households making joint decisions. Using the sample of undergraduate students at GSU responding to comparable questions for longevity risk for men (fin15) and women (fin16) we find lower levels of literacy in the GSU student population. For the question regarding longevity risk in men, the GSU sample had *L* literacy index values of only 0.22 for single individuals and 0.31 for exogenously formed groups. For the question regarding longevity risk in women, the GSU sample had *L* literacy index values of only 0.27 for single individuals and 0.33 for exogenously formed groups. Danes, in general, perform better than GSU students for these questions. The naturally formed household groups in Denmark making joint decisions are more than 1.48 times more literate in the longevity risk of men and almost 1.55 times more literate in the longevity risk of women under this measure than an anonymously paired group of GSU students in the laboratory.

The next three questions address basic health risks in Denmark and included questions about the percentage of deaths attributed to heart disease, cancer, and smoking in a given year. For the heart disease question, we observe that singles and households making decisions separately or jointly all exhibited statistically significant bias from the correct answer. In each case the magnitude of the overestimation was around 15 percentage points higher than the correct answer. For the question about deaths from cancer, we observe that singles and households making a joint decision are unbiased from the correct answer, and that households making joint decisions benefited from having more confidence in their unbiased belief. For the smoking question, we see a similar pattern of bias that is statistically significant for singles and households making separate decisions, overestimating the risk of death from smoking by 6 and 8 percentage points respectively. We see the following levels of *L* literacy indices respectively for singles, households making separate decisions, and households making joint decisions for the following questions: deaths from heart disease of 0.21, 0.18, and 0.18; deaths from cancer of 0.21, 0.21, and 0.32; and deaths from smoking of 0.21, 0.25, and 0.23.

The final two questions ask about the rates of prison sentencing in Denmark for people that were found guilty of property crimes (e.g. burglary and theft) and for "non-property" crimes (e.g. homicide, assault, violence, rape and indecent exposure). For both questions we find that singles, households making separate decisions, and households making joint decisions all exhibit bias from the correct answer and overestimate the actual sentencing rates. However, this general bias is more a result of a lack of confidence than a precisely held, biased belief. These diffuse beliefs lower literacy levels, and is consistent with Danes "knowing that they don't know" what the actual sentencing rate is for either type of crime. We see the following levels of *L* literacy indices respectively for singles, households making separate decisions, and households making joint decisions for the following questions: sentencing rate for property crimes of 0.13, 0.18, and 0.18; and for non-property crimes of 0.08, 0.06, and 0.10. The question regarding the sentencing rate for non-property crime had the lowest levels of literacy out of all nine questions asked.

Finally, Figures 5.21 and 5.22 present some overall displays of bias and confidence. Although not a substitute for the detailed evaluations of marginal and total effects from the main analysis, these serve to provide a high-level view of results. Bias is measured by comparing the average response, denoted by the appropriate label with the true value shown in a dashed red line. The "lack of confidence" is simply the standard deviation of beliefs.

Figure 5.21 shows the overall bias (y-axis) and lack of confidence (x-axis) for each question. These results collate those from Tables 5.1 through 5.9, and refer to the belief estimates pooled over all treatments. In that sense they refer to the pooled responses for this sample (no population weights have been applied), and provide a context to evaluate the comparative treatment effects shown in Figure 5.22.

Statistically significant findings are indicated by a bold label. In Figure 5.21 this refers solely to bias, and indicates an average belief that has a 95% confidence interval that does not include the true value. Hence this is a one-sided hypothesis test at the 5% level. Only three questions generate responses that are *not* significantly biased: the remaining life of men, the fraction of deaths from smoking, and the extent of sentencing for property crime.

In Figure 5.22 statistical significance refers to the joint effect of the household treatment on bias and confidence, and refer to the *p*-values shown in the top panels of Tables 5.10, 5.12, and so on. We only observe the joint household decision as being significantly different from the single individuals control, and for only three questions: the fraction of deaths from heart diseases, the fraction of deaths from cancer, and the extent of sentencing for non-property crime. Of course, these comparative effects are relative to the six questions for which the pooled response is statistically significantly biased, shown in Figure 5.21. Hence, with overall bias for the remaining life of women (Figure 5.21), the fact that there is no statistically significant difference for the household treatments compared to the single individuals control means that there is also bias for households in that instance (Figure 5.22).

The upshot is that there is considerable bias across the different treatments and questions. It is not the case that households make better joint decisions across the board, even if they do make better decisions in certain domains. Moreover, as stressed throughout, one should not examine bias without also examining confidence. The existence of bias is less problematic for decision-making if it is accompanied by an "appropriate" lack of confidence in beliefs. Of course, that conclusion rests on whether the agent exhibiting bias applies the Reduction of Compound Lotteries to their subjective belief distribution: if they do, then any bias is problematic.

In summary, Table 5.28 displays a view of the pooled *L* and *W* indices for our control subjects of single (not married, living alone) individuals, as well as the two treatment groups of married or registered partnership households that were either assigned to make decisions *jointly* together or *separate* from each other. Table 5.29 then focuses only on the *L* index, and assesses the changes when comparing literacy measures between the control group of single individuals against married or registered partnership households making decisions jointly or separate from each other. Over the 4 questions that are classified as numeracy and longevity risk, and have direct overlap the previous thesis chapters, we see a similar story. Joint decisions made by couples resulted in higher *L* indices than either singles or separated couples responding using only their private literacy, for the numeracy questions, here Q1 and Q2. However, this result does not hold when examining longevity risk.

Table 5.1: Interval Regression Estimates for the Question about the Savings Account with 2% Interest

Average: 1103.2 kroner Standard deviation: 18.7 kroner
Correct answer: 1104 kroner N=282 adult Danes
Literacy indices $L = 0.52$ and $W = 0.46$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	1103.17	<0.001	1100.79	1105.56
		LnSigma		
constant	2.93	<0.001	2.81	3.04
		Sigma		
sigma	18.72		16.69	20.99

Figure 5.1: The Savings Account with 2% Interest Question

Average: 1103.2 kroner Standard deviation: 18.7 kroner
Correct answer: 1104 kroner N=282 adult Danes
Literacy indices $L = 0.52$ and $W = 0.46$

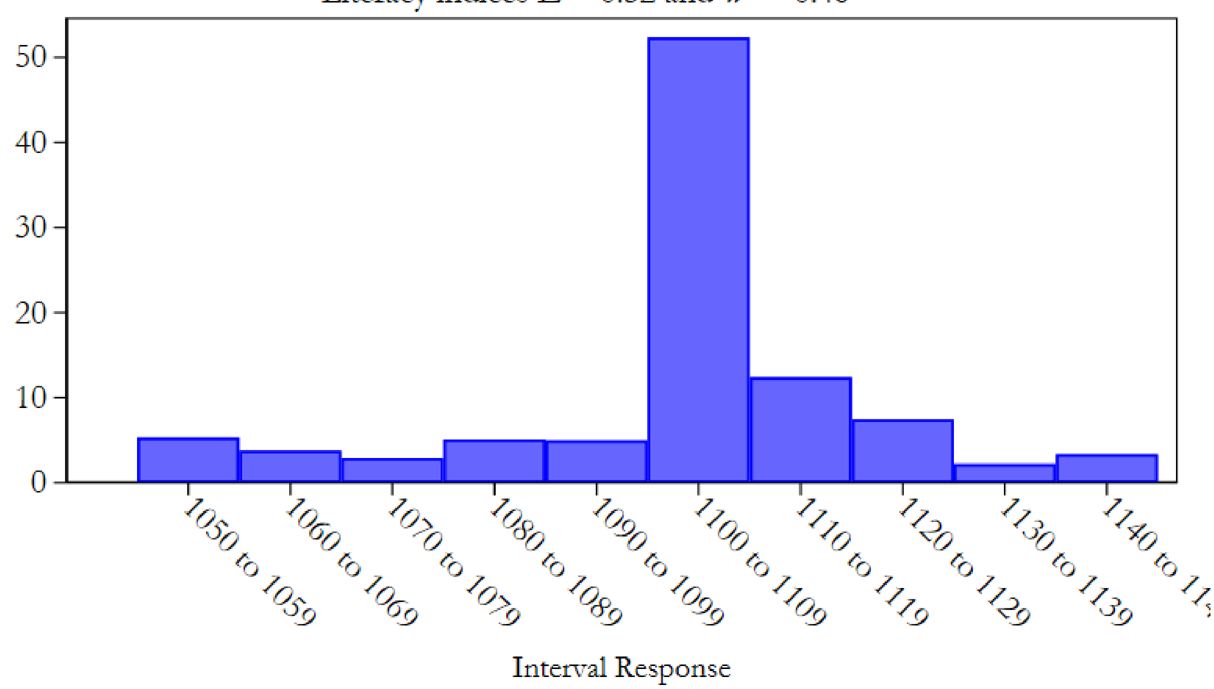


Table 5.2: Interval Regression Estimates for the Question about the Real Interest Rate

Average: 1990.2 kroner Standard deviation: 18.6 kroner
Correct answer: 1980 kroner N=222 adult Danes
Literacy indices $L = 0.36$ and $W = 0.27$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	1990.17	<0.001	1987.83	1992.52
		LnSigma		
constant	2.93	<0.001	2.82	3.04
		Sigma		
sigma	18.65		16.71	20.81

Figure 5.2: The Real Interest Rate Question

Average: 1990.2 kroner Standard deviation: 18.6 kroner
Correct answer: 1980 kroner N=222 adult Danes
Literacy indices $L = 0.36$ and $W = 0.27$

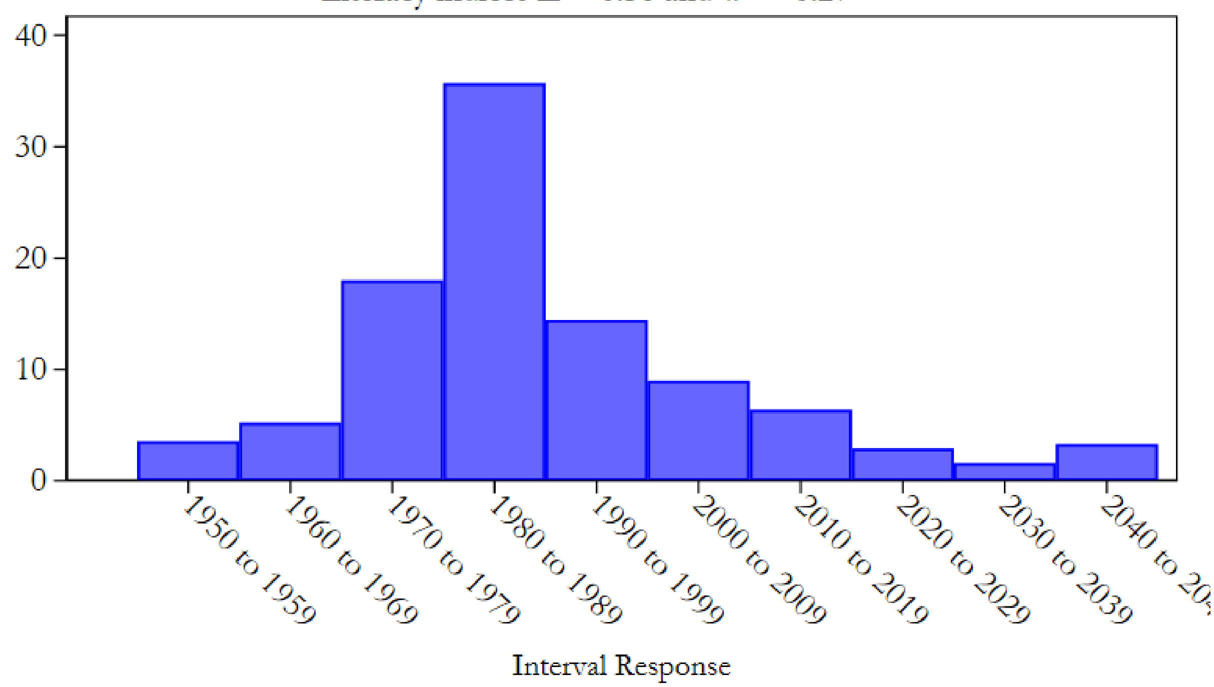


Table 5.3: Interval Regression Estimates for the Question about the Remaining Life for Men

Average: 53.6 years Standard deviation: 18.4 years
Correct answer: 58.4 years N=222 adult Danes
Literacy indices $L = 0.44$ and $W = 0.36$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	53.56	<0.001	51.22	55.89
		LnSigma		
constant	2.91	<0.001	2.79	3.03
		Sigma		
sigma	18.36		16.30	20.67

Figure 5.3: The Remaining Life for Men Question

Average: 53.6 years Standard deviation: 18.4 years
Correct answer: 58.4 years N=222 adult Danes
Literacy indices $L = 0.44$ and $W = 0.36$

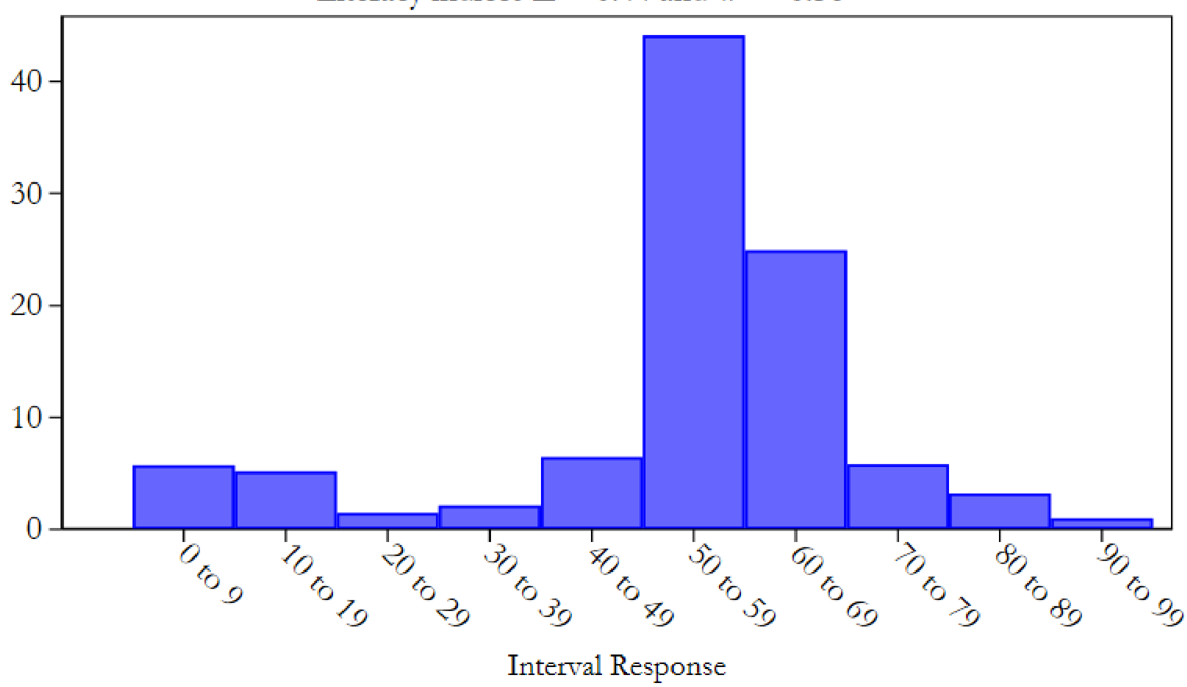


Table 5.4: Interval Regression Estimates for
the Question about the Remaining Life for Women

Average: 57.7 years   Standard deviation: 19.0 years
Correct answer: 62.4 years   N=223 adult Danes
Literacy indices $L = 0.51$ and $W = 0.44$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	57.71	<0.001	55.32	60.10
		LnSigma		
constant	2.94	<0.001	2.82	3.06
		Sigma		
sigma	18.98		16.84	21.39

Figure 5.4: The Remaining Life for Women Question

Average: 57.7 years Standard deviation: 19.0 years
Correct answer: 62.4 years N=223 adult Danes
Literacy indices $L = 0.51$ and $W = 0.44$

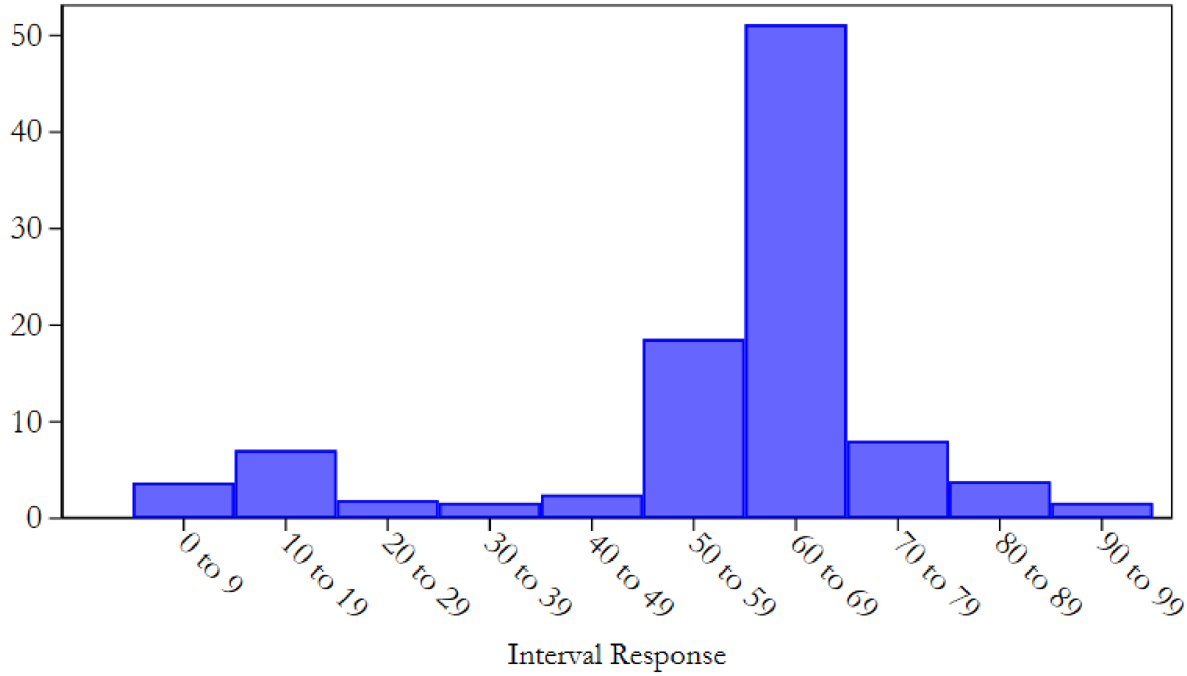


Table 5.5: Interval Regression Estimates for the Question about Deaths from Heart Disease

Average: 29.9% Standard deviation: 16.7%
Correct answer: 14.4% N=223 adult Danes
Literacy indices $L = 0.19$ and $W = 0.12$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	29.95	<0.001	28.00	31.89
		LnSigma		
constant	2.82	<0.001	2.72	2.91
		Sigma		
sigma	16.71		15.21	18.36

Figure 5.5: The Deaths from Heart Disease Question

Average: 29.9% Standard deviation: 16.7%
Correct answer: 14.4% N=223 adult Danes
Literacy indices $L = 0.19$ and $W = 0.12$

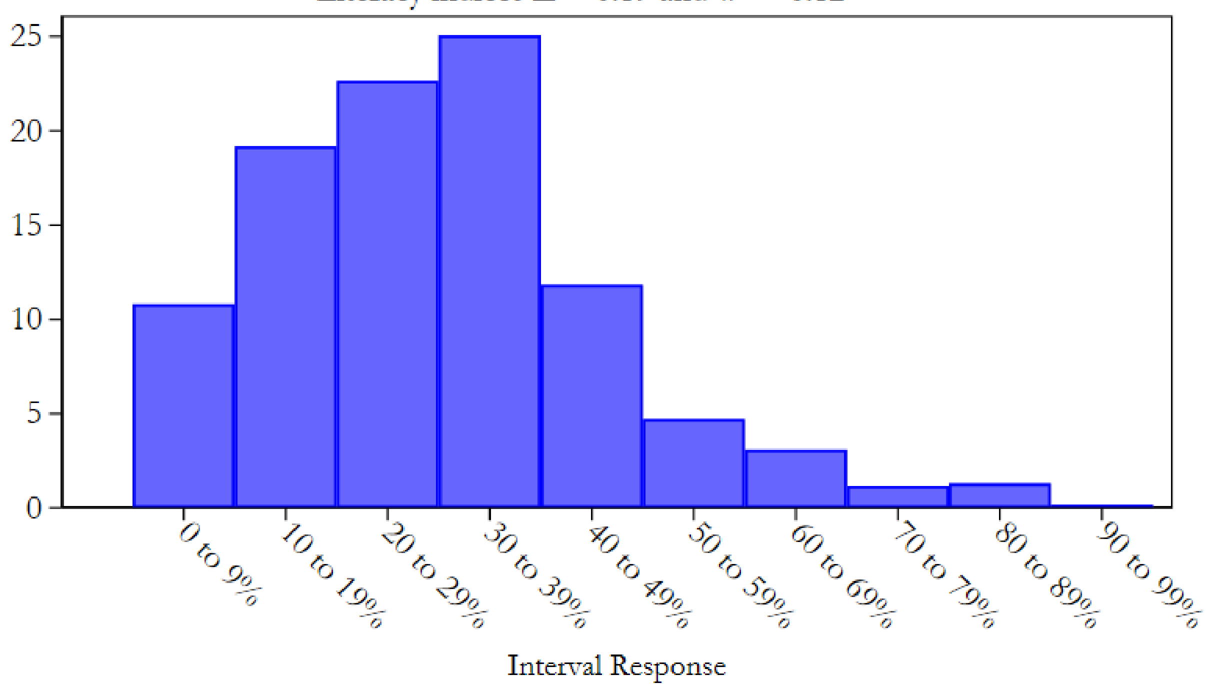


Table 5.6: Interval Regression Estimates for the Question about Deaths from Cancer

