Subjective belief distributions and the characterization of economic literacy

Di Girolamo, Amalia; Harrison, Glenn W.; Lau, Morten I.; Swarthout, J. Todd

DOI:
10.1016/j.socec.2015.08.004

Subjective Belief Distributions and the, Characterization of Economic Literacy

Amalia Di Girolamo , Glenn W. Harrison , Morten I. Lau , J.Todd Swarthout

PII: S2214-8043(15)00101-9
DOI: 10.1016/j.socec.2015.08.004
Reference: JBEE 146

To appear in: *Journal of Behavioral and Experimental*

Received date: 9 November 2014
Revised date: 3 July 2015
Accepted date: 11 August 2015

Please cite this article as: Amalia Di Girolamo , Glenn W. Harrison , Morten I. Lau , J.Todd Swarthout , Subjective Belief Distributions and the, Characterization of Economic Literacy, *Journal of Behavioral and Experimental* (2015), doi: 10.1016/j.socec.2015.08.004

Highlights

- We characterize literacy as the subjective beliefs over possible responses to a question.
- We elicit subjective belief distributions using incentive-compatible methods.
- We measure the confidence that an individual has about their knowledge of some fact.
- We show considerable heterogeneity in literacy levels over economic domains.

Subjective Belief Distributions and the Characterization of Economic Literacy

by

Amalia Di Girolamo, Glenn W. Harrison, Morten I. Lau and J. Todd Swarthout †

July 2015

ABSTRACT.

We characterize the literacy of an individual in a domain by their elicited subjective belief distribution over the possible responses to a question posed in that domain. By eliciting the distribution, rather than just the answers to true/false or multiple choice questions, we can directly measure the confidence that an individual has about their knowledge of some fact. We consider literacy across several financial and economic domains. We find considerable demographic heterogeneity in the degree of literacy. We also measure the degree of consistency within a sample about their knowledge, even when that knowledge is imperfect.

† Department of Economics, University of Birmingham, UK (Di Girolamo); Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, Copenhagen Business School, Denmark (Lau); Department of Accounting and Finance, Durham University, UK (Lau); Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout). Harrison and Lau are also affiliated with the Institute for the Study of Labor (IZA), Germany, and Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contacts: a.digirolamo@bham.ac.uk, gharrison@gsu.edu, mla.eco@cbs.dk and swarthout@gsu.edu. We are grateful to the Danish Social Science Research Council (grant #12-130950) and the Center for Actuarial Excellence Research Fund of the Society of Actuaries for financial support. We are grateful to the reviewers for helpful comments.

Table of Contents

When we say that someone is literate we mean more than that they can just "read and write." The term more generally indicates someone who is educated, whether by formal or informal means, and able to comprehend topics through words.[1] 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 questions, 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.

[1] The *Oxford English Dictionary (Second Edition)* defines the adjective "literate" as someone who is "acquainted with letters or literature; educated, instructed, learned." Remund [2010] offers a balanced account of the many definitions of literacy found in the academic and policy literature. Our focus on financial knowledge corresponds to the first of his categories of conceptual definitions of literacy (p. 279).

Our approach is to elicit the entire subjective belief *distribution* that an individual has, to ascertain how precise their knowledge is in response to some question. This extends and generalizes the prevailing approach to measuring literacy, which considers responses to (hypothetical) multiple choice questions (e.g., Lusardi and Mitchell [2007][2008][2013; §3]). For a specific question or domain, we are able to say "how literate" the person is, rather than just say that they are or are not literate. Of course, by asking a series of questions one can ascertain the fraction of correct answers for an individual with the traditional approach, but that requires one to pool responses over different questions which may span different knowledge domains.

The domains of interest to us are financial and economic knowledge. We consider a mixture of questions in which the correct answer involves the application of logical and grammatical principles, and questions in which the correct answer involves some specific fact. This reflects a trend in the measurement of literacy towards the ability to draw logical or grammatical conclusions from information presented in the question itself, and to also consider awareness of facts that are of importance for the functioning of the individual. We apply and extend two questions on financial planning and literacy by Lusardi and Mitchell [2007][2008] that have been used in the 2004 Health and Retirement Survey (HRS). These questions measure the ability to understand and apply simple economic concepts that are important in economic planning for the future such as compounding of interest and inflation. We also ask individuals two questions on longevity, which is relevant for retirement planning. These questions have been used by Bateman, Eckert, Geweke, Louviere, Thorp and Satchell [2012] and in the HRS since 1992. The four questions are documented and discussed further in section 1.

A byproduct of our characterization is that we can also say something about the degree of common knowledge that a sample of individuals have about some proposition. Quite apart from whether or not a given individual knows the true answer with some precision, we often want to know if a group of individuals have the same degree of knowledge. In effect, we are able to operationalize several interpretations of what it means to have heterogeneous beliefs.

Literacy is an important characteristic of economic behavior in it's own right. It is also something that behavioral and experimental economists should be interested in, since it goes to the heart of whether someone has understood some task or not. If behavioral economics is concerned with decisions that are commonly characterized as "mistakes" relative to some standard model or normative criterion, then it is critical to know if the decision was a "correct" decision for a misunderstood task, or an "incorrect" decision for an understood task. In psychology this is the area of task representation, and literacy is one input into a subject arriving at their representation. Similarly, when experimental economists claim that a subject has responded to a task with certain properties, such as incentive-compatibility, they need to know whether the subject has indeed understood the task and it's properties: again, literacy is one input into this process.[2]

Our results show considerable variation in literacy levels over the financial and economic domains we consider here and across observable demographic characteristics. In particular, older subjects are more literate than younger subjects in the interest compounding domain, and women exhibit higher literacy than men in the inflation domain. We also find that whites are more literate than non-whites regarding the expected remaining lifetime of men, and older subjects are more literate than younger subjects regarding the expected remaining lifetime for women.

In section 1 we describe the experimental task that we developed and employed with a sample of 120 subjects. We review in detail the properties of the subjective belief elicitation procedure in

[2] An important example in experimental economics is the Becker, DeGroot and Marschak [1964] procedure for eliciting certainty-equivalents of lotteries. There remains considerable controversy about whether subjects understand the incentive-compatibility of this procedure (Cason and Plott [2014]).

section 2 and present results on the degree of literacy of our subjects in section 3. In section 4 we consider the consistency of knowledge across subjects, and section 5 concludes.

1. Procedures

A. Literacy

We consider literacy in terms of four specific questions asked of each subject in an experiment. 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 1 presents the interface.[3] The interface implements the Quadratic Scoring Rule discussed in section 2. 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. The instructions explained that they could earn up to £20, as shown in Figure 2, but only by allocating all 100 tokens to one interval *and* that interval containing the true answer: if the true answer was just outside the selected interval, they would in that case receive £0. At the time of the experiments in December 2012, £20 was worth roughly $32.

[3] The instructions are reproduced in full in Appendix A. The interface was initialized with 10 tokens allocated to each bin.

Subjects were rewarded for one of these belief elicitation tasks, with the task selected at random.[4] The correct answer was revealed, and their earnings calculated according to the number of tokens allocated to the true interval in their elicited beliefs. For example, if the respondent had reported the beliefs in Figure 1, she would have been paid £16.25 if the correct answer was between 8% and 9.99%. As it happens, the correct answer here is 7.9%, so the subject would have actually received £11.25 since the correct answer was in the next lower interval, corresponding to unemployment rates between 6% and 7.99%.