Average: 33.8% Standard deviation: 17.2%
Correct answer: 29.0% N=221 adult Danes
Literacy indices $L = 0.23$ and $W = 0.15$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	33.75	<0.001	31.72	35.78
		LnSigma		
constant	2.84	<0.001	2.75	2.94
		Sigma		
sigma	17.17		15.61	18.88

Figure 5.6: The Deaths from Cancer Question

Average: 33.8% Standard deviation: 17.2%
Correct answer: 29.0% N=221 adult Danes
Literacy indices $L = 0.23$ and $W = 0.15$

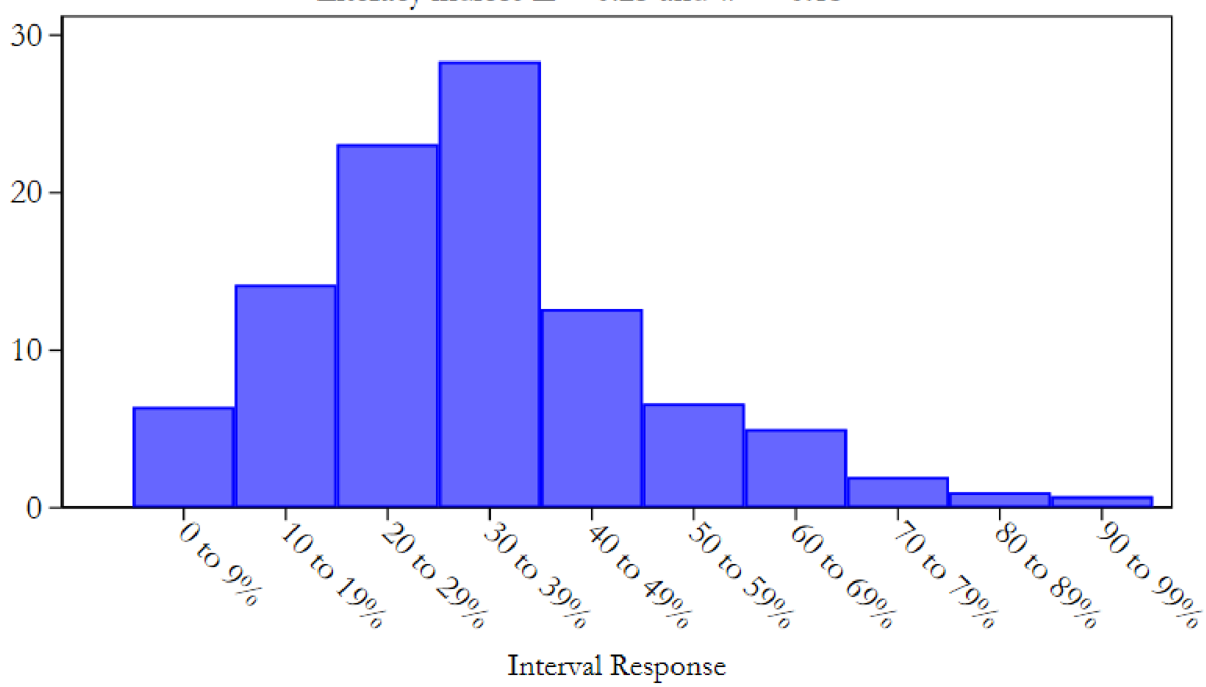


Table 5.7: Interval Regression Estimates for the Question about Deaths from Smoking

Average: 28.3% Standard deviation: 18.4%
Correct answer: 21.5% N=222 adult Danes
Literacy indices $L = 0.23$ and $W = 0.15$

Parameter	Estimate	p-value	95% CI Lower	95% CI Upper
		Average		
constant	28.32	<0.001	26.05	30.60
		LnSigma		
constant	2.91	<0.001	2.81	3.02
		Sigma		
sigma	18.38		16.55	20.41

Figure 5.7: The Deaths from Smoking Question

Average: 28.3% Standard deviation: 18.4%
Correct answer: 21.5% N=222 adult Danes
Literacy indices $L = 0.23$ and $W = 0.15$

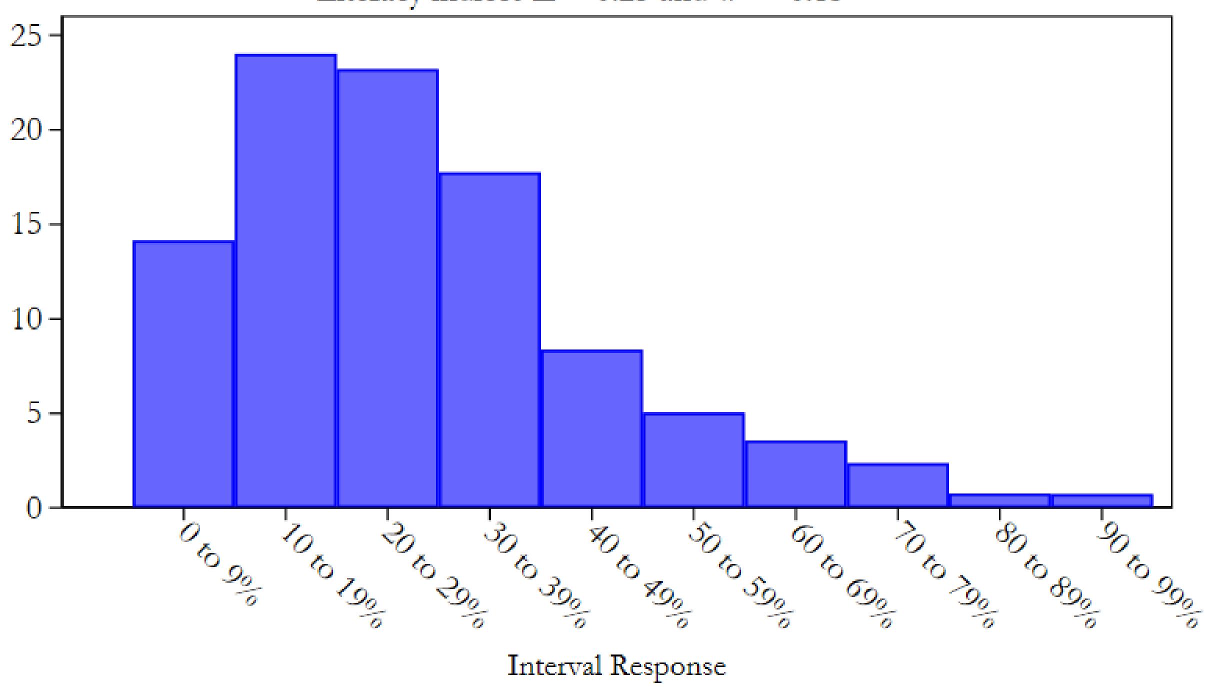


Table 5.8: Interval Regression Estimates for the Question about Property Criminal Activity

Average: 37.0% Standard deviation: 22.9%
Correct answer: 13.9% N=221 adult Danes
Literacy indices $L = 0.16$ and $W = 0.10$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	36.99	<0.001	34.21	39.76
		LnSigma		
constant	3.13	<0.001	3.07	3.20
		Sigma		
sigma	22.92		21.48	24.46

Figure 5.8: The Property Criminal Activity Question

Average: 37.0% Standard deviation: 22.9%
Correct answer: 13.9% N=221 adult Danes
Literacy indices $L = 0.16$ and $W = 0.10$

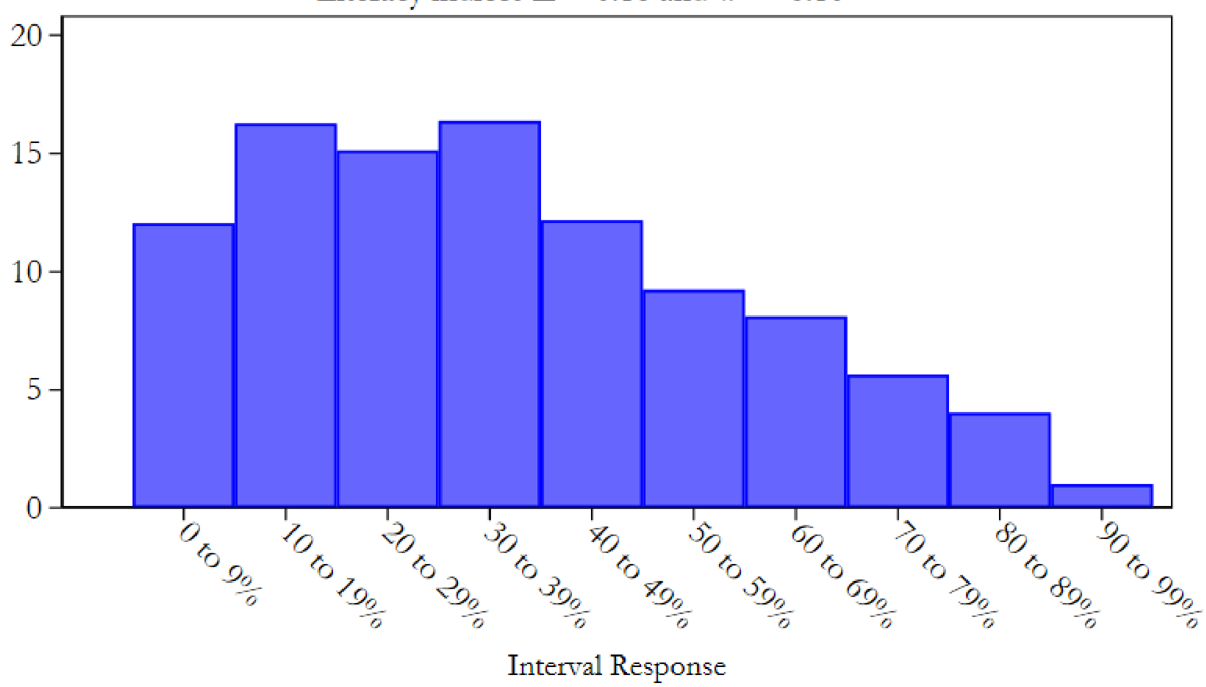


Table 5.9: Interval Regression Estimates for the Question about Non-Property Criminal Activity

Average: 55.6% Standard deviation: 28.4%
Correct answer: 41.8% N=222 adult Danes
Literacy indices $L = 0.08$ and $W = 0.04$

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
constant	55.63	<0.001	52.05	59.22
		LnSigma		
constant	3.35	<0.001	3.29	3.40
		Sigma		
sigma	28.36		26.86	29.95

Figure 5.9: The Non-Property Criminal Activity Question

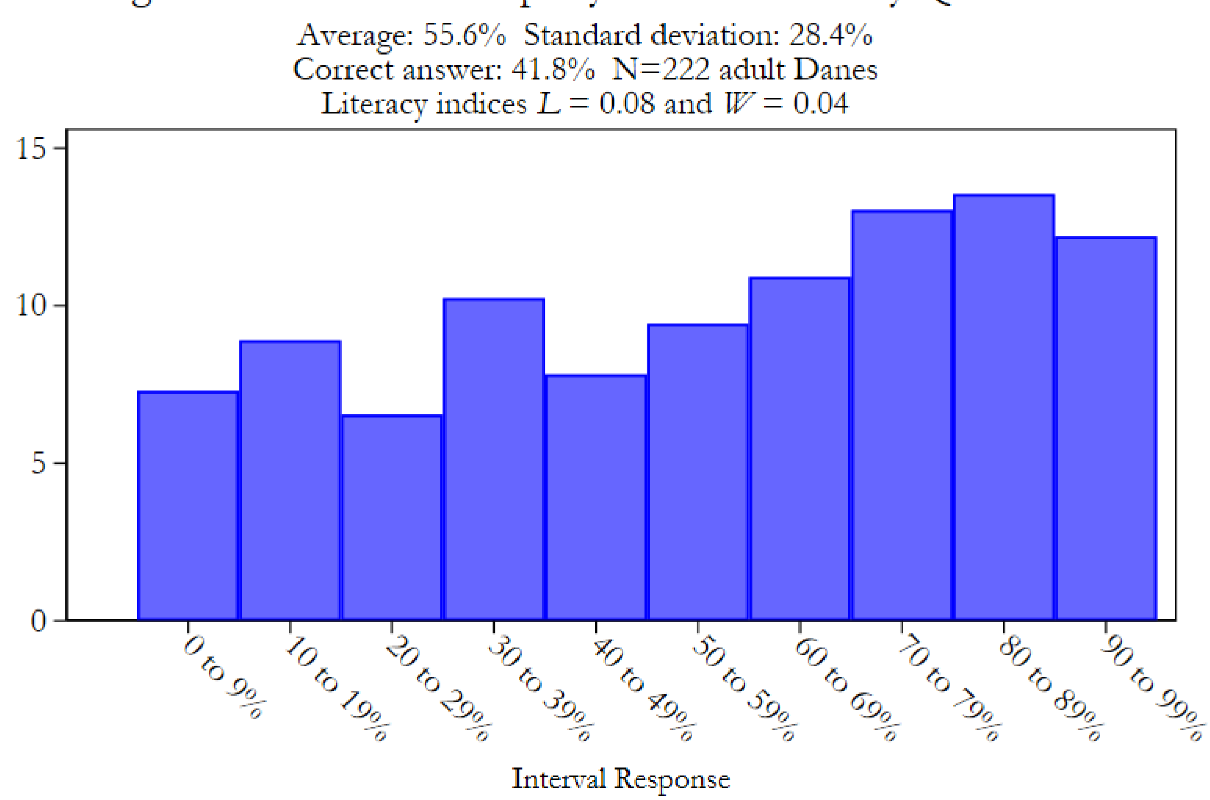


Table 5.10: Effective Households Literacy for the Question about the Savings Account with 2% Interest

Correct answer: 1104 kroner N=222 adult Danes
Singles: Average: 1095.9 kroner Standard deviation: 37.6 kroner
Separate: Average: 1092.5 kroner Standard deviation: 42.4 kroner *p*-value: 0.224
Joint: Average: 1095.1 kroner Standard deviation: 29.4 kroner *p*-value: 0.313

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-3.40	0.18	-8.40	1.61
joint	-0.85	0.71	-5.34	3.64
female	0.72	0.70	-2.88	4.31
young	9.39	<0.001	3.41	15.38
middle	3.64	0.18	-1.66	8.93
old	5.21	0.08	-0.57	10.99
married	-2.08	0.42	-7.18	3.02
highschool	-1.88	0.61	-9.14	5.38
college	-4.04	0.02	-7.55	-0.52
kids	4.52	<0.001	1.86	7.19
inc_rank	10.52	0.07	-0.66	21.70
constant	1095.93	<0.001	1084.86	1107.00
		LnSigma		
separate	0.12	0.57	-0.30	0.54
joint	-0.25	0.35	-0.76	0.27
female	0.11	0.34	-0.12	0.35
young	-0.54	0.02	-1.00	-0.09
middle	-0.22	0.37	-0.69	0.25
old	0.33	0.06	-0.01	0.67
married	0.02	0.91	-0.35	0.40
highschool	-0.12	0.53	-0.48	0.25
college	0.19	0.21	-0.11	0.49
kids	-0.10	0.32	-0.30	0.10
inc_rank	-1.26	<0.001	-1.97	-0.54
constant	3.63	<0.001	2.96	4.30

Figure 5.10: Comparisons of Beliefs for the
Question about the Savings Account with 2% Interest

Correct answer: 1104 kroner

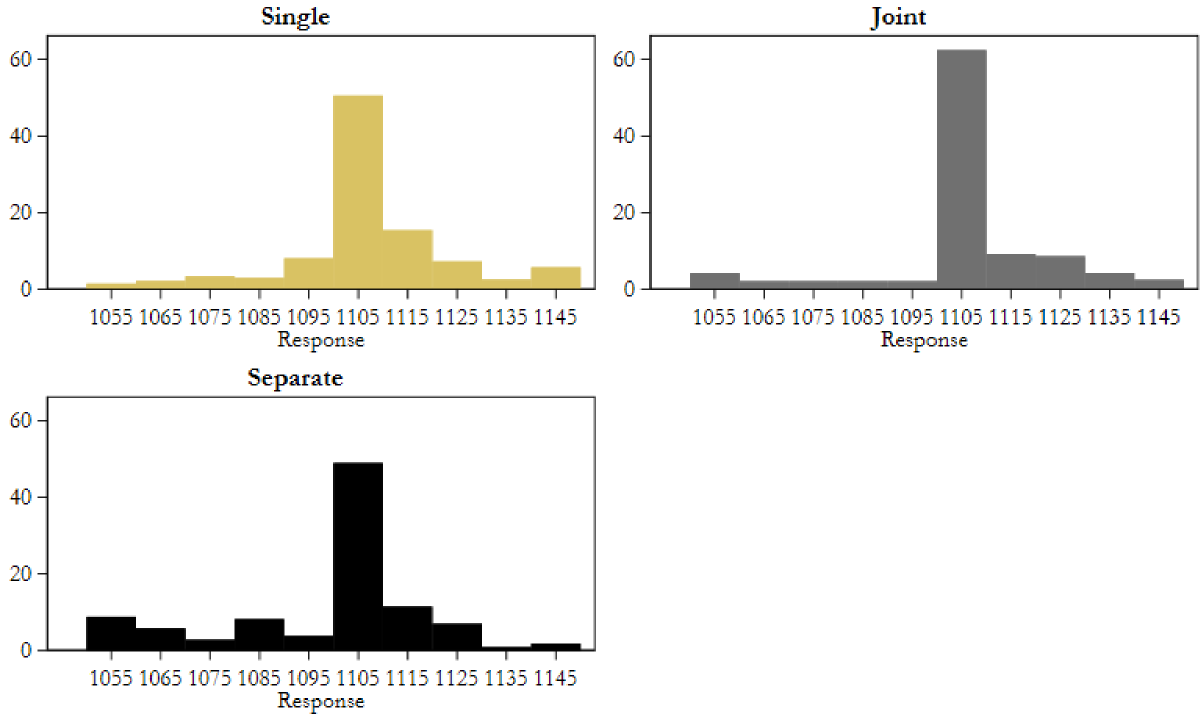


Table 5.11: Literacy Bias and Imprecision, by Demographics, for the Question about the Savings Account with 2% Interest

Bias is relative to the correct answer: 1104 kroner
Additional imprecision is relative to the average imprecision: 18.72
Literacy index L = 0.51 (Singles), 0.49 (Separate) and 0.62 (Joint)
Literacy index W = 0.43 (Singles), 0.42 (Separate) and 0.57 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	2.98	0.10	-0.52	6.48
separate	-5.11	0.01	-8.98	-1.23
joint	1.70	0.32	-3.16	6.56
female	-0.34	0.39	-3.94	3.25
young	2.20	0.22	-1.70	6.10
middle	2.77	0.16	-1.24	6.78
old	-2.58	0.19	-6.68	1.51
married	-2.57	0.15	-6.15	1.02
highschool	-0.53	0.36	-2.99	1.93
college	-1.41	0.27	-4.55	1.72
kids	1.85	0.17	-0.96	4.66
inc_rank	1.12	0.31	-1.97	4.22
		Additional Imprecision of Beliefs		
singles	-2.21	0.18	-5.64	1.23
separate	1.57	0.22	-1.30	4.44
joint	-1.85	0.31	-7.04	3.34
female	1.13	0.30	-1.85	4.10
young	-4.83	0.07	-9.87	0.21
middle	-6.98	0.02	-12.39	-1.57
old	3.10	0.04	0.25	5.96
married	0.93	0.33	-1.98	3.84
highschool	-0.76	0.33	-3.18	1.66
college	-0.79	0.35	-3.81	2.23
kids	-2.33	0.12	-5.27	0.60
inc_rank	-2.70	0.13	-6.23	0.83