The incentivized questions were as follows:

- **Q1: Interest Compounding**. "Suppose you had £100 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 on this account in total?" The correct answer is £110.40, and responses were elicited between £105 and £115 in intervals of £1.
- **Q2: Inflation**. "Suppose you had £200 in a saving account. The interest rate on your saving account was 1% per year and inflation was 2% per year. After 1 year, how much would be the value of the money on this account?" The correct answer is £197.96, and responses were elicited between £195 and £205 in intervals of £1. Due to a rounding error we rewarded subjects as if the correct answer was £198, and treat that as the "correct" answer.
- **Q3: Expected Lifetime for Men**. "Based on 2010 National Statistics, if a man lived to be 20 in the United Kingdom, how many more years would he expect to live? Note that this is not the age he would die at, but how many more years he would expect to live." The correct answer is 59.1 years, and responses were elicited between 0 and 100 years in intervals of 10 years.[5]

---

[4] The experiments also contained 4 belief elicitation questions that posed questions that tested knowledge of statistical principles and more specifically the application of Bayes Rule. The first pair of questions on the risk of breast cancer are taken from the 15 questions considered by Gigerenzer and Hoffrage [1995; Table 2, p. 693]. Our Q5 and Q6 are direct translations of the parameters they use for the same breast cancer risk question. The difference between Q5 and Q6 is that the former uses conventional probability information to set up the problem, and the latter uses natural frequencies to set up the same problem. Gigerenzer and Hoffrage [1995; Figure 3, p. 694] report a dramatic improvement in the correct application of Bayes Rule as their subjects moved from the probability format to the frequency format. Our questions Q7 and Q8 consider another application of Bayes Rule reasoning, but with a financial example having to do with the risk of consumer credit solvency. This issue is particularly important in the United Kingdom, which had relatively high levels of consumer debt coming into the global recession of 2008. Our conclusions do not rely on the inclusion of these 4 belief elicitation questions, which is why we do not report the results here.

[5] At the time of the experiment we did not have access to the correct Life Tables, and instead subtracted 20 from the expected lifetime at birth from the UK Office of National Statistics to pay subjects that had this question selected. The difference for aggregates such as "all men" or "all women" is tiny, and did not affect the payments to any subject given that the bin intervals we used were in 10-year increments: there is a

- **Q4: Expected Lifetime for Women**. "Based on 2010 National Statistics, if a woman lived to be 20 in the United Kingdom, how many more years would she expect to live? Note that this is not the age she would die at, but how many more years she would expect to live." The correct answer is 62.9 years, and responses were elicited between 0 and 100 years in intervals of 10 years.

The order of presentation of questions was randomized for each subject.

difference of 0.6 of a year for men (59.1 versus 58.5), and 0.5 of a year for women (62.9 versus 62.4).

The first two questions are natural extensions of questions asked by Lusardi and Mitchell [2007][2008] in the *Health & Retirement Survey* (HRS) of 2004 in the United States.[6] This survey is naturally representative of Americans over the age of 50. Our Q1 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?" The main difference is that we ask for beliefs about the true answer over a wide range. Our Q2 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?" Lusardi and Mitchell [2012; 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]).[7]

[6] 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 the individuals know how to diversify their investment. 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, Eckert, Geweke, Louviere, Thorp and Satchell [2012] in surveys in Australia.

[7] Bateman, Eckert, Geweke, Louviere, Thorp and Satchell [2012] ask these questions of adult retirement savers in Australia, and find that 78.4% get the inflation question correct and 71.8% get the interest rate question correct.

The final two questions ask about a basic informational input to retirement planning: expected remaining lifetime, conditional on reaching the age of 20. Indeed, 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 easily extend these questions to elicit more precise beliefs about someone who is more similar to the subject.

The most widely used subjective beliefs about longevity come from the *Health and Retirement Survey*, which has asked a simple question since 1992 to respondents under the age of 65: "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 themself, 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, these questions are not incentivized, and do not elicit information on the confidence of the subjective belief. However, 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.

B. Demographics and Additional Measures

Apart from these incentivized subjective belief questions, which are our main focus, we asked subjects several hypothetical questions that have been widely used in the literature on cognitive abilities. One is **Cognitive Reflection Test** (CRT) proposed by Frederick [2005], consisting of three questions:

- 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?
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

The correct answers are £0.05, 5 minutes and 47 days. Many subjects responding to these hypothetical questions fail to "reflect" on some aspect of the information provided.[8] The other hypothetical battery is known as the **Berlin Numeracy Test**, and is due to Cokely, Galesic, Schulz, Ghazal and Garcia-Retamero [2012]:

- Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

[8] It is an interesting question whether one should care if subjects get the CRT questions right when they face incentives. In fact, many subjects do if they get sufficient time to answer the questions, see Borghans, Meijers and ter Weel [2008]. It could be argued that the CRT is designed to detect cognitive propensities to spot "minimally hidden" information in decision settings, and that this propensity is better detected when higher-order, executive brain functions are not engaged to earn money.

- Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? (Please indicate the probability in percentage).
- Imagine we are throwing a loaded die (6 sides) 70 times. The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, how many times would the die show the number 6?
- In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?

The correct answers are 30, 25%, 20 and 50%.

In addition, a standard list of demographic questions were posed. These included age, sex, racial group, field of study, year of study, highest level of formal education expected to complete, current grade, citizenship, marital status, number of people in household, total income of the household, total income of parents, and smoking status.

2. Belief Elicitation

The decision maker in our experiment reports her subjective beliefs with a discrete version of a Quadratic Scoring Rule for continuous distributions, developed by Matheson and Winkler [1976]. 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 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 $\Sigma_{i=1 \ldots K} (r_i) = 1$.

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)]:

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

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, including the correct one.

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.[9] 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:

$$S = (2 \times ¼) - (¼^2 + ¼^2 + ¼^2 + ¼^2) = ½ - ¼ = ¼ < 1.$$

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.

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

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

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.

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

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 Quadratic Scoring Rule when eliciting subjective *probabilities* over *binary* events (Winkler and Murphy [1970], Kadane and Winkler [1988]), and there are various methods for addressing these concerns.[10] Harrison, Martínez-Correa, Swarthout and Ulm [2012] 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:

12. An individual reports having a positive probability for an event only if he has positive subjective probability for the event. So if the individual believes that unemployment is definitely below 12%, we would never see the individual reporting that it could be above 12%. Further, we can infer from Figure 1, for instance, that this subject truly attaches zero weight to the possibility of unemployment above 12%, no matter what his risk attitudes.

13. 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. So if the individual has a true, latent subjective probability of 0.1 that the unemployment rate is between 6% and 7.99%, and a true, latent subjective probability of 0.1 that it is between 10% and 11.99%, then the reported probabilities for these two intervals will

[10] For instance, see Köszegi and Rabin [2008], Holt and Smith [2009], Karni [2009], Andersen, Fountain, Harrison and Rutström [2014] and Harrison, Martínez-Correa and Swarthout [2014].

be the same as well, as in Figure 1 (although typically not 0.1).

14. 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.

15. 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.

16. 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.

17. It is possible to derive the effect of increased risk aversion on the difference between the reported distribution and true distribution. Harrison, Martínez-Correa, Swarthout and Ulm [2012] 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.[11]

[11] Alternatively, one could use a binary lottery procedure to "risk neutralize" subjects, as demonstrated by Harrison, Martínez-Correa, Swarthout and Ulm [2015]. Or one could use the methods developed by Harrison and Ulm [2015] to recover latent subjective belief distributions from Expected Utility Theory or Rank-Dependent Utility characterizations of risk preferences.

3. The Measurement of Economic Literacy

A. Description of Results

In December 2012 we recruited 120 subjects from Durham University. The majority had major fields of study in Finance or some other Business area, and were completing a Master of Science degree. The average age was 24.4, 67% were women, and 85% were single and had never been married. Just over 73% were non-EU citizens, and 14% were current smokers.

Figure 3 provides a quick helicopter tour of the aggregate beliefs we elicited. More formal statistical tests are provided below. We observe very precise beliefs for the interest compounding question, which was relatively easy for our sample. Far less precision is observed for the other questions. Aggregate beliefs for the economic literacy questions tended to be unimodal, with most subjects having some sense of where the correct answer was, but with varying precision.

Figures 4 and 5 begin the evaluation of individual responses for the two financial literacy questions about interest compounding and inflation and the value of money. In each case we report the correct answer, a "literacy index" and "concordance index," the responses of three individuals selected to illustrate some differences in individual behavior, and the pooled distribution. We discuss the concordance index in the next section.