Table 5.12: Effective Households Literacy for the Question about the Real Interest Rate

Correct answer: 1980 kroner N=222 adult Danes
Singles: Average: 1998.9 kroner Standard deviation: 23.8 kroner
Separate: Average: 1993.4 kroner Standard deviation: 23.3 kroner *p*-value: 0.398
Joint: Average: 1993.2 kroner Standard deviation: 17.4 kroner *p*-value: 0.200

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-5.54	0.16	-13.19	2.11
joint	-5.69	0.16	-13.61	2.23
female	4.13	0.09	-0.68	8.93
young	-1.44	0.74	-10.07	7.19
middle	-0.41	0.90	-7.09	6.26
old	-1.29	0.71	-8.06	5.49
married	-1.52	0.68	-8.73	5.69
highschool	1.39	0.71	-5.97	8.76
college	-5.87	0.02	-10.64	-1.09
kids	1.25	0.53	-2.66	5.17
inc_rank	-5.46	0.30	-15.80	4.89
constant	1998.90	<0.001	1988.91	2008.88
		LnSigma		
separate	-0.02	0.91	-0.41	0.37
joint	-0.31	0.19	-0.78	0.16
female	0.17	0.20	-0.09	0.42
young	0.18	0.44	-0.27	0.64
middle	0.18	0.33	-0.19	0.55
old	0.09	0.59	-0.24	0.41
married	0.04	0.79	-0.28	0.37
highschool	0.05	0.79	-0.28	0.37
college	-0.27	0.04	-0.52	-0.02
kids	0.06	0.39	-0.08	0.21
inc_rank	-0.57	0.04	-1.10	-0.04
constant	3.17	<0.001	2.66	3.67

Figure 5.11: Comparisons of Beliefs for the Question about the Real Interest Rate

Correct answer: 1980 kroner

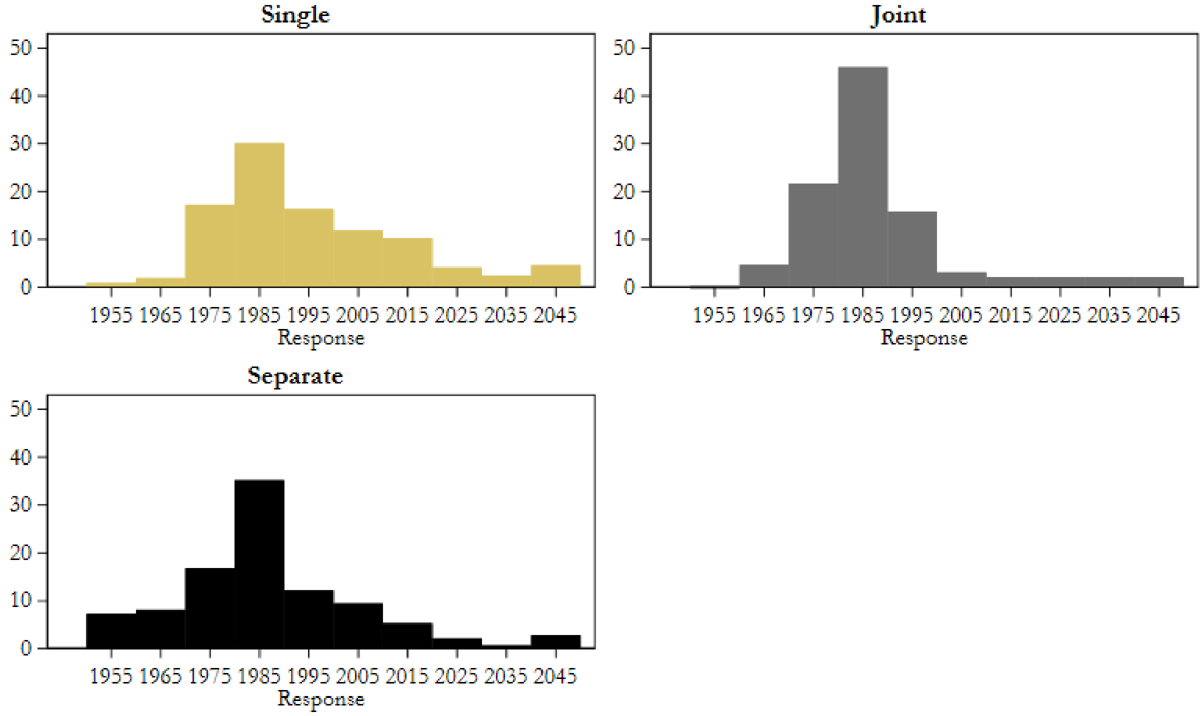


Table 5.13: Literacy Bias and Imprecision, by Demographics, for the Question about the Real Interest Rate

Bias is relative to the correct answer: 1980 kroner
Additional imprecision is relative to the average imprecision: 18.65
Literacy index L = 0.30 (Singles), 0.35 (Separate) and 0.46 (Joint)
Literacy index W = 0.21 (Singles), 0.27 (Separate) and 0.37 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	15.37	<0.001	11.37	19.38
separate	7.09	<0.001	3.56	10.61
joint	7.92	<0.001	3.50	12.33
female	13.10	<0.001	9.43	16.78
young	11.98	<0.001	6.31	17.65
middle	11.95	<0.001	4.14	19.75
old	7.77	<0.001	4.46	11.08
married	6.97	<0.001	3.82	10.12
highschool	9.77	<0.001	7.33	12.22
college	7.19	<0.001	4.60	9.78
kids	10.85	<0.001	7.12	14.58
inc_rank	7.33	<0.001	4.27	10.38
		Additional Imprecision of Beliefs		
singles	0.55	0.37	-2.38	3.48
separate	0.09	0.40	-2.99	3.17
joint	-3.27	0.20	-8.71	2.17
female	1.75	0.18	-0.95	4.45
young	1.48	0.33	-3.08	6.04
middle	2.38	0.33	-5.04	9.79
old	-0.72	0.36	-3.69	2.25
married	-1.15	0.30	-4.16	1.87
highschool	-0.63	0.34	-2.85	1.58
college	-3.46	0.01	-6.02	-0.91
kids	0.71	0.35	-1.98	3.40
inc_rank	-3.97	0.02	-7.22	-0.72

Table 5.14: Effective Households Literacy for the Question about the Remaining Life for Men

Correct answer: 58.4 years N=223 adult Danes
Singles: Average: 44.1 years Standard deviation: 26.3 years
Separate: Average: 42.1 years Standard deviation: 29.7 years *p*-value: 0.399
Joint: Average: 45.1 years Standard deviation: 29.2 years *p*-value: 0.399

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-2.03	0.54	-8.58	4.51
joint	1.00	0.79	-6.52	8.51
female	-1.66	0.40	-5.49	2.18
young	2.51	0.42	-3.56	8.59
middle	-4.26	0.31	-12.45	3.94
old	-3.35	0.16	-8.05	1.35
married	5.47	0.09	-0.94	11.89
highschool	5.44	0.28	-4.38	15.26
college	5.71	<0.001	1.56	9.85
kids	-1.97	0.18	-4.87	0.93
inc_rank	4.08	0.36	-4.63	12.79
constant	44.11	<0.001	31.11	57.10
		LnSigma		
separate	0.12	0.56	-0.28	0.52
joint	0.10	0.66	-0.35	0.56
female	0.12	0.38	-0.15	0.40
young	0.05	0.83	-0.43	0.54
middle	0.33	0.31	-0.31	0.98
old	0.37	0.03	0.03	0.71
married	-0.37	0.05	-0.75	0.00
highschool	-0.39	0.01	-0.69	-0.08
college	-0.20	0.18	-0.50	0.09
kids	0.15	0.26	-0.11	0.42
inc_rank	-0.38	0.26	-1.04	0.28
constant	3.27	<0.001	2.68	3.87

Figure 5.12: Comparisons of Beliefs for the Question about the Remaining Life for Men

Correct answer: 58.4 years

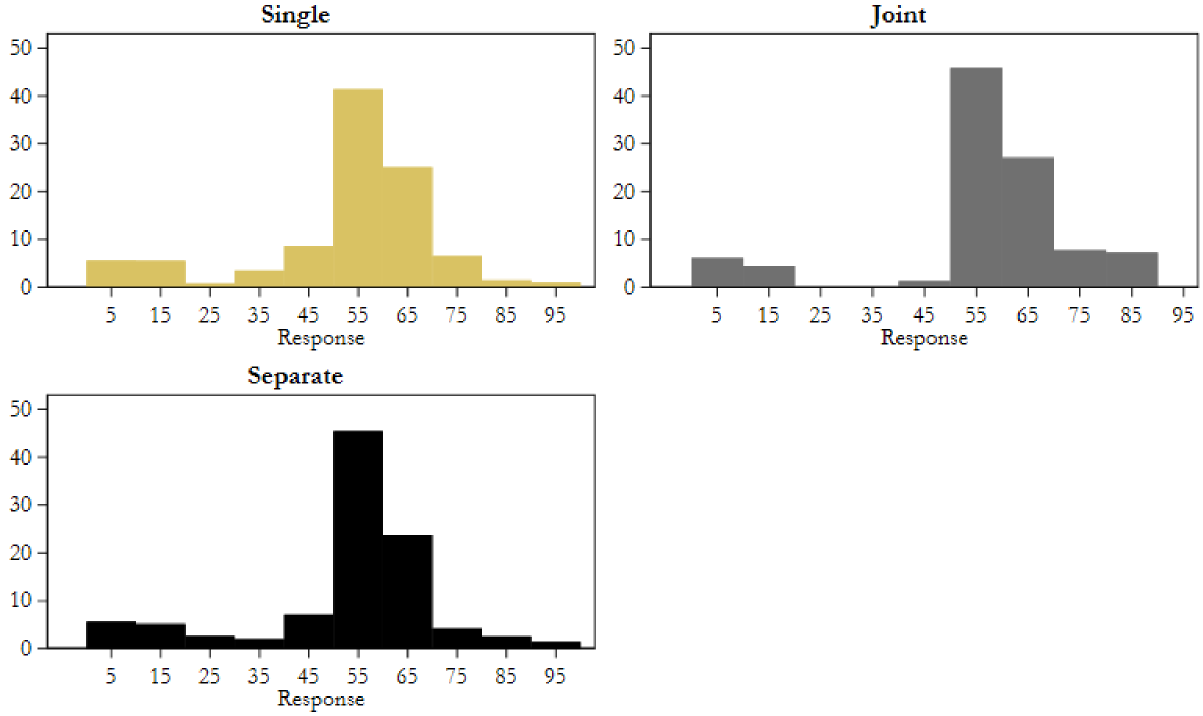


Table 5.15: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Men

Bias is relative to the correct answer: 58.4 years
Additional imprecision is relative to the average imprecision: 18.36
Literacy index L = 0.41 (Singles), 0.45 (Separate) and 0.46 (Joint)
Literacy index W = 0.33 (Singles), 0.37 (Separate) and 0.37 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	-5.57	<0.001	-9.37	-1.76
separate	-5.64	<0.001	-9.17	-2.11
joint	-1.97	0.30	-7.24	3.29
female	-6.35	<0.001	-9.97	-2.73
young	-3.27	0.16	-8.08	1.55
middle	-6.76	0.04	-12.77	-0.74
old	-5.97	<0.001	-9.82	-2.11
married	-3.65	0.04	-6.95	-0.35
highschool	-3.52	<0.001	-5.74	-1.30
college	-1.16	0.27	-3.69	1.38
kids	-5.31	<0.001	-7.95	-2.66
inc_rank	-3.21	0.07	-6.51	0.10
		Additional Imprecision of Beliefs		
singles	-0.20	0.40	-3.67	3.28
separate	-0.03	0.40	-3.27	3.21
joint	0.10	0.40	-5.34	5.54
female	1.58	0.23	-1.42	4.58
young	-0.98	0.37	-5.65	3.68
middle	-2.19	0.32	-8.91	4.52
old	2.20	0.15	-0.86	5.26
married	-0.34	0.39	-3.64	2.96
highschool	-1.95	0.13	-4.48	0.58
college	-3.79	0.03	-7.12	-0.46
kids	-1.14	0.31	-4.36	2.07
inc_rank	-2.00	0.22	-5.58	1.58

Table 5.16: Effective Households Literacy for
the Question about the Remaining Life for Women

Correct answer: 62.4 years N=223 adult Danes
Singles: Average: 45.2 years Standard deviation: 43.2 years
Separate: Average: 41.2 years Standard deviation: 43.2 years *p*-value: 0.380
Joint: Average: 43.3 years Standard deviation: 48.6 years *p*-value: 0.299

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-4.02	0.15	-9.55	1.52
joint	-1.90	0.61	-9.17	5.37
female	-1.49	0.42	-5.09	2.12
young	1.29	0.65	-4.37	6.96
middle	-2.81	0.56	-12.16	6.55
old	-3.94	0.13	-9.06	1.18
married	8.43	0.01	1.90	14.96
highschool	10.10	0.06	-0.48	20.68
college	4.97	0.06	-0.25	10.19
kids	-2.77	0.13	-6.37	0.82
inc_rank	3.53	0.40	-4.68	11.74
constant	45.23	<0.001	32.06	58.41
		LnSigma		
separate	0.00	1.00	-0.37	0.37
joint	0.12	0.60	-0.32	0.56
female	-0.09	0.58	-0.39	0.22
young	-0.33	0.14	-0.77	0.11
middle	0.28	0.41	-0.40	0.97
old	0.23	0.25	-0.16	0.61
married	-0.34	0.08	-0.72	0.04
highschool	-0.38	0.02	-0.69	-0.06
college	-0.32	0.08	-0.68	0.03
kids	0.16	0.31	-0.15	0.46
inc_rank	-0.57	0.12	-1.29	0.15
constant	3.77	<0.001	3.21	4.32

Figure 5.13: Comparisons of Beliefs for the Question about the Remaining Life for Women

Correct answer: 62.4 years

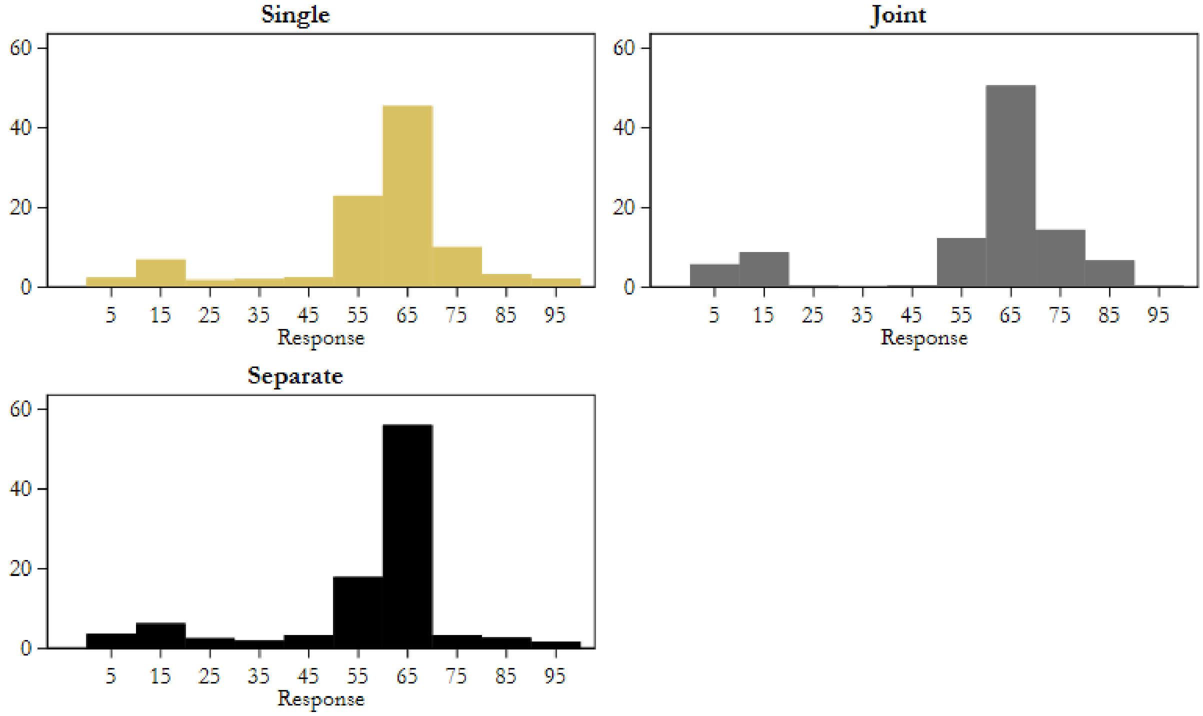


Table 5.17: Literacy Bias and Imprecision, by Demographics, for the Question about the Remaining Life for Women

Bias is relative to the correct answer: 62.4 years
Additional imprecision is relative to the average imprecision: 18.98
Literacy index L = 0.45 (Singles), 0.56 (Separate) and 0.51 (Joint)
Literacy index W = 0.37 (Singles), 0.49 (Separate) and 0.43 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	-4.21	0.04	-8.10	-0.32
separate	-5.50	<0.001	-8.91	-2.09
joint	-3.82	0.18	-9.82	2.19
female	-6.66	<0.001	-10.21	-3.12
young	-4.54	0.06	-9.11	0.03
middle	-4.91	0.12	-11.06	1.25
old	-5.49	<0.001	-9.38	-1.60
married	-3.76	0.04	-7.23	-0.30
highschool	-3.07	0.01	-5.33	-0.80
college	-0.97	0.29	-3.39	1.44
kids	-5.06	<0.001	-7.90	-2.23
inc_rank	-2.98	0.09	-6.40	0.43
		Additional Imprecision of Beliefs		
singles	-0.70	0.37	-4.36	2.96
separate	-0.54	0.38	-3.92	2.83
joint	2.08	0.30	-3.28	7.44
female	0.72	0.36	-2.26	3.71
young	-2.11	0.26	-6.65	2.44
middle	-2.69	0.30	-9.68	4.31
old	2.07	0.18	-1.18	5.32
married	0.36	0.39	-3.10	3.83
highschool	-2.21	0.10	-4.79	0.37
college	-4.67	0.01	-8.17	-1.17
kids	-0.84	0.35	-3.98	2.30
inc_rank	-2.01	0.23	-5.71	1.69

Table 5.18: Effective Households Literacy for the Question about Deaths from Heart Disease

Correct answer: 14.4% N=221 adult Danes
Singles: Average: 33.9% Standard deviation: 20.4%
Separate: Average: 38.3% Standard deviation: 20.8% *p*-value: 0.352
Joint: Average: 38.1% Standard deviation: 18.5% *p*-value: 0.045

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	4.40	0.19	-2.13	10.93
joint	4.15	0.26	-3.14	11.44
female	1.48	0.44	-2.31	5.27
young	0.23	0.96	-8.06	8.52
middle	-6.46	0.06	-13.25	0.33
old	-1.47	0.62	-7.30	4.36
married	-2.88	0.42	-9.88	4.12
highschool	-2.76	0.36	-8.66	3.13
college	3.04	0.17	-1.32	7.41
kids	-1.30	0.49	-5.04	2.43
inc_rank	-4.95	0.35	-15.26	5.36
constant	33.95	<0.001	24.37	43.53
		LnSigma		
separate	0.02	0.88	-0.23	0.27
joint	-0.09	0.54	-0.40	0.21
female	0.04	0.64	-0.14	0.23
young	0.00	0.99	-0.30	0.31
middle	-0.39	<0.001	-0.65	-0.13
old	-0.12	0.38	-0.38	0.15
married	-0.26	0.05	-0.52	-0.00
highschool	0.08	0.58	-0.21	0.38
college	-0.06	0.60	-0.29	0.16
kids	0.05	0.35	-0.05	0.15
inc_rank	-0.13	0.52	-0.54	0.27
constant	3.01	<0.001	2.61	3.42

Figure 5.14: Comparisons of Beliefs for the Question about Deaths from Heart Disease