We construct a simple index of literacy, $L \in [0, 1]$, given by the fraction of 100 tokens that the individual allocates to the interval containing the true answer.[12] This index does not need to be estimated: it is a direct transformation of the observed data. Thus we see a value of L=1 for subject #1

[12] There are many other literacy indices that could be constructed from these data. For instance, a natural metric would be the amount of *money* that the subject had "left on the table" by not allocating all 100 tokens to the correct answer. This reflects the QSR underlying the incentives provided to subjects to be more precise, and is therefore a non-linear transformation of all token allocations. More generally, there is one way to be completely correct about these questions, and 4,263,421,511,270 ways to be wrong. If there are t tokens and b bins, then there are (t+b-1)!/t!(b-1)! possible allocations in each of our elicitation tasks. Only one of these is completely correct. If someone has a literacy index L=0 then there are still (100+9-1)!/100!(9-1)! = 352,025,629,371 ways to respond.

in Figure 5, the interest compounding domain, since this subject allocated all 100 tokens to the interval containing the correct answer. Many subjects did exactly the same thing in this case, but subject #10 and subject #11 show how a few hedged their bets, quite literally. For subject #10, 60 of the 100 tokens were in the interval containing the correct answer, so L=0.60 for this subject.

In Figure 5 we see that subject #1 has a literacy index of zero, since she allocated all 100 tokens to the interval just to the left of the correct answer for the inflation question. In this domain, subject #5 has a literacy index of 0.26 since 26 out of 100 tokens were allocated to the correct interval. By being less dogmatic, subject #5 exhibited greater literacy than subject #1. Of course, one might want to argue that subject #1 was very close to the correct answer, but countering that claim is the subject's choice, implicitly saying that she was certain of her answer. If indeed she has some imprecision, that should have led her to report a non-degenerate distribution.

Figures 6 and 7 consider the other questions, about expected remaining lifetime for men and women. Figure 6 shows detailed responses for 11 individuals to the question about men, since we observe considerable heterogeneity in this domain compared to the financial literacy questions. The imprecision for subject #3, #5 and #7 is substantial, and leads one to speculate if it shows up in their savings behavior or retirement planning. Figure 7 shows the differences in the aggregate distribution between the questions about men and women, to gauge if the longer expected lifetime of women compared to men is generally understood. Indeed, we see that this increment is detected in aggregate, so we can say that we observe a reasonable "literacy of crowds" on this issue.

Finally, Figure 8 collates information on the distribution of the literacy indices L across the domains considered. The vertical, dashed line is the average of each distribution, for reference. The distribution for the interest compounding question, in the top left panel of Figure 8, is what one would normatively like to see: almost universal high-literacy. However, one can visually infer that this is the exception across these domains.

Our measure of literacy is sensitive to the range of possible answers and the intervals of the 10 bins in the belief elicitation task. If one follows the prevailing approach the literacy index is either 1 or 0 depending on whether the individual provides the correct answer to the question or not. By eliciting the distribution of beliefs we move away from a binary measure of literacy at the level of the individual and the measure of literacy is now a discrete variable determined by the fraction of tokens allocated to the interval that contains the correct answer. Increasing the size of the intervals in the belief elicitation task increases the likelihood that the individual will put a larger fraction of tokens on the interval with the correct answer, and thus the likelihood of being characterized as more literate. A related issue is the symmetry of the interval around the correct answer. If the individual is uncertain about the correct answer to the question then our measure of literacy is sensitive to where the correct answer falls in the interval, and we are more likely to report a higher level of literacy if the correct answer is the midpoint of the interval instead of one of the two endpoints. There is no correct way to design the belief elicitation task and we chose a design with 10 symmetric intervals of the same size, but where the correct answer is not the midpoint of the interval.