Correct answer: 14.4%

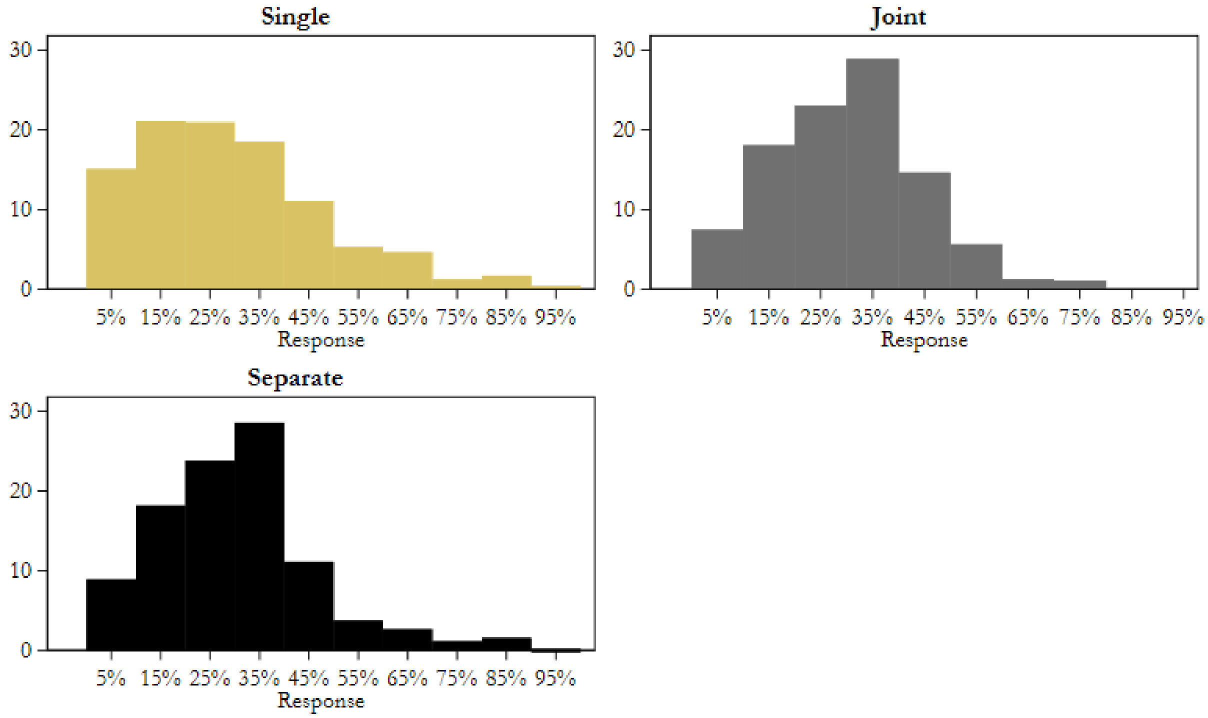


Table 5.19: Literacy Bias and Imprecision, by Demographics, for the Question about Deaths from Heart Disease

Bias is relative to the correct answer: 14.4%
Additional imprecision is relative to the average imprecision: 16.71
Literacy index L = 0.21 (Singles), 0.18 (Separate) and 0.18 (Joint)
Literacy index W = 0.13 (Singles), 0.11 (Separate) and 0.10 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	14.95	<0.001	11.29	18.62
separate	15.88	<0.001	13.08	18.67
joint	15.85	<0.001	12.10	19.60
female	16.72	<0.001	13.78	19.66
young	17.47	<0.001	12.44	22.49
middle	9.75	<0.001	5.46	14.05
old	15.45	<0.001	12.79	18.10
married	15.30	<0.001	12.98	17.62
highschool	15.30	<0.001	13.22	17.38
college	16.22	<0.001	13.67	18.76
kids	15.04	<0.001	12.04	18.03
inc_rank	13.43	<0.001	10.47	16.39
		Additional Imprecision of Beliefs		
singles	2.09	0.12	-0.59	4.77
separate	-0.64	0.35	-3.16	1.87
joint	-2.67	0.04	-5.17	-0.17
female	0.75	0.33	-1.72	3.21
young	3.12	0.04	0.21	6.03
middle	-3.63	<0.001	-5.26	-2.00
old	-1.58	0.20	-4.27	1.10
married	-2.70	0.01	-4.68	-0.72
highschool	-0.09	0.40	-1.74	1.56
college	-0.74	0.30	-2.65	1.17
kids	0.24	0.38	-1.45	1.94
inc_rank	-1.60	0.14	-3.76	0.56

Table 5.20: Effective Households Literacy for the Question about Deaths from Cancer

Correct answer: 29.0% N=222 adult Danes
Singles: Average: 39.3% Standard deviation: 23.2%
Separate: Average: 44.1% Standard deviation: 26.0% *p*-value: 0.259
Joint: Average: 40.5% Standard deviation: 15.4% *p*-value: < 0.001

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	4.81	0.17	-2.09	11.72
joint	1.26	0.71	-5.42	7.94
female	1.75	0.35	-1.91	5.40
young	-2.73	0.49	-10.48	5.03
middle	-2.58	0.49	-9.95	4.79
old	-2.29	0.40	-7.67	3.09
married	-3.11	0.35	-9.57	3.35
highschool	-1.15	0.73	-7.67	5.37
college	1.30	0.53	-2.80	5.40
kids	-0.90	0.40	-3.02	1.21
inc_rank	-6.70	0.11	-14.94	1.53
constant	39.27	<0.001	29.27	49.26
		LnSigma		
separate	0.12	0.37	-0.14	0.37
joint	-0.41	<0.001	-0.68	-0.14
female	-0.03	0.74	-0.20	0.14
young	-0.03	0.84	-0.34	0.27
middle	-0.13	0.40	-0.43	0.17
old	-0.15	0.22	-0.39	0.09
married	-0.13	0.34	-0.40	0.14
highschool	-0.22	0.09	-0.49	0.04
college	0.06	0.61	-0.15	0.26
kids	-0.17	0.02	-0.32	-0.03
inc_rank	0.06	0.78	-0.39	0.52
constant	3.14	<0.001	2.74	3.55

Figure 5.15: Comparisons of Beliefs for the Question about Deaths from Cancer

Correct answer: 29.0%

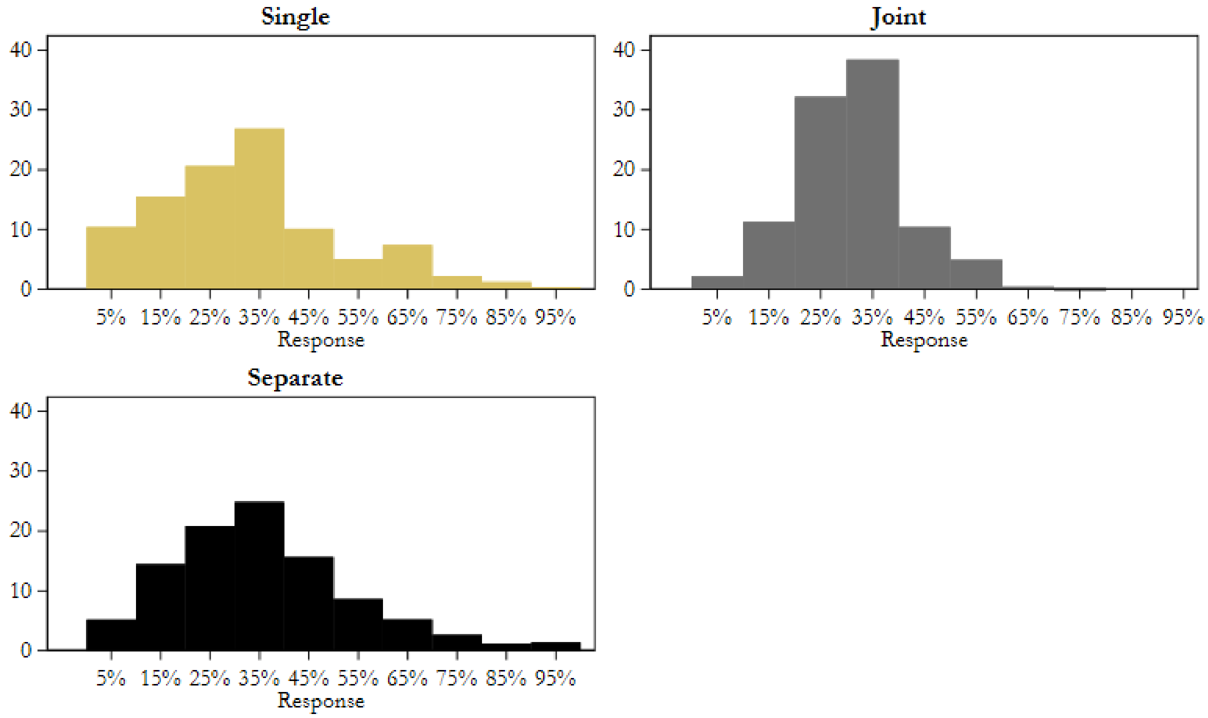


Table 5.21: Literacy Bias and Imprecision, by Demographics, for the Question about Deaths from Cancer

Bias is relative to the correct answer: 29.0%
Additional imprecision is relative to the average imprecision: 17.17
Literacy index L = 0.21 (Singles), 0.21 (Separate) and 0.32 (Joint)
Literacy index W = 0.13 (Singles), 0.13 (Separate) and 0.22 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	3.72	0.06	-0.00	7.43
separate	6.83	<0.001	3.60	10.07
joint	2.07	0.13	-0.60	4.75
female	5.79	<0.001	3.02	8.55
young	5.82	0.04	0.50	11.15
middle	3.38	0.20	-2.18	8.93
old	4.70	<0.001	1.92	7.48
married	4.65	<0.001	2.13	7.18
highschool	4.33	<0.001	2.24	6.43
college	4.60	<0.001	1.89	7.31
kids	4.11	<0.001	1.94	6.28
inc_rank	2.25	0.12	-0.63	5.12
		Additional Imprecision of Beliefs		
singles	1.31	0.25	-1.29	3.90
separate	1.17	0.26	-1.30	3.64
joint	-6.74	<0.001	-8.68	-4.79
female	-0.56	0.36	-3.11	1.98
young	2.35	0.19	-1.46	6.17
middle	-1.48	0.28	-4.93	1.97
old	-1.03	0.28	-3.40	1.34
married	-1.95	0.09	-4.16	0.27
highschool	-0.68	0.30	-2.51	1.14
college	-0.20	0.39	-2.65	2.26
kids	-2.01	<0.001	-3.44	-0.58
inc_rank	-1.89	0.10	-4.10	0.33

Table 5.22: Effective Households Literacy for the Question about Deaths from Smoking

Correct answer: 21.5% N=221 adult Danes
Singles: Average: 36.6% Standard deviation: 14.2%
Separate: Average: 36.4% Standard deviation: 12.3% *p*-value: 0.374
Joint: Average: 31.5% Standard deviation: 9.9% *p*-value: 0.381

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-0.11	0.97	-5.78	5.56
joint	-5.05	0.14	-11.78	1.68
female	5.28	0.03	0.41	10.15
young	-2.94	0.41	-9.99	4.11
middle	-1.12	0.79	-9.39	7.15
old	2.59	0.41	-3.61	8.79
married	3.85	0.19	-1.91	9.61
highschool	-2.08	0.50	-8.15	3.99
college	-0.28	0.91	-5.21	4.65
kids	0.60	0.61	-1.72	2.91
inc_rank	-14.29	<0.001	-24.83	-3.76
constant	36.55	<0.001	26.93	46.17
		LnSigma		
separate	-0.14	0.23	-0.38	0.09
joint	-0.36	0.05	-0.72	0.00
female	0.26	0.02	0.04	0.48
young	-0.01	0.93	-0.35	0.32
middle	0.02	0.91	-0.32	0.36
old	-0.11	0.47	-0.39	0.18
married	0.35	<0.001	0.08	0.62
highschool	0.24	0.10	-0.04	0.51
college	0.02	0.91	-0.25	0.28
kids	-0.16	0.11	-0.35	0.04
inc_rank	-0.12	0.63	-0.63	0.38
constant	2.65	<0.001	2.16	3.14

Figure 5.16: Comparisons of Beliefs for the Question about Deaths from Smoking

Correct answer: 21.5%

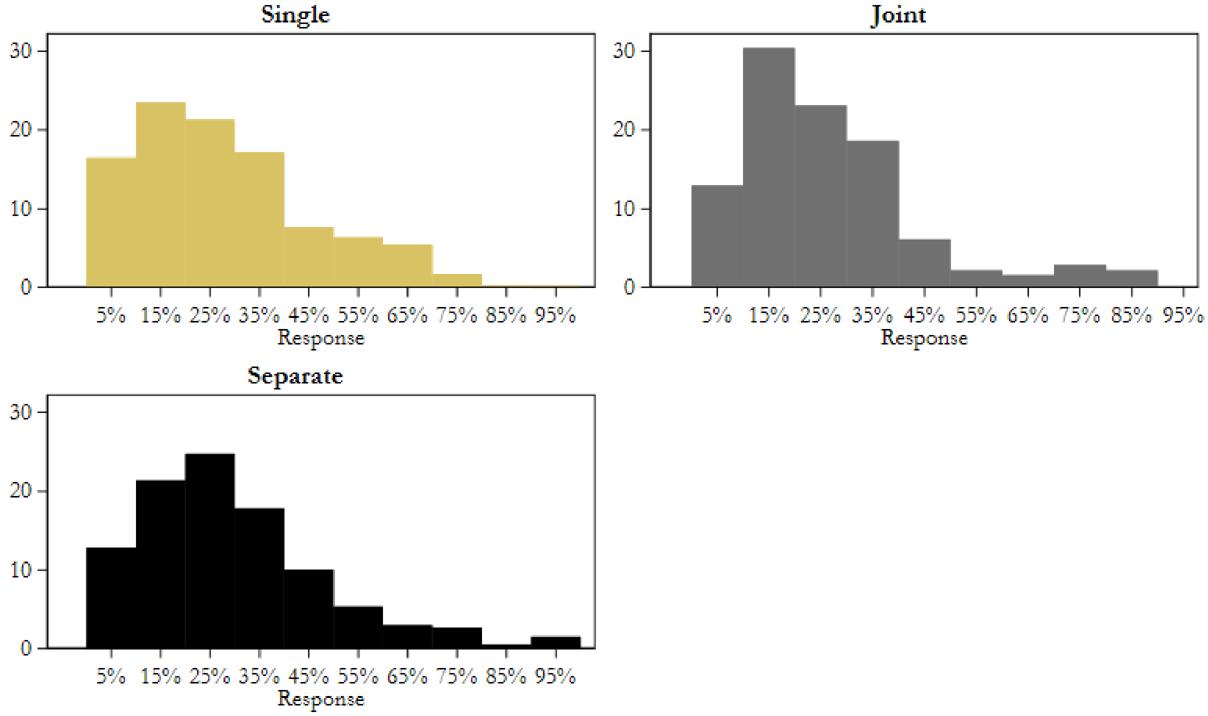


Table 5.23: Literacy Bias and Imprecision, by Demographics, for the Question about Deaths from Smoking

Bias is relative to the correct answer: 21.5%
Additional imprecision is relative to the average imprecision: 18.38
Literacy index $L = 0.21$ (Singles), 0.25 (Separate) and 0.23 (Joint)
Literacy index $W = 0.13$ (Singles), 0.16 (Separate) and 0.15 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	6.28	<0.001	2.59	9.98
separate	8.14	<0.001	4.60	11.68
joint	4.97	0.06	0.07	9.86
female	9.66	<0.001	6.19	13.13
young	3.18	0.13	-0.97	7.34
middle	3.15	0.24	-2.97	9.26
old	9.50	<0.001	5.98	13.01
married	8.54	<0.001	5.07	12.01
highschool	6.02	<0.001	3.56	8.47
college	5.57	<0.001	2.43	8.71
kids	7.05	<0.001	4.41	9.69
inc_rank	3.56	0.06	0.04	7.08
		Additional Imprecision of Beliefs		
singles	-0.40	0.38	-2.80	2.01
separate	0.57	0.37	-2.54	3.69
joint	-0.78	0.38	-5.81	4.24
female	1.29	0.24	-1.27	3.85
young	-2.64	0.12	-6.01	0.72
middle	-1.59	0.31	-6.05	2.88
old	0.82	0.35	-2.23	3.87
married	0.93	0.34	-2.19	4.05
highschool	-0.03	0.40	-2.23	2.17
college	-0.16	0.40	-2.90	2.58
kids	-1.01	0.29	-3.51	1.48
inc_rank	-0.86	0.34	-3.98	2.26

Table 5.24: Effective Households Literacy for the Question about Property Criminal Activity

Correct answer: 13.9% N=222 adult Danes
Singles: Average: 38.5% Standard deviation: 28.0%
Separate: Average: 38.2% Standard deviation: 28.7% *p*-value: 0.250
Joint: Average: 32.1% Standard deviation: 23.7% *p*-value: 0.089

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-0.31	0.95	-9.46	8.85
joint	-6.42	0.24	-17.16	4.32
female	-3.70	0.23	-9.70	2.30
young	1.65	0.74	-8.18	11.49
middle	1.93	0.72	-8.76	12.63
old	3.21	0.42	-4.59	11.01
married	-3.29	0.47	-12.18	5.59
highschool	5.30	0.28	-4.27	14.86
college	3.04	0.38	-3.73	9.81
kids	3.07	0.10	-0.63	6.76
inc_rank	-8.08	0.27	-22.35	6.19
constant	38.52	<0.001	24.20	52.84
		LnSigma		
separate	0.03	0.80	-0.17	0.22
joint	-0.17	0.25	-0.46	0.12
female	-0.16	0.03	-0.31	-0.02
young	0.09	0.46	-0.15	0.33
middle	0.13	0.24	-0.09	0.35
old	0.01	0.91	-0.19	0.22
married	0.08	0.47	-0.14	0.29
highschool	-0.05	0.69	-0.32	0.21
college	0.01	0.89	-0.16	0.18
kids	-0.01	0.90	-0.13	0.12
inc_rank	-0.24	0.27	-0.66	0.19
constant	3.33	<0.001	2.96	3.70

Figure 5.17: Comparisons of Beliefs for the Question about Property Criminal Activity

Correct answer: 13.9%

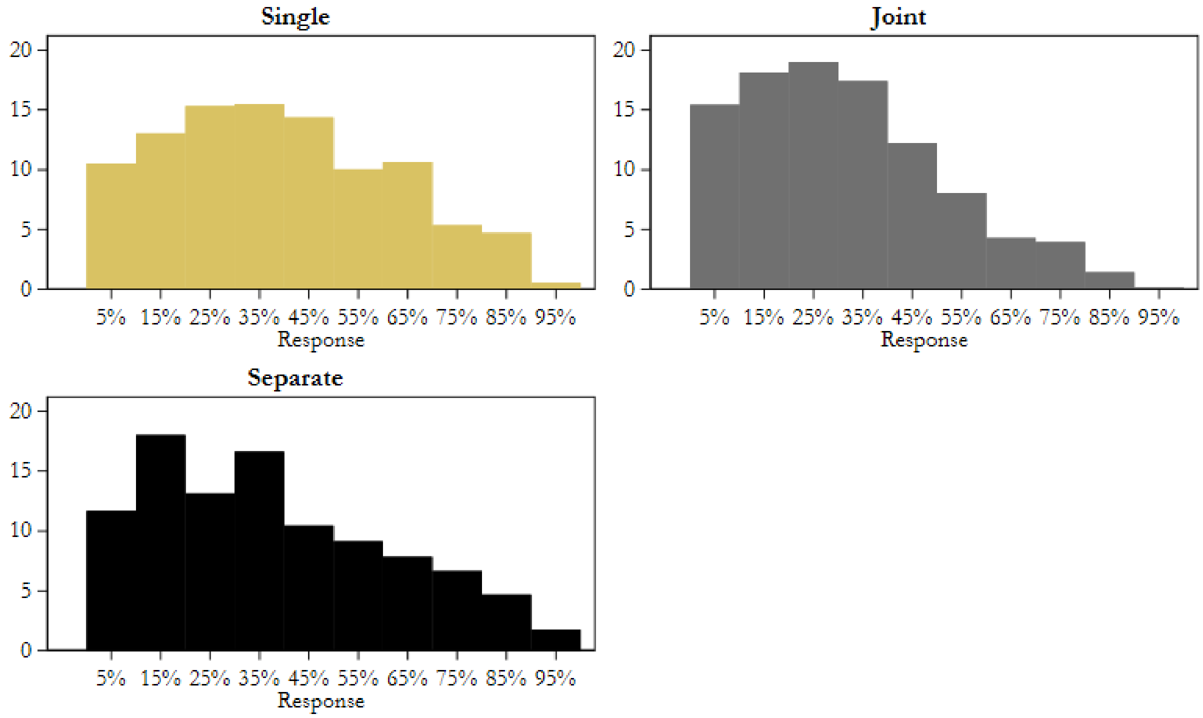


Table 5.25: Literacy Bias and Imprecision, by Demographics, for the Question about Property Criminal Activity

Bias is relative to the correct answer: 13.9%
Additional imprecision is relative to the average imprecision: 22.92
Literacy index L = 0.13 (Singles), 0.18 (Separate) and 0.18 (Joint)
Literacy index W = 0.07 (Singles), 0.11 (Separate) and 0.11 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	25.28	<0.001	20.75	29.81
separate	23.98	<0.001	19.60	28.37
joint	17.47	<0.001	12.01	22.92
female	21.60	<0.001	17.93	25.27
young	25.05	<0.001	18.45	31.64
middle	26.82	<0.001	18.44	35.21
old	21.70	<0.001	17.66	25.75
married	20.85	<0.001	16.83	24.87
highschool	23.91	<0.001	21.01	26.82
college	24.91	<0.001	21.24	28.58
kids	24.88	<0.001	21.17	28.58
inc_rank	22.85	<0.001	18.60	27.11
		Additional Imprecision of Beliefs		
singles	-0.33	0.38	-2.60	1.94
separate	1.13	0.25	-1.16	3.43
joint	-3.02	0.09	-6.44	0.40
female	-1.59	0.12	-3.61	0.43
young	1.68	0.23	-1.43	4.80
middle	0.68	0.37	-2.76	4.13
old	-0.50	0.37	-2.98	1.98
married	0.02	0.40	-2.35	2.39
highschool	-0.50	0.33	-2.04	1.04
college	-0.80	0.29	-2.78	1.17
kids	-0.16	0.39	-2.25	1.93
inc_rank	-0.93	0.31	-3.49	1.64

Table 5.26: Effective Households Literacy for the Question about Non-Property Criminal Activity

Correct answer: 41.8% N=221 adult Danes
Singles: Average: 42.0% Standard deviation: 30.3%
Separate: Average: 38.5% Standard deviation: 31.9% *p*-value: 0.263
Joint: Average: 40.5% Standard deviation: 25.1% *p*-value: 0.015

Parameter	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Average		
separate	-3.50	0.58	-15.83	8.83
joint	-1.46	0.83	-14.53	11.60
female	-1.72	0.66	-9.39	5.96
young	-2.05	0.78	-16.21	12.10
middle	-4.91	0.47	-18.30	8.48
old	0.36	0.95	-11.51	12.23
married	12.40	0.05	0.21	24.60
highschool	9.12	0.12	-2.32	20.56
college	0.02	1.00	-8.27	8.31
kids	-1.47	0.68	-8.51	5.57
inc_rank	5.18	0.59	-13.42	23.78
constant	41.95	<0.001	23.23	60.68
		LnSigma		
separate	0.05	0.59	-0.13	0.24
joint	-0.19	0.10	-0.41	0.04
female	-0.11	0.13	-0.25	0.03
young	-0.09	0.49	-0.36	0.17
middle	-0.10	0.40	-0.32	0.13
old	0.01	0.95	-0.26	0.28
married	-0.07	0.48	-0.27	0.12
highschool	0.02	0.85	-0.18	0.21
college	-0.03	0.67	-0.19	0.12
kids	0.01	0.94	-0.16	0.17
inc_rank	0.03	0.83	-0.27	0.34
constant	3.41	<0.001	3.06	3.76

Figure 5.18: Comparisons of Beliefs for the
Question about Non-Property Criminal Activity

Correct answer: 41.8%

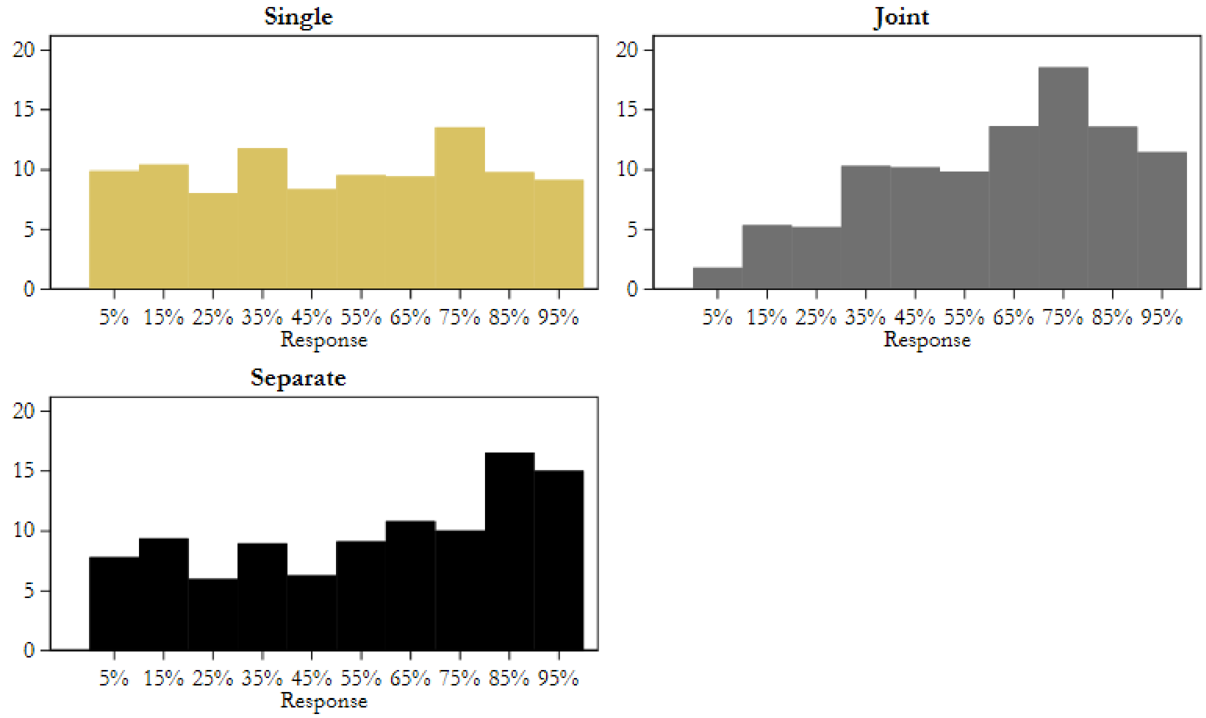


Table 5.27: Literacy Bias and Imprecision, by Demographics, for the Question about Non-Property Criminal Activity

Bias is relative to the correct answer: 41.8%
Additional imprecision is relative to the average imprecision: 28.36
Literacy index $L = 0.08$ (Singles), 0.06 (Separate) and 0.10 (Joint)
Literacy index $W = 0.04$ (Singles), 0.03 (Separate) and 0.06 (Joint)

Demographic	Estimate	*p*-value	95% CI Lower	95% CI Upper
		Bias from the True Response		
singles	8.72	<0.001	2.66	14.78
separate	15.38	<0.001	9.76	21.00
joint	19.24	<0.001	12.57	25.91
female	12.82	<0.001	7.91	17.72
young	7.38	0.06	0.03	14.72
middle	11.24	0.04	0.73	21.76
old	16.43	<0.001	11.06	21.80
married	19.39	<0.001	14.45	24.33
highschool	15.10	<0.001	11.27	18.93
college	15.65	<0.001	10.83	20.47
kids	14.55	<0.001	9.88	19.22
inc_rank	17.82	<0.001	12.23	23.41
		Additional Imprecision of Beliefs		
singles	0.13	0.40	-2.33	2.60
separate	1.11	0.26	-1.27	3.49
joint	-4.32	0.02	-7.63	-1.01
female	-1.00	0.26	-3.17	1.16
young	-1.46	0.25	-4.42	1.51
middle	0.25	0.40	-3.73	4.23
old	0.33	0.38	-2.00	2.67
married	-1.15	0.27	-3.74	1.43
highschool	-0.26	0.38	-1.98	1.46
college	-0.66	0.33	-2.81	1.49
kids	-0.30	0.39	-2.79	2.19
inc_rank	-0.26	0.39	-2.84	2.33

Figure 5.19: Utility Functions of Singles and Households

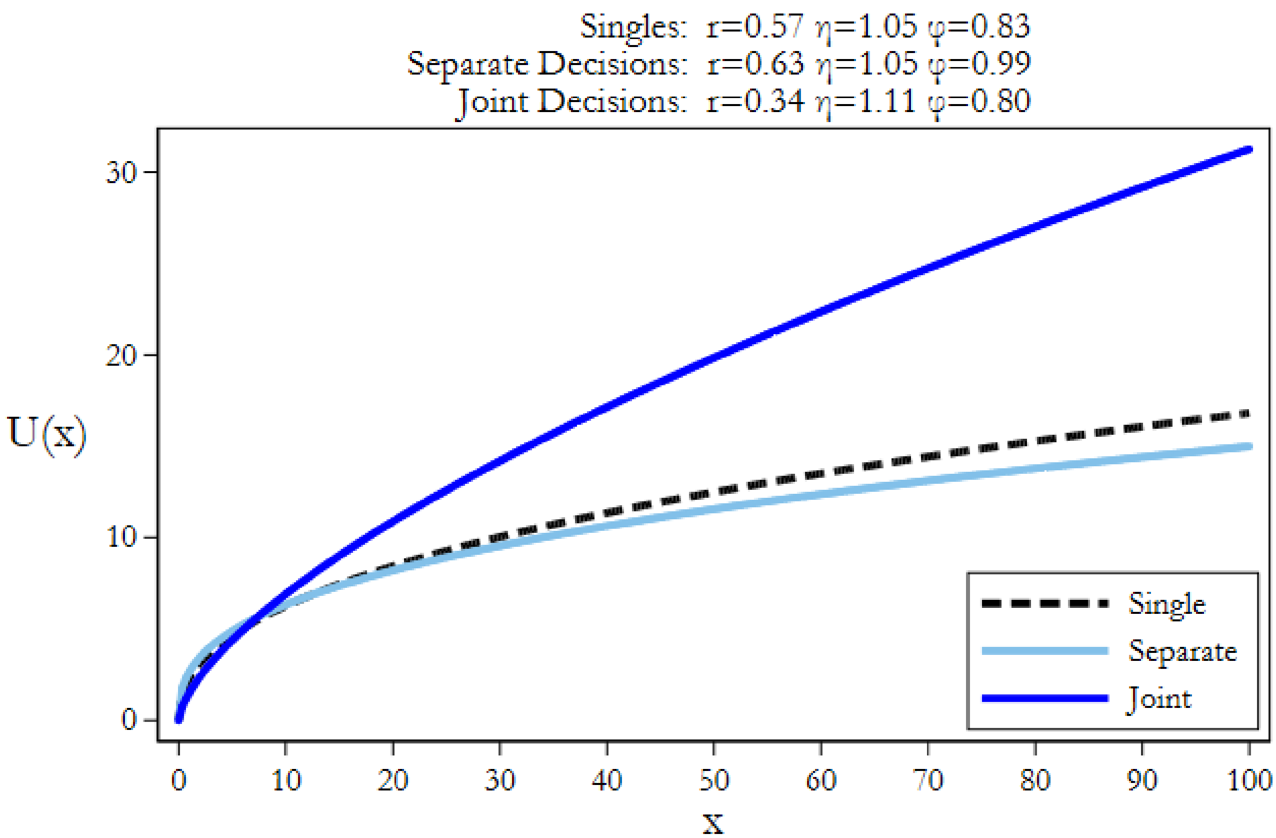


Figure 5.20 Probability Weighting of Singles and Households

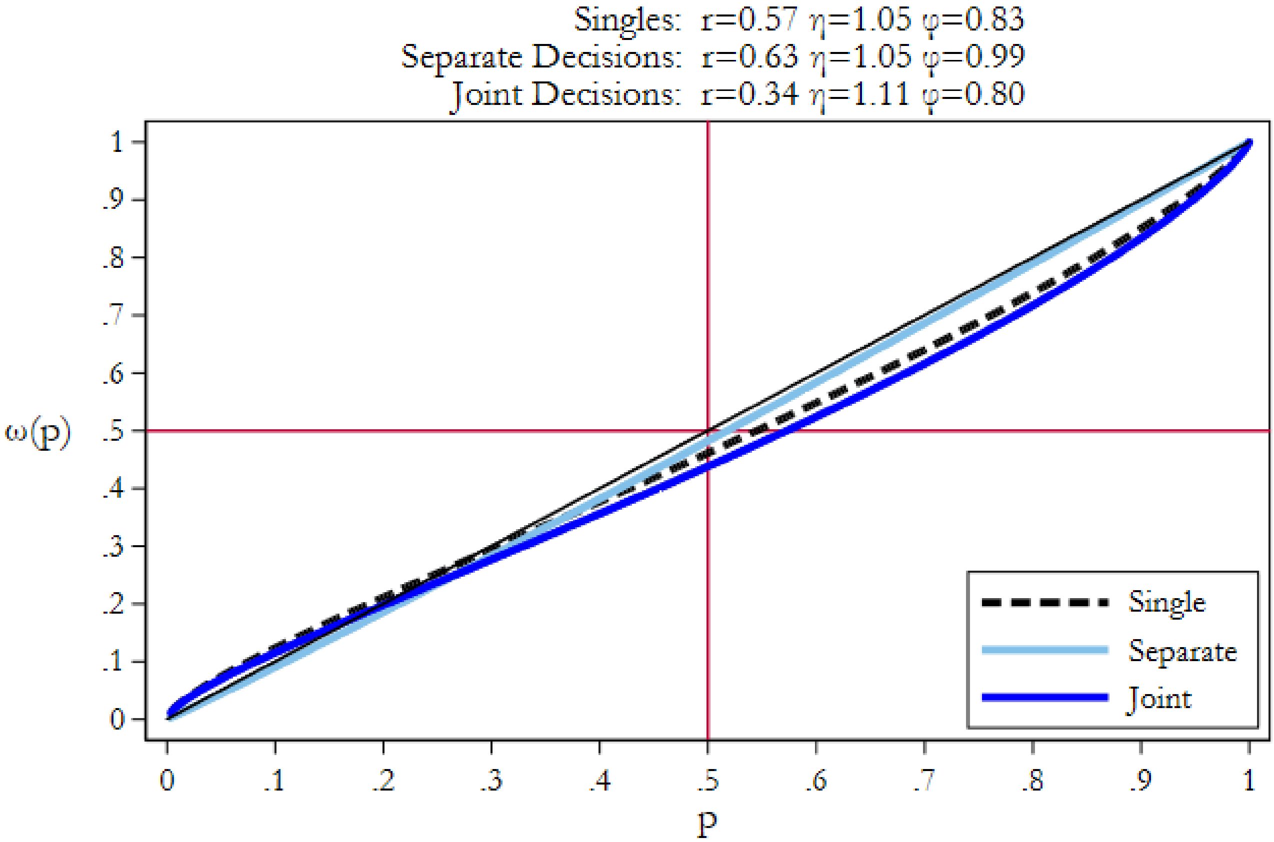


Figure 5.21: Overall Bias and Confidence

Bold indicates statistically significant bias at the 5% significance level

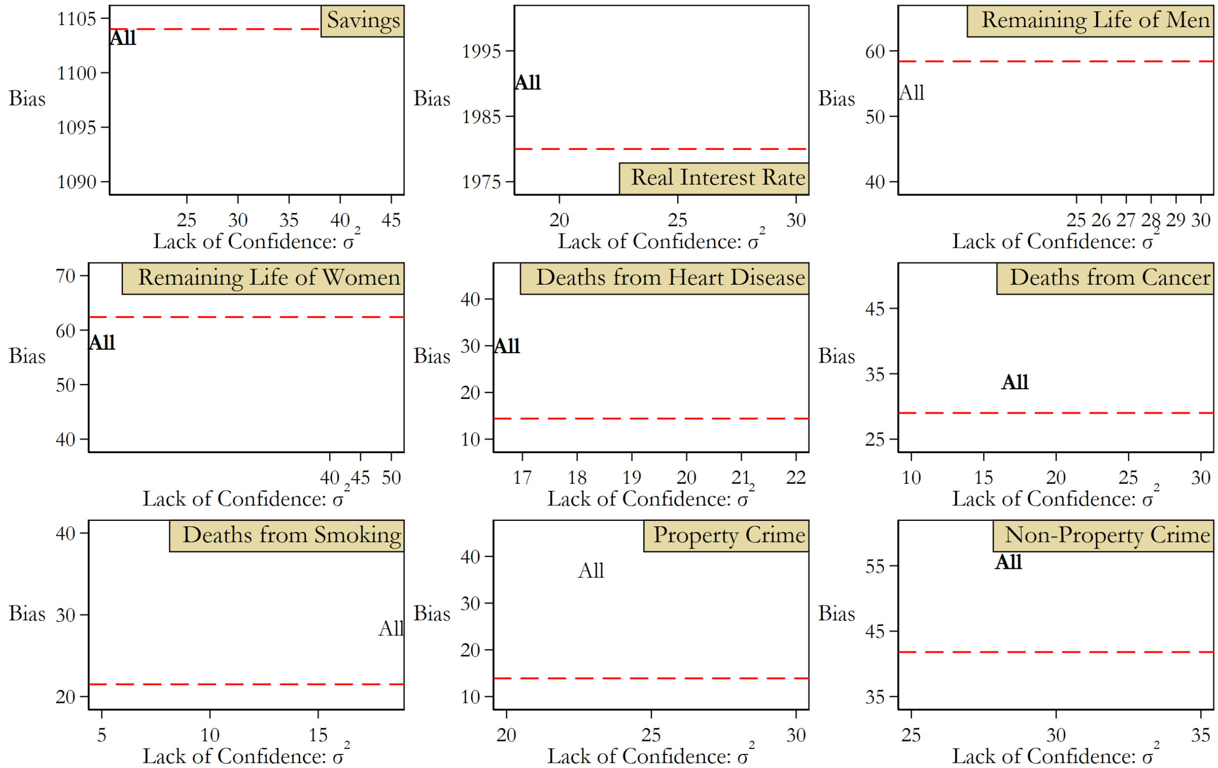


Figure 5.22: Comparative Bias and Confidence

Bold indicates statistically significant bias *and* confidence at the 5% significance level

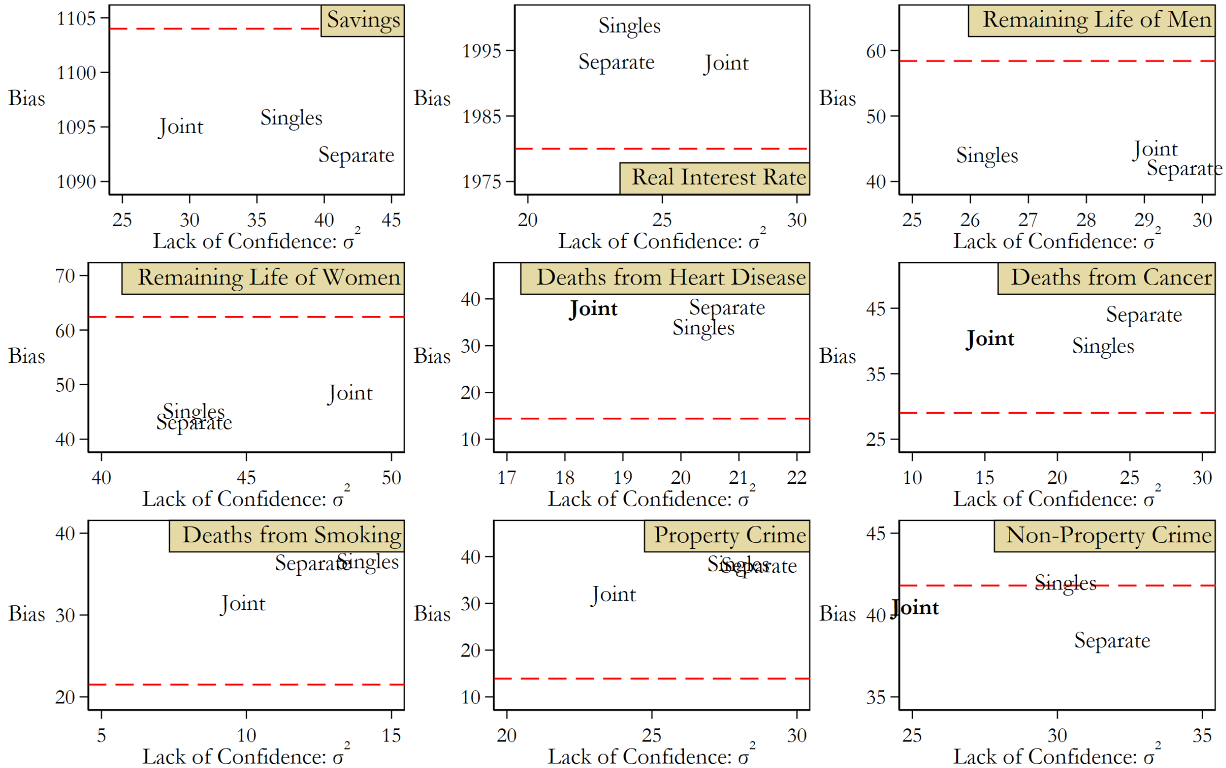


Table 5.28: Pooled Measures of *L* and *W* Indices, Single Individual and Households Making Joint or Separated Decisions