B. Statistical Analysis of Results

A natural statistical model for directly evaluating the beliefs data is interval regression. In this specification the dependent variable refers to the intervals given by our elicitation "bins."[13] Each of the 10 intervals in the belief elicitation task is weighted by the number of tokens that the subject puts on the interval, and we use frequency weights in the interval regression model to indicate the number

[13] Interval regression allows one to identify clopen intervals with -∞ as a lower bound *or* +∞ as an upper bound, but that is not generally appropriate for our elicitation tasks. We evaluate the effects of using such clopen bounds for the interest compounding and inflation questions, and it makes no difference to our conclusions.

of duplicated observations (tokens). In all cases we control for sex, age, marital status, race, whether a Finance major, whether a non-EU citizen, whether a current smoker, and the scores on the CRT and Berlin Numeracy Test. Unless otherwise noted, the age and test scores are normalized to have mean zero and unit standard deviation.

Detailed estimates from the interval regression models for each belief question are reported in Tables 1 through 4. The estimated constant for the interest compounding question (Table 1) is 109.7 which is lower than the true response of 110.4. Although there is very little variation in the dependent variable here (Figure 4), we find that the marginal effect of age is positive and significantly different from 0 (*p*-value of 0.040), indicating that older individuals are slightly more literate with respect to interest compounding. A similar conclusion has been found by Delavande, Rohwedder and Willis [2008], Lusardi, Mitchell and Curto [2010] and Lusardi and Mitchell [2014].[14] We find that females exhibit higher literacy than men in the inflation domain, contrary to the general findings in the literature using hypothetical surveys.

We find that whites are more literate than non-whites on the expected remaining life years for men (Table 3), and that older individuals are more literate on the expected remaining life years for women (Table 4). These outcomes are in line with the literature using hypothetical surveys (Lusardi and Mitchell [2014] and Bissonette and de Bresser [2015]).

[14] Delavande, Rohwedder and Willis [2008] use two sources of data: the *Cognitive Economic Survey*, an internet survey interviewing people aged 51 and above, and the *American Life Panel*, an internet panel interviewing people aged 18 and above. These two surveys have an identical group of 25 questions on financial literacy, which allows them to merge the responses. Their regression analysis leads to the conclusion that age has a *negative* correlation with financial literacy, as measured by them, even though years of formal education have a *positive* correlation with financial literacy. Lusardi, Mitchell and Curto [2010] analyze data from the 1997 *National Longitudinal Survey of Youth*, a sample of young adults aged 23 to 28 years old, and find that formal education level is positively correlated with correct answers in both the compound interest rate question and the inflation rate question. We do not control for years of education and find that age is *positively* correlated with literacy.

We evaluate the systematic effect of demographics more formally by pooling across measures of literacy *L* in different domains. We estimate a generalized linear model of the literacy index *L*, using the specification proposed by Papke and Wooldridge [1996] for fractional response variables in the unit interval. Table 5 reports the effects of the listed covariates on the literacy index. Focusing only on statistically significant effects at the 10% level, in this sample marriage is associated with an *increase in illiteracy*; being a Finance major is associated with heightened literacy; being a non-EU citizen is associated with a much lower level of literacy compared to EU citizens; better scores on the Cognitive Reflection Test are associated with lower literacy; and better scores on the Berlin Numeracy Test is associated with a significantly higher probability of being literate. These significant effects are a mix of the expected (e.g., the Berlin Numeracy Test) and the unexpected (e.g., being married). There are also some effects that the previous literature suggested to be significant, which we do not find (e.g., gender and age). The significance of these effects also points to the heterogeneity of literacy across these domains.

4. The Consistency of Knowledge

Although literacy is a capacity that is naturally measured for the individual, it obviously impacts the extent to which knowledge about something is shared.[15] If someone has a poor level of literacy in some domain, the natural question is whether that is consistent with the knowledge that others have. The immediate consequences for behavior when there are heterogeneous beliefs are by now well-studied, such as in models of asset pricing in finance (e.g., Shefrin [2008]), game theoretical interaction, and rational expectations.