Question	Type	Correct Answer	Single Literacy Measures		Joint Literacy Measures		Separate Literacy Measures	
			L	*W*	*L*	*W*	*L*	*W*
Q1 - Savings Account 2%	Numeracy	1,104 kroner	0.51	0.43	0.62	0.57	0.49	0.42
Q2 - Real Interest Rate	Numeracy	1,980 kroner	0.30	0.21	0.46	0.37	0.35	0.27
Q3 - Remaining Life for Men	Longevity Risk	58.4 years	0.41	0.33	0.46	0.37	0.45	0.37
Q4 - Remaining Life for Women	Longevity Risk	62.4 years	0.45	0.37	0.51	0.43	0.56	0.49
Q5 - Deaths from Heart Disease	Health	14.40%	0.21	0.13	0.18	0.10	0.18	0.11
Q6 - Deaths from Cancer	Health	29.0%.	0.21	0.13	0.32	0.22	0.21	0.13
Q7 - Deaths from Smoking	Health	21.50%	0.21	0.13	0.23	0.15	0.25	0.16
Q8 - Property Criminal Activity	Crime	14%	0.13	0.07	0.18	0.11	0.18	0.11
Q9 - Non-Property Criminal Activity	Crime	41.80%	0.08	0.04	0.10	0.06	0.06	0.03

*Pooled literacy measures of *L* and *W* are initally reported for each question in their respective table in Chapter 5. "Single" (not married, living alone) control subjects. "Joint" (married or registered partnership) making decisions together jointly. "Separate" (married or registered partnership) making decisions separately from one another

Table 5.29: Pooled Measures of *L* Index, Comparing the Single Individual to Households Making Joint or Separated Decisions Measures

Question	Type	Correct Answer	Single Literacy Measures		Joint Literacy Measures		Separate Literacy Measures	
			L		*L*	+/-	*L*	+/-
Q1 - Savings Account 2%	Numeracy	1,104 kroner	0.51		0.62	0.11	0.49	-0.02
Q2 - Real Interest Rate	Numeracy	1,980 kroner	0.30		0.46	0.16	0.35	0.05
Q3 - Remaining Life for Men	Longevity Risk	58.4 years	0.41		0.46	0.05	0.45	0.04
Q4 - Remaining Life for Women	Longevity Risk	62.4 years	0.45		0.51	0.06	0.56	0.11
Q5 - Deaths from Heart Disease	Health	14.40%	0.21		0.18	-0.03	0.18	-0.03
Q6 - Deaths from Cancer	Health	29.0%.	0.21		0.32	0.11	0.21	0.00
Q7 - Deaths from Smoking	Health	21.50%	0.21		0.23	0.02	0.25	0.04
Q8 - Property Criminal Activity	Crime	14%	0.13		0.18	0.05	0.18	0.05
Q9 - Non-Property Criminal Activity	Crime	41.80%	0.08		0.10	0.02	0.06	-0.02

*Pooled literacy measures of *L* are initally reported for each question in their respective table. The "+/-" column is the difference of the treatment compared to the single (not married, living alone) "control" subjects. "Joint" (married or registered partnership) making decisions together jointly. "Separate" (married or registered partnership) making decisions separately from one another

Appendix G: Danish Households, Sample Frame for Experiments

1. Allocation of Subjects to Sampling Frames

We distinguish between 3 groups in our laboratory experiment with a total goal of 360 subjects:

Group 1 (G1): Single adults (**120 subjects from Fam1**)
Group 2 (G2): Households - Individual Decisions (**120 subjects from Fam 2 and Fam 3**)
Group 3 (G3): Households - Joint Decisions (**120 subjects from Fam 2 and Fam 3**)

2. Allocation to Recruitment Treatment

The 120 individuals in **G1** are randomly allocated across three recruitment treatments:

- **I_1500**, in which each individual will have a 1 in 4 chance of receiving 1,500 kroner
- **I_2000**, in which each individual will have a 1 in 4 chance of receiving 2,000 kroner
- **I_2500**, in which each individual will have a 1 in 4 chance of receiving 2,500 kroner

The 240 individuals (120 households) in **G2** and **G3** are randomly allocated across three recruitment treatments:

- **H_1500**, in which each individual in the household will have a 1 in 4 chance of receiving 1,500 kroner
- **H_2000**, in which each individual in the household will have a 1 in 4 chance of receiving 2,000 kroner
- **H_2500**, in which each individual in the household will have a 1 in 4 chance of receiving 2,500 kroner

Hence, we operate with one set of recruitment treatments for individuals in **G1** and another set of treatments for households (**G2** and **G3**), but where each individual is randomly allocated across the same three treatments.

We make a distinction between sessions in which subjects, including household couples, make individual decisions (**G1** and **G2**) and sessions in which couples in a households make joint decisions (**G3**). For logistical reasons we separate sessions with individual and joint household decision tasks.

3. Summary of the Design: Target Number of Subjects per Group and Treatment

26 sessions were run over the course of 13 days: 20 sessions with individual decisions and 5 sessions with joint household decisions. The planned sessions by day consisted of the following:

Day 1
The sample in each session is composed of 12 subjects from **G1**. The subjects are paid the following recruitment fees:
Day 1: Each individual gets 1 in 4 chance of receiving 1,500 kr. **(I_1500)**

Day 2
The sample in each session is composed of 12 subjects (6 households) from **G2**. The subjects are paid the following recruitment fees:
Day 2: Each individual gets 1 in 4 chance of receiving 2,500 kr. (**H_2500** & **I_2500**)

Day 3 - Day 6
The sample in each session is composed of 12 subjects from **G1**. The subjects are paid the following recruitment fees:

Day 3:	Each individual gets 1 in 4 chance of receiving 1,500 kr. (**I_1500**)
Day 4:	Each individual gets 1 in 4 chance of receiving 2,000 kr. (**I_2000**)
Day 5:	Each individual gets 1 in 4 chance of receiving 2,500 kr. (**I_2500**)
Day 6:	Each individual gets 1 in 4 chance of receiving 1,500 kr. (**I_1500**)

Day 7 - Day 10
The sample in each session is composed of 12 subjects (6 households) from **G2**. The subjects are paid the following recruitment fees:

Day 7:	Each individual gets 1 in 4 chance of receiving 1,500 kr. (**H_1500** & **I_1500**)
Day 8:	Each individual gets 1 in 4 chance of receiving 2,000 kr. (**H_2000** & **I_2000**)
Day 9:	Each individual gets 1 in 4 chance of receiving 2,500 kr. (**H_2500** & **I_2500**)
Day 10:	Each individual gets 1 in 4 chance of receiving 1,500 kr. (**H_1500** & **I_1500**)

Day 11 - Day 13
The sample in each session is composed of 20 subjects (10 households) from **G3**. The subjects are paid the following recruitment fees:

Day 11:	Each individual gets 1 in 4 chance of receiving 1,500 kr. (**H_1500**)
Day 12:	Each individual gets 1 in 4 chance of receiving 2,000 kr. (**H_2000**)
Day 13:	Each individual gets 1 in 4 chance of receiving 2,500 kr. (**H_2500**)

Table G1 quickly summarizes the design envisioned by group and treatments.

Table G1 - Summary of the Design: Target Number of Subjects per Group and Treatment

TREATMENT	Number of Subjects by Group		
	G1	G2	G3
I_1500	72	0	0
I_2000	24	0	0
I_2500	24	0	0
H_1500	0	48	40
H_2000	0	24	40
H_2500	0	48	40
Total	120	120	120

4. Recruitment, Acceptance and Show Up Rates

Statistics Denmark mailed several thousand letters to potential participants inviting them to be in the study. Table G2 shows how many letters were sent out, and the acceptance and show up rate by group and treatment.

Table G2 - Summary of Recruitment, Acceptance and Show Up Rates

TREATMENT	Number of Invitations by Group			Number of Accepted by Group			Number Showed Up by Group		
	G1	G2	G3	G1	G2	G3	G1	G2	G3
I_1500	1,269	0	0	44	0	0	35	0	0
I_2000	520	0	0	31	0	0	25	0	0
I_2500	660	0	0	51	0	0	42	0	0
H_1500	0	1,600	1,800	0	40	52	0	38	41
H_2000	0	1,600	840	0	38	30	0	37	19
H_2500	0	908	1,372	0	38	66	0	31	62
Total	2,449	4,108	4,012	126	116	148	102	106	126

Appendix H: Danish Household, Experimental Parameters

There are five decision tasks in the experimental design that are presented in the following order to all individual subjects and households: (1) atemporal risk aversion choice tasks, (2) time preference choice tasks, (3) intertemporal risk aversion choice tasks, (4) subjective belief questions, and (5) survey questions. We focus here solely on the atemporal risk aversion task. The subjective belief elicitation task is documented in the main text.

We have a battery of 60 lottery pairs, presented to subjects in random order.

Wakker, Erev, and Weber (1994) constructed lotteries to carefully test the "comonotonic independence" axiom of RDU. Their main lottery pairs consist of 6 sets of 4 pairs. The logic of their design can be seen by considering the first set, from Wakker, Erev, and Weber (1994, Fig. 3.1). The second and third prizes in each pair stay the same within the set of 4 choice pairs. The only thing that varies from pair to pair is the monetary value of the first prize, and that is common to the "safe" and "risky" lottery within each pair.[42] Since the first listed prize is a common consequence in both lotteries within a pair, it should not affect choices under EUT. In the 1st pair the first prize is only $0.50, and is the lowest ranked prize for both lotteries. The first prize increases to $3.50 for the 2nd pair, and is again the lowest ranked prize for both lotteries: so rank-dependence should have no effect on choice patterns as the subject moves from the 1st to the 2nd pair. But when we come to the 3rd pair the first prize is $6.50, which makes it the second highest ranked prize for both lotteries; this is where RDU *could* have a different prediction than EUT, depending on the extent and nature of probability weighting. Finally, in the 4th pair the common consequence is the highest ranked prize for both lotteries, again *allowing* RDU to predict something different from EUT (and from the choices in the 3rd pair). Note that this design does not formally *require* an RDU decision-maker to choose differently than an EUT decision-maker; it simply encourages it for a *priori* reasonable levels of probability weighting. We employ all 24 of their main lottery pairs, and scale the prizes considerably.

Loomes and Sugden (1998) pose an important design feature for common ratio tests: variation in the "gradient" of the EUT-consistent indifference curves within a Marschak-Machina (MM) triangle. The reason for this is to generate some choice patterns that are more powerful tests of EUT for any given risk attitude. That is, under EUT the slope of the indifference curve within a MM triangle is a measure of risk aversion. So there always exists some risk attitude such that the subject is indifferent, and evidence of common ratio violations has virtually zero power (EUT does not, then, predict 50:50 choices, as some casually claim). We use 30 lottery pairs from their design.

A final battery of 6 lottery pairs is designed to test the premise of the calibration puzzle posed by Hansson (1988) and Rabin (2000): that subjects exhibit small-stakes risk aversion for "all wealth" (or for a large enough finite range of wealth levels). Cox and Sadiraj (2008, pg. 33) proposed a very simple test for this idea. You give people choices between safe and risky lotteries, where the safe lotteries are certain amounts of money, and the risky lotteries are a 50:50 chance of +x/-y either side of the certain amount of money in the safe lottery. Hold x and y constant for choices that vary the safe prize level, and let x>y so that the Expected Value of this risky lottery is slightly above the safe lottery level. The idea here is to see the safe lottery as "lab wealth," and then see if subjects are risk averse as we vary lab wealth. For instance, one might have +x/-y as +$15/-$10, then consider one binary choice in which the safe lottery is $20 and one binary choice in which the safe lottery is $100. So the subject would make two choices: take $20 for certain, or take a 50:50 chance of $10 or $35; and take $100 for certain, or take a 50:50 chance of $90 or $115. The

[42] What is "safe" and what is "risky" is not so obvious when one allows for probability weighting, but this is how the lotteries are labeled.

Hansson-Rabin premise is that one gets risk aversion in both cases, with a majority of people picking the safe lottery.[43]

The experimental design includes three treatments for individuals and two treatments for households, where the treatments vary the scale of prizes in lottery tasks. The three treatments for individuals are designed to give a risk neutral individual approximately 850 kroner in the low income treatment, 1,225 kroner in the medium income treatment and 1,700 kroner in the high income treatment. The two treatments for households are designed to give a risk neutral household approximately 1,700 kroner in the low income treatment and 2,550 kroner in the high income treatment. These treatments are randomly allocated between individuals and households. Each individual and household is given a 10% chance of being paid in this part of the experiment.

[43] Wilcox (2013) independently implemented the same idea with Chapman University students, and found striking evidence against the calibration premise. We have replicated his design with GSU students, and obtained the same findings, reported in Harrison, Lau, Ross, and Swarthout (2017). Cox, Sadiraj, Vogt and Uteeyo (2013) implemented this design in two experiments in Calcutta. A third experiment, involving a casino in Europe, entailed some experimental procedures that are non-standard. The raw data for the two Calcutta experiments had sample sizes of 30 and 40, respectively. In one case they observed clear evidence of risk neutrality or risk-loving behavior; this is their +30/-20 case. In the other case they observed less clear evidence that depends on how one interprets indifference choices; this is their +90/-50 case. They interpret their data as saying that 43% to 48% of subjects in the +30/20 case satisfy the calibration premise, and that 81% or so satisfy it in the +90/-50 case. We disagree with these interpretations, for reasons that are not critical here.

Chapter 6 – Summary and Extensions for Future Research

6.1 Summary

The concept of literacy has grown from "reading literacy" to now encompass many different domain-specific topics and skill sets, such as health literacy, financial literacy, and computer literacy. The way concepts of literacy are discussed, examined, measured, and communicated has also evolved. Reading literacy measures began as a simple metric of counting the number of individuals in a country that could read certain text, and dividing that count by the total population to compute the percentage of literate individuals. However, this approach ignores situations in which an illiterate person has access to a literate person that could read to them. This was the premise of research in development economics that introduced the measure of effective literacy, which accounts for potential positive externalities that could arise from an illiterate person having access to a literate person. In the psychology literature this is referred to as a social scaffold.

This dissertation expands on the idea of effective literacy and introduces a concept of extended literacy in Chapter 1, which applies to a decision-maker having access to an external scaffold during the decision-making process. The scaffolds we consider include access to the internet, access to an anonymous person as part of a group, and access to a household member. The research presented here measures extended financial literacy under these various scaffolds and compares those measures to a group of participants that did not have access to a scaffold, thus were responding using only their "private" literacy. The main contribution of this dissertation is that it is the first and only extension to consider the effects of scaffolds as it pertains to financial literacy.

Financial literacy reflects an individual's knowledge about financial matters, including the management of risks. The research presented here assesses subjects' knowledge about interest and inflation, budgeting, and longevity risk. The techniques used to measure literacy move beyond multiple choice questions or "fill in the blank" responses, and reflects state-of-the-art advances in subjective belief elicitation that allow for the recovery of each decision-maker's entire underlying subjective distribution, while controlling for their attitude to risk. This method ascertains how precise an individual's knowledge is in response to a question to a known, true answer. This method allows the construction of literacy and welfare measures in addition to a rich characterization of the belief distribution underlying an individual's elicited response, a test for any bias, and an evaluation of their confidence of response with respect to a known, true, answer.

Table 6.1 displays a summary view of the financial literacy questions and the associated pooled *L* and *W* indices across the results previously shown as standalones in Tables 2.39, 3.34, and 4.23 from Chapters 2, 3 and 4, respectively. The financial literacy questions are further classified as either pertaining to numeracy, procedural, and longevity risk, to allow a more nuanced evaluation of which scaffolds perform better under what circumstances. The numeracy questions ask people to perform various calculations relating to financial concepts that include simple interest, compound interest, the real rate of interest adjusting for inflation, and budgeting. The procedural questions relate to financial literacy with a range of policy implications: the age that one is eligible for Medicare or social security, and consumer protection issues around fraudulent charges on debit and credit cards. The longevity risk questions are important for retirement planning purposes.

Table 6.2 is like Table 6.1, but it focuses only on the *L* index and the changes to it when compared to the Individual literacy measures. Table 6.2 is an aggregation of Tables 3.35 and 4.24. In the fourth column of Table 6.2 the pooled measures of the *L* index are for individuals responding with only their private literacy. In the sixth and eighth columns are the *L* index measures for the Internet and the Group treatments, respectively. Column seven (nine) is the "+/-" difference between *L* index for those with Internet (Group) access compared to individuals without. We see "+/-" values for the Internet treatment that range from -0.02 for fin14, "the interest rate" question, all the way to +0.66 for fin3, "the Medicare eligibility" question. And for the Group treatment we see values ranging from -0.05 for fin2 to +0.17 for fin7. It appears that the Internet scaffold enhances literacy the most for procedural questions, while the Group scaffolds enhances literacy the most for numeracy questions. Neither scaffold appears to have diminished literacy in an economically appreciable manner.

Tables 6.3 through 6.13 display a summary view of how measures of literacy are enhanced or diminished with respect to various demographics and treatments. While the *L* index accounted for pooled results only, the thesis also examined bias from the true, objective response and additional imprecision of beliefs held by different demographics. These results were displayed in the Tables in Chapters 2, 3, and 4, and are collated here. Take for example the estimates shown in Table 6.3 for fin1 the "savings account at 2%" question. We know that the correct answer is $110.41, and would like to know if there are any particular demographics that are biased from the correct answer. To do that, in the top panel we estimate the "total effect" of the covariate, and then compare the estimated average belief of the covariate to the correct answer. Consider the covariate female: we see that the estimated average of being female is an upward bias of $1.33 over the correct answer of $110.41, and that this result is statistically significant at the 5% level. We can read down this column and also see statistical evidence of bias from the correct answer for Asians, Black, Christians and Seniors, and no evidence of bias for those with a high GPA or juniors.

The bottom panel of Table 6.3 displays the additional imprecision of beliefs, where we compare the estimated standard deviation for each covariate to the overall pooled standard deviation. The pooled imprecision over all individuals is 3.38. The estimated standard deviation for the covariate female is statistically significant different at the 5% level, and is 1.24 less than the overall. Thus females exhibit a greater confidence in beliefs, but that increased confidence may be costly because they are also biased from the true answer.

Moving over to the Internet scaffold in Table 6.3, we interpret these results similarly. Here we note that evidence of statistical bias from the correct answer is beginning to disappear for several demographics: Asian, Black, Christian, and Seniors, for instance. It is still there for female, although the absolute bias from the true answer has decreased and even flipped signs. The additional imprecision of beliefs is 3.18 for the pooled Internet treatment, hence it seems that access to the Internet as a scaffold increases confidence. Finally, in the Group scaffold section of Table 6.3 we notice for each demographic that all evidence of statistical bias from the correct answer has disappeared. For this numeracy type question, being in a group greatly enhances literacy and reduces imprecision.

Tables 6.4 through 6.13 present the results for fin2 through fin16 in the same manner. Using controlled laboratory and artefactual field experiments with real rewards and incentivized elicitation of beliefs, we find that the scaffolds investigated in this research *reliably enhance literacy as compared with "private literacy" measures of participants without access to a scaffold.*

6.2 Extensions

There are many possible extensions of the research presented in this thesis.

One extension would be to calibrate differences between the elicited responses about beliefs from incentivized experiments with survey responses to hypothetical questions. One reason to do this would be to judge if the latter are reliable indicators of the former, although that is not the most important reason for making this comparison *a priori*, given the wealth of evidence of hypothetical bias across a wide range of elicitation tasks.[44] Rather, the challenge is to design complementary surveys and experiments that can be used together, with the latter being used to calibrate the former to correct for hypothetical bias. It is typically not feasible, logistically or financially, to undertake experiments with thousands of respondents, although it is feasible to conduct surveys with larger samples.[45] The idea is to take a sub-sample and conduct experiments in which one considers hypothetical bias and the connection to observable demographics, and then use those data to statistically calibrate the respondents that only did the survey (e.g., Blackburn,

[44] There is a vast literature on this topic. For instance, see Neill et al. (1994); Cummings, Harrison, and Rutström (1995); Cummings et al. (1997); Harrison et al. (1999); Nape et al. (2003); Harrison (2006a, 2006b, 2006c, 2014b).
[45] There exists a scalable piece of software that is already built and lives on an internet server that can implement this logistically. As long as there is an internet connection available and a computer running a modern internet browser, someone could participate in an elicitation anywhere across the globe.

Harrison, and Rutström (1994), Harrison, Beekman, Brown, Clements, McDaniel, Odom and Williams (1999), Harrison (2006a, 2006b)).

Another extension would be to examine the interaction between the cognitive demand of the elicitation task and the literacy measures observed by participants with varying levels of cognitive skill. In Chapters 2, 3, and 4, for example, the elicited responses were obtained from university students at Georgia State, who have passed high school and are enrolled at an institution of higher education. Thus, one would presume these are individuals with higher cognitive skills. While a cognitive skills test was beyond the scope of this research, it would be interesting to capture that data and analyze if students with lower cognitive skills performed differently than those students with higher cognitive skills. After controlling for cognitive ability, one could also compare the responses obtained by the subjective belief elicitation framework to other methods (multiple choice, fill in the blank, yes/no, etc.) that are also incentivized.

Another extension would be to consider households making joint decisions, and augment those experiments with additional tasks that allow one to identify the "bargaining power" of each adult member of the household. Recognizing that this bargaining power might be domain-specific, in the sense that one household member might be more influential when it comes to financial decisions and the other household member might be more influential when it comes to lifestyle decisions, one can then better explain how joint decisions are arrived at. This information might also help identify more efficient interventions to improve household literacy. This extension would be particularly important for applications in developing countries, where the bargaining power of women is often sharply diminished by cultural and historical circumstances (e.g., Summers (1994)).

Another extension would be to use the insights gained in the experimental treatments of chapters 3, 4 and 5 to design "behaviorally smart" interventions to improve literacy. We know from these results that individuals with certain demographics have a *statistical* tendency to be less literate than others: can we then efficiently target individuals with those demographics with interventions to improve their literacy, avoiding the time and pecuniary cost of interventions for others? We could also test an individual's awareness of bias, or lack thereof, to speak to the level of naivety or sophistication they have. More generally, can we examine *which interventions* (e.g., access to internet or access to someone in a group) might be more effective for certain demographic sub-samples and levels of sophistication?

Another extension would be to examine literacy over the life-cycle. There is evidence that individuals have relatively well calibrated beliefs about mortality risks for people their age and younger, but relatively poorly calibrated beliefs about mortality risks for people much older than them (Harrison and Rutström 2006). This is not surprising, given the role that peer experience and own experience plays in forming risk perceptions. But it points to a dimension in which literacy interventions might focus: events of significance in the future. The debate over risk perceptions about the health consequences of smoking is an important example. Most people start smoking when they are around 13 or 14, but bear the morbidity and mortality consequences many decades later. How might one improve literacy at the age when decisions of this kind are made, particularly when they lead to decades of dependence (e.g., (Harrison (2017))?

Finally, it would be valuable to extend the approach adopted here to health literacy, which has a large literature of its own (e.g., Baker (2006); Benjamin (2010); Huber, Shapiro II, and Gillaspy (2012)). Clearly some of the "financial literacy" questions we considered, such as longevity risk, overlap with "health literacy." But a more detailed investigation of health literacy with the same methods would be warranted. One significant extension of the methodology used here would be to consider the elicitation of bivariate risks at the same time: what is the subjective covariance (or correlation) between financial risks and health risks? Or between mortality and morbidity risks, since things that do not kill someone often leave them needing medical care.

With the tools and the methods of subjective belief elicitation developed and firmly in place to rigorously measure literacy there are many interesting questions to be researched in various domains.

Table 6.1: Pooled Measures of L and W Indices by Treatment

Question	Type	Correct Answer	Individual Literacy Measures		Internet Literacy Measures		Group Literacy Measures	
			L	W	L	W	L	W
fin1 - Savings Account 2%	Numeracy	$110.41	0.53	0.50	0.68	0.67	0.67	0.62
fin2 - Social Security Start Age	Procedural	62	0.26	0.21	0.87	0.84	0.21	0.18
fin5 - Medicare Eligibility	Procedural	65	0.16	0.16	0.82	0.80	0.31	0.37
fin7 - Real Interest Rate	Numeracy	$98.98	0.28	0.29	0.28	0.28	0.45	0.38
fin9 - Savings Horizon	Numeracy	4 months	0.77	0.74	0.84	0.83	0.92	0.93
fin10 - Stolen Credit Card	Procedural	$50	0.06	0.06	0.66	0.64	0.05	0.07
fin11 - Stolen Debit Card	Procedural	$500	0.16	0.12	0.61	0.59	0.21	0.15
fin13 - Nominal Interest	Numeracy	5%	0.80	0.76	0.81	0.83	0.93	0.92
fin14 - Interest Rate	Numeracy	$102	0.74	0.68	0.72	0.71	0.90	0.86
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22	0.18	0.48	0.42	0.31	0.22
fin16 - Remaining Life for Wome	Longevity Risk	61.7 years	0.27	0.22	0.51	0.53	0.33	0.27

*Pooled literacy measures of L and W are initally reported for each question in their respective section. Chapter 2 for indvidual measures, Chapter 3 for Internet, and Chapter 4 for Group measures.

Table 6.2: Pooled Measures of L Index, Comparison to the Individual Measures by Treatment

Question	Type	Correct Answer	Individual Literacy Measures		Internet Literacy Measures		Group Literacy Measures	
			L		L	+/-	L	+/-
fin1 - Savings Account 2%	Numeracy	$110.41	0.53		0.68	0.15	0.67	0.14
fin2 - Social Security Start Age	Procedural	62	0.26		0.87	0.61	0.21	-0.05
fin5 - Medicare Eligibility	Procedural	65	0.16		0.82	0.66	0.31	0.15
fin7 - Real Interest Rate	Numeracy	$98.98	0.28		0.28	0.00	0.45	0.17
fin9 - Savings Horizon	Numeracy	4 months	0.77		0.84	0.07	0.92	0.15
fin10 - Stolen Credit Card	Procedural	$50	0.06		0.66	0.60	0.05	-0.01
fin11 - Stolen Debit Card	Procedural	$500	0.16		0.61	0.45	0.21	0.05
fin13 - Nominal Interest	Numeracy	5%	0.80		0.81	0.01	0.93	0.13
fin14 - Interest Rate	Numeracy	$102	0.74		0.72	-0.02	0.90	0.16
fin15 - Remaining Life for Men	Longevity Risk	57.1 years	0.22		0.48	0.26	0.31	0.09
fin16 - Remaining Life for Wome	Longevity Risk	61.7 years	0.27		0.51	0.24	0.33	0.06

*Pooled literacy measures of L are initally reported for each question in their respective section. Chapter 2 for indvidual measures, Chapter 3 for Internet, and Chapter 4 for Group measures. The "+/-" column is the difference of the treatment compared to the Individual "control" subjects.

Table 6.3: Summary of Results for the "Savings Account 2%" Question (fin1)

Bias is realative to the correct answer of $110.41
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-1.07 **	
InternetC		-0.43	
Group			0.12
female	1.33 **	-0.40 *	0.16
asian	2.02 ***	-0.15	0.52
black	1.30 **	-0.23	0.22
christian	1.34 ***	-0.19	0.27
gpaHI	-0.10	-0.39 **	-0.22
junior	0.57	-0.35	-0.11
senior	1.08 **	-0.66	0.07
Additional Imprecision of Beliefs			
Impercision	**3.38**	**3.18**	**2.95**
InternetNC		-0.69	
InternetC		-0.13	
Group			-0.94 ***
female	-1.24 **	0.10	0.06
asian	-0.33	-0.06	-0.40
black	0.34	0.28	0.25
christian	-0.62	0.23	0.21
gpaHI	-1.45 ***	-0.59 **	-0.34
junior	-1.86 ***	-0.98 ***	-0.62 **
senior	-1.49 ***	0.17	-0.34

NB: Estimates taken from thesis Tables: 2.18 (Individuals), 3.3 (Internet), and 4.2 (Group)

Table 6.4: Summary of Results for the "Social Security Start Age" Question (fin2)

Bias is realative to the correct answer of 62 years old
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		0.30 *	
InternetC		0.63 ***	
Group			3.22 ***
female	4.67 ***	1.99 ***	3.63 ***
asian	2.67 ***	1.79 ***	3.44 ***
black	3.03 ***	2.22 ***	3.50 ***
christian	4.33 ***	2.24 ***	3.54 ***
gpaHI	4.55 ***	2.02 ***	3.72 ***
junior	4.04 ***	1.58 ***	3.22 ***
senior	3.38 ***	1.75 ***	2.98 ***
Additional Imprecision of Beliefs			
Impercision	**3.58**	**3.17**	**3.24**
InternetNC		-1.51 **	
InternetC		-1.43 ***	
Group			-0.75 ***
female	-0.31	-0.01	0.00
asian	-0.89	-0.11	-0.10
black	0.35	0.12	0.03
christian	-0.14	0.08	-0.01
gpaHI	1.06	0.11	0.25
junior	-1.21 ***	-0.46	-0.59 **
senior	-0.95 **	-0.48 **	-0.52 *

NB: Estimates taken from thesis Tables: 2.20 (Individuals), 3.6 (Internet), and 4.4 (Group)

Table 6.5: Summary of Results for the "Medicare Eligibility" Question (fin5)

Bias is realative to the correct answer of 65 years old
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-1.47 **	
InternetC		-0.71 ***	
Group			-5.96 ***
female	-4.05 **	-3.78 ***	-6.97 ***
asian	-6.05 ***	-3.98 ***	-6.60 ***
black	-7.19 ***	-4.61 ***	-7.36 ***
christian	-6.42 ***	-4.48 ***	-7.22 ***
gpaHI	-6.05 ***	-3.98 ***	-6.94 ***
junior	-6.23 ***	-3.93 ***	-7.34 ***
senior	-4.25 ***	-3.10 ***	-5.86 ***
Additional Imprecision of Beliefs			
Impercision	**5.30**	**5.38**	**5.30**
InternetNC		-1.39	
InternetC		-2.98 ***	
Group			-0.18
female	-0.81	-0.16	-0.12
asian	-0.51	0.04	-0.11
black	-0.13	0.24	0.12
christian	-0.69	0.08	0.04
gpaHI	-0.93	-0.04	-0.04
junior	-0.26	0.39	-0.22
senior	0.07	-0.60	-0.06

NB: Estimates taken from thesis Tables: 2.22 (Individuals), 3.9 (Internet), and 4.6 (Group)

Table 6.6: Summary of Results for the "Real Interest Rate" Question (fin7)

Bias is realative to the correct answer of $98.98
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		1.59 ***	
InternetC		1.64 ***	
Group			1.00 ***
female	2.28 ***	2.08 ***	1.80 ***
asian	1.86 ***	1.55 ***	1.78 ***
black	1.50 ***	2.10 ***	1.63 ***
christian	2.31 ***	2.14 ***	1.81 ***
gpaHI	0.88	1.55 ***	1.44 ***
junior	0.64	1.61 ***	1.52 ***
senior	1.45	1.47 ***	1.56 ***
Additional Imprecision of Beliefs			
Impercision	**2.24**	**2.12**	**2.19**
InternetNC		-0.14	
InternetC		-0.15	
Group			-0.26 *
female	0.36	-0.01	0.00
asian	0.20	-0.10	-0.13
black	0.37	0.05	0.02
christian	0.24	0.04	0.03
gpaHI	-0.17	-0.17 *	-0.14
junior	-0.54 *	-0.16	-0.28 *
senior	0.23	0.04	-0.02

NB: Estimates taken from thesis Tables: 2.24 (Individuals), 3.12 (Internet), and 4.8 (Group)

Table 6.7: Summary of Results for the "Savings Horizon" Question (fin9)

Bias is realative to the correct answer of 4 months
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		0.28 **	
InternetC		0.47 ***	
Group			0.33 *
female	-0.15	0.63 ***	0.69 ***
asian	0.06	0.52 **	0.42
black	0.31	0.86 ***	0.92 ***
christian	0.35 **	0.87 ***	0.87 ***
gpaHI	-0.24	0.66 ***	0.70 ***
junior	0.29	0.96 ***	0.89 **
senior	0.09	0.64 ***	0.72 **
Additional Imprecision of Beliefs			
Impercision	**2.25**	**1.82**	**2.00**
InternetNC		-0.91 ***	
InternetC		-0.51 **	
Group			-0.60
female	-2.07 ***	-0.11	-0.07
asian	-1.80 ***	-0.18	-0.41
black	-1.46 ***	0.21	0.26
christian	-1.69 ***	0.21	0.16
gpaHI	-2.08 ***	-0.03	-0.02
junior	-1.56 ***	0.54	0.47
senior	-1.88 ***	-0.06	0.02

NB: Estimates taken from thesis Tables: 2.26 (Individuals), 3.15 (Internet), and 4.10 (Group)

Table 6.8: Summary of Results for the "Stolen Credit Card" Question (fin10)

Bias is realative to the correct answer of $50
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		41.20 **	
InternetC		25.74 **	
Group			82.84 ***
female	49.49	76.52 ***	118.03 ***
asian	48.64	46.88 *	69.63 *
black	94.04	89.27 ***	127.33 ***
christian	121.69 **	92.77 ***	138.05 ***
gpaHI	50.58	82.14 ***	123.67 ***
junior	-32.14	30.92 *	75.22 **
senior	-11.28	53.71 **	63.07 *
Additional Imprecision of Beliefs			
Impercision	**283.08**	**218.67**	**274.87**
InternetNC		-64.55	
InternetC		-91.62 ***	
Group			-17.62
female	-88.13	9.34	12.85
asian	-94.16	-39.01	-43.83
black	-45.69	24.08	17.27
christian	-68.02	21.87	24.13
gpaHI	-113.26	11.26	20.94
junior	-135.51 *	-62.84 *	-25.73
senior	-118.13	-17.79	-39.43

NB: Estimates taken from thesis Tables: 2.28 (Individuals), 3.18 (Internet), and 4.12 (Group)

Table 6.9: Summary of Results for the "Stolen Debit Card" Question (fin11)

Bias is realative to the correct answer of $500
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-22.45	
InternetC		-30.57	
Group			-5.95
female	-262.98	177.29 **	252.70 ***
asian	-48.68	-8.44	20.21
black	274.27 *	248.62 ***	321.56 ***
christian	334.97 *	282.50 ***	371.02 ***
gpaHI	-53.61	224.01 *	349.18 **
junior	-63.79	31.71	68.56
senior	413.88	393.23 *	586.43
Additional Imprecision of Beliefs			
Impercision	**1495.55**	**1073.09**	**1241.33**
InternetNC		-667.21 ***	
InternetC		-653.20 ***	
Group			-715.55 ***
female	-1075.55 ***	-32.15	-71.33
asian	-1120.63 ***	-565.52 ***	-660.74 ***
black	-823.68 ***	42.15	-22.14
christian	-349.52	180.75	158.59
gpaHI	-821.32 ***	147.12	170.35
junior	-906.41 ***	-395.22 **	-463.01 **
senior	-42.25	639.51	880.90

NB: Estimates taken from thesis Tables: 2.30 (Individuals), 3.21 (Internet), and 4.14 (Group)

Table 6.10: Summary of Results for the "Nominal Interest" Question (fin13)

Bias is realative to the correct answer of 5%
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-0.57 ***	
InternetC		-0.56 ***	
Group			-0.19 **
female	-0.66 **	-0.58 ***	-0.47 ***
asian	-0.59 *	-0.46 ***	-0.36 ***
black	-0.43 *	-0.63 ***	-0.47 ***
christian	-0.68 ***	-0.59 ***	-0.44 ***
gpaHI	-0.25	-0.34 ***	-0.33 ***
junior	-0.03	-0.60 ***	-0.14 ***
senior	-0.11	-0.38 ***	-0.25 ***
Additional Imprecision of Beliefs			
Impercision	**1.24**	**1.30**	**1.09**
InternetNC		0.06	
InternetC		0.06	
Group			-0.36 *
female	-0.06	-0.08	0.02
asian	0.17	-0.11	-0.14
black	-0.22	0.07	0.07
christian	0.38	0.02	0.01
gpaHI	-0.56 *	-0.41 ***	-0.19
junior	-0.75 ***	0.25	-0.52 ***
senior	-0.42	-0.25	-0.13

NB: Estimates taken from thesis Tables: 2.32 (Individuals), 3.24 (Internet), and 4.16 (Group)

Table 6.11: Summary of Results for the "Interest Rate" Question (fin14)

Bias is realative to the correct answer of $102
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-0.39 *	
InternetC		-0.31 **	
Group			-0.14 ***
female	-0.06	-0.25 ***	-0.22 ***
asian	-0.42 *	-0.26 *	-0.31 *
black	-0.11	-0.33 ***	-0.20 ***
christian	-0.31 **	-0.35 ***	-0.26 ***
gpaHI	0.04	-0.14 *	-0.18 **
junior	0.02	-0.25	-0.23 *
senior	0.10	-0.13	-0.07
Additional Imprecision of Beliefs			
Impercision	**1.25**	**1.26**	**1.06**
InternetNC		0.03	
InternetC		-0.01	
Group			-0.42 ***
female	-0.52 **	-0.03	0.06
asian	-0.06	-0.11	0.07
black	-0.67 ***	0.09	0.00
christian	-0.02	0.14	0.10
gpaHI	-0.80 ***	-0.22 *	-0.04
junior	-0.92 ***	-0.05	-0.19
senior	-0.83 ***	-0.22	-0.24

NB: Estimates taken from thesis Tables: 2.34 (Individuals), 3.27 (Internet), and 4.18 (Group)

Table 6.12: Summary of Results for the "Remaining Life for Men" Question (fin15)

Bias is realative to the correct answer of 57.1 years
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-7.81 ***	
InternetC		-6.52 ***	
Group			-7.06 ***
female	-8.19 ***	-8.25 ***	-9.77 ***
asian	-7.15 ***	-8.42 ***	-7.87 ***
black	-7.56 ***	-9.19 ***	-9.31 ***
christian	-5.49 ***	-8.43 ***	-8.84 ***
gpaHI	-1.81	-7.22 ***	-8.02 ***
junior	-3.24	-7.49 ***	-7.64 ***
senior	-6.18 **	-7.70 ***	-10.02 ***
Additional Imprecision of Beliefs			
Impercision	**9.70**	**9.17**	**9.08**
InternetNC		-1.18 *	
InternetC		-0.52	
Group			-1.46 **
female	-1.78	0.00	0.41
asian	-1.68	0.28	-0.48
black	-2.17 *	0.52	0.38
christian	-2.60 **	0.08	-0.03
gpaHI	-5.34 ***	-0.97 *	-1.10 *
junior	-3.08 *	-0.39	-1.34
senior	-2.03	0.16	1.25

NB: Estimates taken from thesis Tables: 2.36 (Individuals), 3.30 (Internet), and 4.20 (Group)

Table 6.13: Summary of Results for the "Remaining Life for Women" Question (fin16)

Bias is realative to the correct answer of 61.7 years
Additional imprecision is relative to the average imprecision

	Individual	Internet	Group
Bias from True Response			
InternetNC		-8.97 ***	
InternetC		-7.84 ***	
Group			-8.45 ***
female	-11.55	-9.94 ***	-11.51 ***
asian	-12.58	-11.90 ***	-11.26 ***
black	-9.57	-10.23 ***	-10.34 ***
christian	-8.65	-9.81 ***	-10.18 ***
gpaHI	-4.53	-8.64 ***	-9.67 ***
junior	-7.15	-8.39 ***	-8.81 ***
senior	-8.78	-9.10 ***	-11.00 ***
Additional Imprecision of Beliefs			
Impercision	**9.53**	**9.35**	**8.86**
InternetNC		-0.43	
InternetC		-0.46	
Group			-1.70 **
female	-1.86	0.23	0.54
asian	0.68	1.82	1.33
black	-2.59 **	-0.19	-0.23
christian	-2.65 **	-0.04	-0.35
gpaHI	-5.18 ***	-0.46	-0.42
junior	-3.67 ***	-1.05	-2.11 ***
senior	-2.04	0.49	1.55

NB: Estimates taken from thesis Tables: 2.38 (Individuals), 3.33 (Internet), and 4.22 (Group)

References

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

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

———. 2017. "A Method for Evaluating the Quality of Financial Decision Making, with an Application to Financial Education." *Working Paper 20618*. National Bureau of Economic Research.

Andersen, Steffen, John Fountain, Glenn W. Harrison, Arne Risa Hole, and E. Elisabet Rutström. 2012. "Inferring Beliefs as Subjectively Imprecise Probabilities." *Theory and Decision* 73 (1): 161–84.

Andersen, Steffen, John Fountain, Glenn W. Harrison, and E. Elisabet Rutström. 2014. "Estimating Subjective Probabilities." *Journal of Risk and Uncertainty* 48 (3): 207–29.

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

———. 2008. "Eliciting Risk and Time Preferences." *Econometrica* 76 (3): 583–618.

———. 2014. "Discounting Behavior: A Reconsideration." *European Economic Review* 71: 15–33.

———. 2018. "Multiattribute Utility Theory, Intertemporal Utility, and Correlation Aversion." *International Economic Review* 59 (2): 537–55.

ANZ. 2008. "ANZ Survey of Adult Financial Literacy in Australia." https://www.anz.com/Documents/AU/Aboutanz/AN_5654_Adult_Fin_Lit_Report_08_Web_Report_full.pdf.

Atkinson, Adele, Stephen McCay, Elaine Kempson, and Collard Sharon. 2006. "Levels of Financial Capability in the UK: Results of a Baseline Survey." U.K. Financial Services Authority, Consumer Research 47.

Atkinson, Adele, and Flore-Anne Messy. 2012. "Measuring Financial Literacy: Results of the OECD / International Network on Financial Education (INFE) Pilot Study." *OECD Working Papers on Finance, Insurance and Private Pensions*, no. 15: 1,3,6-71.