[15] When measuring the literacy of a *household*, how does one account for the heterogeneity of levels of literacy within the household? The concept of effective literacy, developed by Basu and Foster [1998], considers this important dimension of what might also be called "social literacy." Obviously the measurement issues goes

beyond the household, and includes any social network used by individuals for making decisions. One approach focuses on potential literacy, and defines literacy in terms of the most literate person in the household. This adjustment to naïve measures of literacy at the individual level is easy to make, and provides a valuable upper bound. One could, however, similarly define a lower bound if the household power relationships lead to the least literate person imposing his or her will on the household decision in some setting. Using our tools for characterizing the production function for social literacy in these settings is an important extension.

These ideas are also familiar from linguistics. The process of learning a language involves the disambiguation of utterances (Allen [1995]). And many linguists discuss language use as intentionally constrained by norms of communicating understanding, which is to say greater literacy (Grice [1989], Clark [1992]). Hence one naturally seeks some measure of shared literacy. Is the uncertainty over some fact in a given domain shared, or is it a domain in which one can clearly identify "experts" and "novices?" We propose a simple measure that can allow us to address that question.

Any measuring instrument can be compared against another measuring instrument. Examples include weight scales, political opinion polls, or medical judgements about diagnoses. In our case we consider the subjective beliefs about some fact, and seek to measure their consistency. In the biostatistics literature a popular concordance index ρ_c has been developed by Lin [1989][2000]. It combines the familiar notion of correlation from a Pearson inter-class correlation coefficient with allowance for bias, and is virtually identical to measures of intra-class correlation (Nickerson [1997]). The concordance index is bounded between ±1, with the usual interpretation that $\rho_c = 1$ indicates perfect concordance, and smaller values indicate poorer concordance.[16]

In Figure 4, for instance, we evaluate the concordance index for each subject with respect to the pooled belief distribution on the interest compounding question, and then also report the average value of the index over all 120 subjects. Even though subject #1 in that setting had a literacy index value of 1, since she gave the correct responses, her concordance with the group was slightly less than 1 (0.972) because some people in the group did not have perfect literacy in this domain (e.g., subjects #10 and #11, shown in Figure 4). Taking a less extreme case, such as the inflation and value of money question in Figure 5, we see much lower levels of concordance. Subject #5, even though less precise

[16] The concordance correlation coefficient is defined as $\rho_c = 2\rho\sigma_x\sigma_y / (\sigma_x^2+\sigma_y^2+(\mu_x-\mu_y)^2)$, where μ_x and μ_x are the means of the two variables, σ_x and σ_y are the standard deviations, and ρ is the correlation coefficient between the two variables.

than subject #1 and subject #3, was more consistent with the beliefs that everyone else had.

Moving to the distribution of concordance indices in each domain, Figure 9 shows the heterogeneity of beliefs we elicited. The interest compounding question, in Figure 9, is again an outlier, showing considerable literacy in this sample (from Figure 8) and hence considerable consistency.

5. Conclusions

Literacy is a concept that is widely discussed, and clearly at the core of understanding economic behavior in modern societies. We propose a characterization of literacy using the familiar notion of a subjective belief distribution over questions in a certain domain.[17] We demonstrate how these belief distributions can be elicited in an operational, incentive-compatible manner from individuals. We show that there is considerable heterogeneity in literacy levels over economic and financial domains, and across observable demographics.

[17] The modern policy literature on literacy stresses the concept of "capability," which is the extent to which individuals use their knowledge, as distinct from being able to answer abstract questions successfully. The concept of capability seeks to characterize if someone is able to function in a certain domain. This raises many subtle, interesting issues. First, it is not obvious that someone must know the right answer in order to be able to function in some domain robustly. There are some task domains where the payoffs are very "flat," in the sense that large errors in the specific choice lead to virtually the same expected payoff as more refined choices. Second, it is also not obvious that someone must infer the right answer by applying grammatical or logical algorithms in order to make good choices: heuristics might do very well in many domains, whether or not there is a flat payoff at work in the region of choice. Third, the concept of capability raises the issue of domain-specific knowledge, which goes beyond the "reasoning from first principles and the information in front of you" approach that characterizes most analyses of literacy. Someone might be a wizard at applying Bayes Rule, but simply have an incorrect prior belief about some base rate. Such a person would typically be deemed statistically literate but not capable.