Baker, David W. 2006. "The Meaning and the Measure of Health Literacy." *Journal of General Internal Medicine* 21 (8): 878–83.

Basu, Kaushik, and James E. Foster. 1998. "On Measuring Literacy." *The Economic Journal* 108 (451): 1733–49.

Bateman, Hazel, Christine Eckert, John Geweke, Jordan Louviere, Susan Thorp, and Stephen Satchell. 2012. "Financial Competence and Expectations Formation: Evidence from Australia." *Economic Record* 88 (280): 39–63.

Beal, Diana J., and Sarath B. Delpachitra. 2003. "Financial Literacy Among Australian University Students." *Economic Papers: A Journal of Applied Economics and Policy* 22 (1): 65–78.

Benjamin, Regina M. 2010. "Improving Health by Improving Health Literacy." *Public Health Reports* 125 (6): 784–85.

Benoît, Jean-Pierre, and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79 (5): 1591–1625.

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

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

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

Blackburn, McKinley, Glenn W. Harrison, and E. Elisabet Rutström. 1994. "Statistical Bias Functions and Informative Hypothetical Surveys." *American Journal of Agricultural Economics* 76 (5): 1084–88.

Bowen, Cathy Faulcon. 2002. "Financial Knowledge of Teens and Their Parents." *Financial Counseling and Planning* 13 (2): 93–102.

Brehm, John, and Scott Gates. 1993. "Donut Shops and Speed Traps: Evaluating Models of Supervision on Police Behavior." *American Journal of Political Science* 37 (2): 555–81.

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

Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge University Press.

Bruner, Jerome. 1968. "Processes of Cognitive Growth: Infancy." *Clark University Press* Heinz Werner Lectures (January). https://commons.clarku.edu/heinz-werner-lectures/20.

Centers for Medicare and Medicaid Services. 2017. "2017 Annual Report of the Boards of Trustees of the Federal Hospital Insurance Trust Fund and the Federal Supplementary Medical Insurance Trust Fund."

Chalmers, David, and Andy Clark. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19.

Clark, Andy. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass: MIT Press.

———. 2011. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford Univ. Press.

Clark, Robert L., Melinda S. Morrill, and Steven G. Allen. 2011. "Pension Plan Distributions: The Importance of Financial Literacy." In *Financial Literacy: Implications for Retirement Security and the Financial Market Place*, 40–58.

Coller, Maribeth, and Melonie B. Williams. 1999. "Eliciting Individual Discount Rates." *Experimental Economics* 2 (2): 107–27.

Condorcet, Jean-Antoine-Nicolas de Caritat. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix ([Reprod.])*. Paris: de l'Impr. royale.

Cook, Douglas O., Robert Kieschnick, and B. D. McCullough. 2008. "Regression Analysis of Proportions in Finance with Self Selection." *Journal of Empirical Finance* 15 (5): 860–67.

Cosmides, Leda, and John Tooby. 1996. "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty." *Cognition* 58 (1): 1–73.

Courchane, Marsha J., and Peter M. Zorn. 2005. "Consumer Literacy and Creditworthiness." In *Federal Reserve System's Community Affairs Research Conference, Promises and Pitfalls: As Consumer Finance Options Multiply, Who Is Being Served and At What Cost?* Proceedings 950: Federal Reserve Bank of Chicago.

Cox, James C., and Vjollca Sadiraj. 2008. "Risky Decisions in the Large and in the Small: Theory and Experiment." In *Risk Aversion in Experiments*, edited by James C. Cox and Glenn W. Harrison, 9–40. Research in Experimental Economics 12. Bingley: Emerald JAI.

Cox, James C., Vjollca Sadiraj, Bodo Vogt, and Utteeyo Dasgupta. 2013. "Is There a Plausible Theory for Decision under Risk? A Dual Calibration Critique." *Economic Theory* 54 (2): 305–33.

Cude, Brenda J, Frances C Lawrence, Angela C Lyons, Kaci Metzger, Emily LeJeune, Loren Marks, and Krisanna Machtmes. 2006. "College Students and Financial Literacy: What They Know and What We Need to Learn." In *Eastern Family Economics and Resource Management Association Conference Proceedings*. Knoxville, Tennessee.

Cummings, Ronald G., Steven Elliott, Glenn W. Harrison, and James Murphy. 1997. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* 105 (3): 609–21.

Cummings, Ronald G., Glenn W. Harrison, and E. Elisabet Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review* 85 (1): 260–66.

Denny, Colleen C., and Christine Grady. 2007. "Clinical Research with Economically Disadvantaged Populations." *Journal of Medical Ethics* 33 (7): 382–85.

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

Dreyfus, Hubert L. 1965. *Alchemy and Artificial Intelligence*. P3244. Santa Monica, CA: RAND Corporation.

———. 1972. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row.
Edmonds, Rachel. n.d. "Fast Facts & Figures About Social Security, 2017." *SSA Publication No.13 17785*, Washington, DC: Social Security Administration.
Elder, Todd E. 2013. "The Predictive Validity of Subjective Mortality Expectations: Evidence from the Health and Retirement Study." *Demography* 50 (2): 569–89.
European Commission | eurostat. 2017. "Europe in Figures - Eurostat Yearbook - Statistics Explained." 2017. http://ec.europa.eu/eurostat/statistics-explained/index.php/Europe_in_figures_-_Eurostat_yearbook.
Fernandes, Daniel, John G. Lynch, and Richard G. Netemeyer. 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors." *Management Science* 60 (8): 1861–83.
Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31 (7): 799–815.
Galton, Francis. 1907. "Vox Populi." *Nature* 75 (1949): 450–51.
Gigerenzer, Gerd, and Ulrich Hoffrage. 1995. "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats." *Psychological Review* 102 (4): 684–704.
Graubard, Stephen Richards, ed. 1988. *The Artificial Intelligence Debate: False Starts, Real Foundations*. Cambridge, Mass.: MIT Press.
Hansson, Bengt. 1988. "Risk Aversion as a Problem of Conjoint Measurement." In *Decision, Probability, and Utility: Selected Readings*, edited by Peter Gärdenfors and Nils-Eric Sahlin. New York: Cambridge University Press.
Harrison, Glenn W. 2014b. "Real Choices and Hypothetical Choices." In *Handbook of Choice Modelling*, in Stephane Hess and Andrew Daly (ed.), 236–54. Cheltenham, UK: Edward Elgar.
———. 2006a. "Hypothetical Bias Over Uncertain Outcomes." In *Using Experimental Methods in Environmental and Resource Economics*, in John A. List (ed.), 41–69. New Horizons in Environmental Economics. Cheltenham, UK; Northampton, MA: Edward Elgar.
———. 2006b. "Making Choice Studies Incentive Compatible." In *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, edited by Barbara J. Kanninen, 65–108. The Economics of Non-Market Goods and Resources. Springer Netherlands.
———. 2006c. "Experimental Evidence on Alternative Environmental Valuation Methods." *Environmental and Resource Economics* 34 (1): 125–62.
———. 2014a. "Hypothetical Surveys or Incentivized Scoring Rules for Eliciting Subjective Belief Distributions?" *Working Paper 2014-05*, Center for the Economic Analysis of Risk, Georgia State University.
———. 2017. "Behavioral Responses to Surveys about Nicotine Dependence." *Health Economics* 26: 114–23.
Harrison, Glenn W., Robert L. Beekman, Lloyd B. Brown, Leianne A. Clements, T. M. McDaniel, Sherry L. Odom, and M. B. Williams. 1999. "Environmental Damage Assessment with Hypothetical Surveys: The Calibration Approach." In *Topics in Environmental Economics*, edited by Mattias Boman, Runar Brännlund, and Bengt Kriström, 217–40. Economy & Environment. Dordrecht: Springer Netherlands.
Harrison, Glenn W., Ronald M. Harstad, and E. Elisabet Rutström. 2004. "Experimental Methods and Elicitation of Values." *Experimental Economics* 7 (2): 123–40.
Harrison, Glenn W., Morten I. Lau, Don Ross, and J. Todd Swarthout. 2017. "Small Stakes Risk Aversion in the Laboratory: A Reconsideration." *Economics Letters* 160 (November): 24–28.
Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. "Estimating Individual Discount Rates in Denmark: A Field Experiment." *American Economic Review* 92 (5): 1606–17.
Harrison, Glenn W., Jimmy Martínez-Correa, and J. Todd Swarthout. 2014. "Eliciting Subjective Probabilities with Binary Lotteries." *Journal of Economic Behavior & Organization* 101: 128–40.
Harrison, Glenn W., Jimmy Martínez-Correa, J. Todd Swarthout, and Eric R. Ulm. 2015. "Eliciting Subjective Probability Distributions with Binary Lotteries." *Economics Letters* 127 (February): 68–71.

Harrison, Glenn W., Jimmy Martinez-Correa, J. Todd Swarthout, and Eric R. Ulm. 2017. “Scoring Rules for Subjective Probability Distributions.” *Journal of Economic Behavior & Organization* 134: 430–48.
Harrison, Glenn W., and Jia Min Ng. 2016. “Evaluating the Expected Welfare Gain from Insurance.” *Journal of Risk and Insurance* 83 (1): 91–120.
Harrison, Glenn W., and Richard D. Phillips. 2014. “Subjective Beliefs and Statistical Forecasts of Financial Risks: The Chief Risk Officer Project.” In *Contemporary Challenges in Risk Management*, edited by Torben Juul Andersen, 163–202. London: Palgrave Macmillan UK.
Harrison, Glenn W., and Don Ross. 2017. “The Empirical Adequacy of Cumulative Prospect Theory and Its Implications for Normative Assessment.” *Journal of Economic Methodology* 24 (2): 150–65.
———. 2018. “Varieties of Paternalism and the Heterogeneity of Utility Structures.” *Journal of Economic Methodology* 25 (1): 42–67.
Harrison, Glenn W., Thomas F. Rutherford, and Ian Wooton. 1993. “An Alternative Welfare Decomposition for Customs Unions.” *Canadian Journal of Economics* 26 (4): 961–68.
Harrison, Glenn W., and E. Elisabet Rutström. 2008. “Risk Aversion in the Laboratory.” *Research in Experimental Economics* 12 (June): 41–196.
Harrison, Glenn W., and Elisabet E. Rutström. 2006. “Eliciting Subjective Beliefs about Mortality Risk Orderings.” *Environmental & Resource Economics* 33 (3): 325–46.
Harrison, Glenn W., and Eric R. Ulm. 2016. “Recovering Subjective Probability Distributions.” *Working Paper 2016-01*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.
Hilgert, Marianne A., Jeanne M. Hogarth, and Sondra G. Beverly. 2003. “Household Financial Management: The Connection between Knowledge and Behavior.” *Federal Reserve Bulletin* 89: 309.
Holzmann, Robert. 2010. “Bringing Financial Literacy and Education to Low and Middle Income Countries: The Need to Review, Adjust, and Extend Current Wisdom.” *IZA Discussion Paper 5114*, Institute for the Study of Labor (IZA).
Huber, Jeffrey T., Robert M. Shapiro II, and Mary L. Gillaspy. 2012. “Top Down versus Bottom Up: The Social Construction of the Health Literacy Movement.” *The Library Quarterly* 82 (4): 429–51.
Huston, Sandra J. 2010. “Measuring Financial Literacy.” *Journal of Consumer Affairs* 44 (2): 296–316.
Janis, Irving L. 1971. “Groupthink.” *Psychology Today* 5 (6): 43–46, 74–76.
Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Boston: Houghton, Mifflin.
———. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. 2nd ed. Boston: Houghton Mifflin.
Johnson, Bridget. 2010. “Obama names April National Financial Literacy Month.” The Hill. April 2, 2010. http://thehill.com/blogs/on-the-money/economy/65901-obama-names-april-national-financial-literacy-month.
Jump$tart Coalition. 2007. *National Standards in K-12 Personal Finance Education: With Benchmarks, Knowledge Statements, and Glossary*. Jump$tart Coalition for Personal Financial Literacy. https://www.jumpstart.org/what-we-share/financial-literacy-facts/.
Kadane, Joseph B., and Robert L. Winkler. 1988. “Separating Probability Elicitation from Utilities.” *Journal of the American Statistical Association* 83 (402): 357–63.
Kennickell, Arthur B., Martha Starr-McCluer, and Annika E. Sundén. 1996. “Saving and Financial Planning: Some Findings from a Focus Group.” *SSRN Electronic Journal* Available at SSRN.
Kim, J. 2001. “Financial Knowledge and Subjective and Objective Financial Wellbeing.” *Consumer Interests Annual* 47 (January).
King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Techniques in Political Analysis. Ann Arbor: University of Michigan Press.

Kripalani, Sunil, Rachel Bengtzen, Laura E. Henderson, and Terry A. Jacobson. 2008. "Clinical Research in Low-Literacy Populations: Using Teach-Back to Assess Comprehension of Informed Consent and Privacy Information." *IRB: Ethics & Human Research* 30 (2): 13–19.

Lin, Judy T., Christopher Bumcrot, Tippy Ulicny, Annamaria Lusardi, Gary Mottola, Christine Kieffer, and Gerri Walsh. 2015. "National Financial Capability Study in the United States: Data from the 2015 State-by-State Survey." http://www.usfinancialcapability.org/downloads.php.

Loomes, Graham, and Robert Sugden. 1998. "Testing Different Stochastic Specifications of Risky Choice." *Economica* 65 (260): 581–98.

Lusardi, Annamaria. 2003. "Planning and Saving for Retirement." *Working Paper Department of Economics*, Dartmouth College.

Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth." *Journal of Monetary Economics* 54 (1): 205–24.

———. 2008. "Planning and Financial Literacy: How Do Women Fare?" *American Economic Review* 98 (2): 413–17.

Lusardi, Annamaria, and Olivia S Mitchell. 2011. "Financial Literacy and Planning: Implications for Retirement Well-Being." In *Financial Literacy: Implications for Retirement Security and the Financial Market Place*, by Annamaria Lusardi and Olivia S. Mitchell, 17–39. Oxford University Press.

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

Lusardi, Annamaria, Olivia S. Mitchell, and Vilsa Curto. 2010. "Financial Literacy Among the Young." *Journal of Consumer Affairs* 44 (2): 358–380.

Mandell, Lewis. 2008. "The Financial Literacy of Young American Adults: Results of the 2008 National Jump$tart Coalition Survey of High School Seniors and College Students." *Report Prepared for Jump$tart Coalition*.

Manning, Willard G., Anirban Basu, and John Mullahy. 2005. "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data." *Journal of Health Economics* 24 (3): 465–88.

Matheson, James E., and Robert L. Winkler. 1976. "Scoring Rules for Continuous Probability Distributions." *Management Science* 22 (10): 1087–96.

McKenzie, Craig R.M, and Michael J Liersch. 2011. "Misunderstanding Savings Growth: Implications for Retirement Savings Behavior." *Journal of Marketing Research* 48: S1–13.

Merkle, Christoph, and Martin Weber. 2011. "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence." *Organizational Behavior and Human Decision Processes* 116 (2): 262–71.

Mitchell, Olivia S., Justine Hastings, and Eric Chyn. 2011. "Fees, Framing, and Financial Literacy in the Choice of Pension Manager." In *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, by Annamaria Lusardi and Olivia S. Mitchell, 101–15. Oxford; New York: Oxford University Press.

Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–17.

Nape, Steven, Peter Frykblom, Glenn Harrison, and James C. Lesley. 2003. "Hypothetical Bias and Willingness to Accept." *Economics Letters* 78 (3): 423–30.

Neill, Helen R., Ronald G. Cummings, Philip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin. 1994. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70 (2): 145–54.

Nigolian, Cynthia J., and Karin L. Miller. 2011. "SUPPORTING Family Caregivers: Teaching Essential Skills to Family Caregivers." *The American Journal of Nursing* 111 (11): 52–58.

Noctor, Michael, Sheila Stoney, and Robert Stradling. 1992. "Financial Literacy: A Discussion of Concepts and Competences of Financial Literacy and Opportunities for Its Introduction into Young People's Learning." Report Prepared for the National Westminster Bank, National Foundation for Education Research. London.

Oehlert, Gary W. 1992. "A Note on the Delta Method." *The American Statistician* 46 (1): 27–29.
Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, And Societies*. Princeton, NJ: Princeton Univ. Press.
Paolino, Philip. 2001. "Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables." *Political Analysis* 9 (4): 325–46.
Peck, Connie. 1996. *The United Nations as a Dispute Settlement System: Improving Mechanisms for the Prevention and Resolution of Conflict*. The Hague; Boston, MA: Kluwer Law International.
Perozek, Maria. 2008. "Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something We Don't Know?" *Demography* 45 (1): 95–113.
Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica* 66 (3): 497–527.
Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" In *Language, Mind, and Knowledge*, by Keith Gunderson, 131–93. Volume 7: Minnesota Studies in Philosophy of Science. University of Minnesota Press.
Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior & Organization* 3 (4): 323–43.
Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68 (5): 1281–92.
Remund, David L. 2010. "Financial Literacy Explicated: The Case for a Clearer Definition in an Increasingly Complex Economy." *Journal of Consumer Affairs* 44 (2): 276–295.
Rooij, Maarten van, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101 (2): 449–72.
Saito, Tetsuya. 2018. "Theoretical Model: Condorcet's Jury Theorem, Part 1." Wolfram Demonstrations Project. April 4, 2018.
Savage, Leonard J. 1954. *The Foundations of Statistics*. John Wiley and Sons.
———. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (336): 783–801.
Savage, Leonard J. 1972. *The Foundations of Statistics*. Second Edition. New York: Dover Publications.
Seley, Jane Jeffrie, and Katie Weinger. 2007. "Executive Summary: The State of the Science on Nursing Best Practices for Diabetes Self-Management." *The American Journal of Nursing* 107 (6): 6–11.
Servon, Lisa J., and Robert Kaestner. 2008. "Consumer Financial Literacy and the Impact of Online Banking on the Financial Behavior of Lower-Income Bank Customers." *Journal of Consumer Affairs* 42 (2): 271–305.
Smith, V. Kerry, Donald H. Taylor, and Frank A. Sloan. 2001. "Longevity Expectations and Death: Can People Predict Their Own Demise?" *American Economic Review* 91 (4): 1126–34.
Summers, Lawrence H. 1994. "Investing In All People: Educating Women In Developing Countries." EDI Seminar Paper, No. 45: World Bank Institute (WBI). Washington, D.C.: The World Bank.
Sunstein, Cass R., and Reid Hastie. 2015. *Wiser: Getting beyond Groupthink to Make Groups Smarter*. Boston, Massachusetts: Harvard Business Review Press.
Task Force on Financial Literacy, Office of the Canadian Minister of Finance. 2009. "About Financial Literacy: Definition." 2009. http://www.financialliteracyincanada.com.
Turnham, Jennifer. 2010. "Attitudes to Savings and Financial Education Among Low-Income Populations:" *Working Paper Series* University of Wisconsin-Madison Center for Financial Security.
Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323.
U.S. Financial Literacy and Education Commission. 2006. "Taking Ownership of the Future: The National Strategy for Financial Literacy."
Vitt, Lois A., Carol Anderson, Jamie Kent, Deanna M. Lyter, Jurg K. Siegenthaler, and Ward Jeremy. 2000. "Personal Finance and the Rush to Competence: Financial Literacy Education in the U.S." Middleburg, Virginia: Institute for Socio-Financial Studies.
Vygotskij, Lev S. 1962. *Thought and Language*. Cambridge, Mass: MIT Press.

———. 1987. “Thinking and Speech.” In *The Collected Works of L.S. Vygotsky, Volume 1: Problems of General Psychology*, in R.W. Rieber & A.S. Carton (Eds.), 39–285. New York: Plenum Press. (Original work published 1934.).
Wakker, Peter, Ido Erev, and Elke U. Weber. 1994. “Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories.” *Journal of Risk and Uncertainty* 9 (3): 195–230.
Weisstein, Eric W. 2018. “Condorcet’s Jury Theorem.” Text. April 4, 2018. http://mathworld.wolfram.com/CondorcetsJuryTheorem.html.
White, Michelle J. 2009. “Bankruptcy: Past Puzzles, Recent Reforms, and the Mortgage Crisis.” *American Law and Economics Review* 11 (1): 1–23.
Wilcox, Nathaniel T. 2013. Is the Premise of Risk Calibration Theorems Plausible? Presentation, CEAR Workshop, Durham University, September 17, 2013.
Winkler, Robert L., and Allan H. Murphy. 1970. “Nonlinear Utility and the Probability Score.” *Journal of Applied Meteorology* 9 (1): 143–48.
Yoong, Annamaria. 2011. “Financial Illiteracy and Stock Market Participation: Evidence from the RAND American Life Panel.” In *Financial Literacy: Implications for Retirement Security and the Financial Market Place*, by Annamaria Lusardi and Olivia S. Mitchell, 76–97.