Immediate extensions of our approach are to consider broader samples and other domains of literacy, as well as the effects of controlled interventions on the distribution of literacy. Evaluation of the consequences of imperfect illiteracy can be undertaken by studying the choices made, or avoided, in related tasks that rely on literacy in that domain.[18] Do semi-literate individuals avoid welfare-improving choice domains for fear that they might make serious mistakes? Straightforward extensions to consider social literacy are also important and natural using our characterization.

[18] Another byproduct of our approach to literacy is the ability to formally characterize what it means for individuals in a sample to be "overconfident" about their literacy, in the sense that they overplace themselves relative to others. Moore and Healy [2008; p.508], Merkle and Weber [2011; p.264] and Benoit and Dubra [2011; p. 1605] explain why elicited subjective belief *distributions* are needed to evaluate such hypotheses. These characterizations would then have immediate application to the related "unskilled and unaware of it" hypothesis of Kruger and Dunning [1999] that those that tend to be less literate also tend to be the least aware of their relative disability.

Figure 1: Belief Elicitation Interface

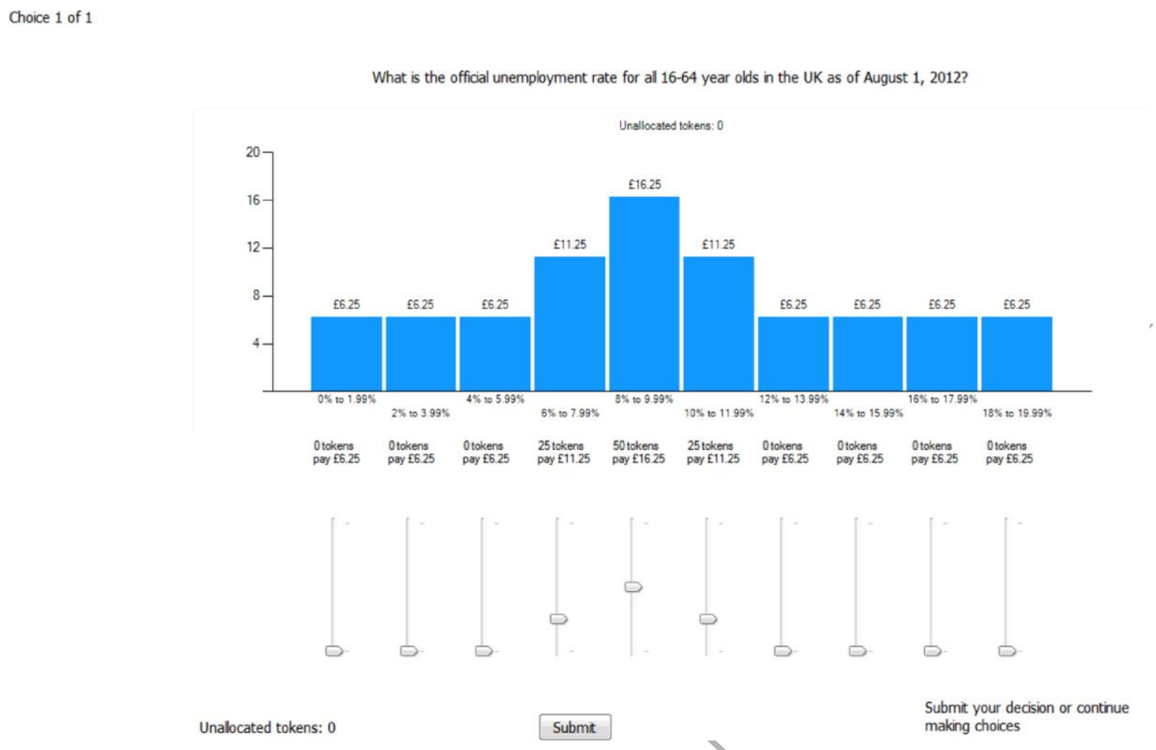

Figure 2: Possible Belief Elicitation Response

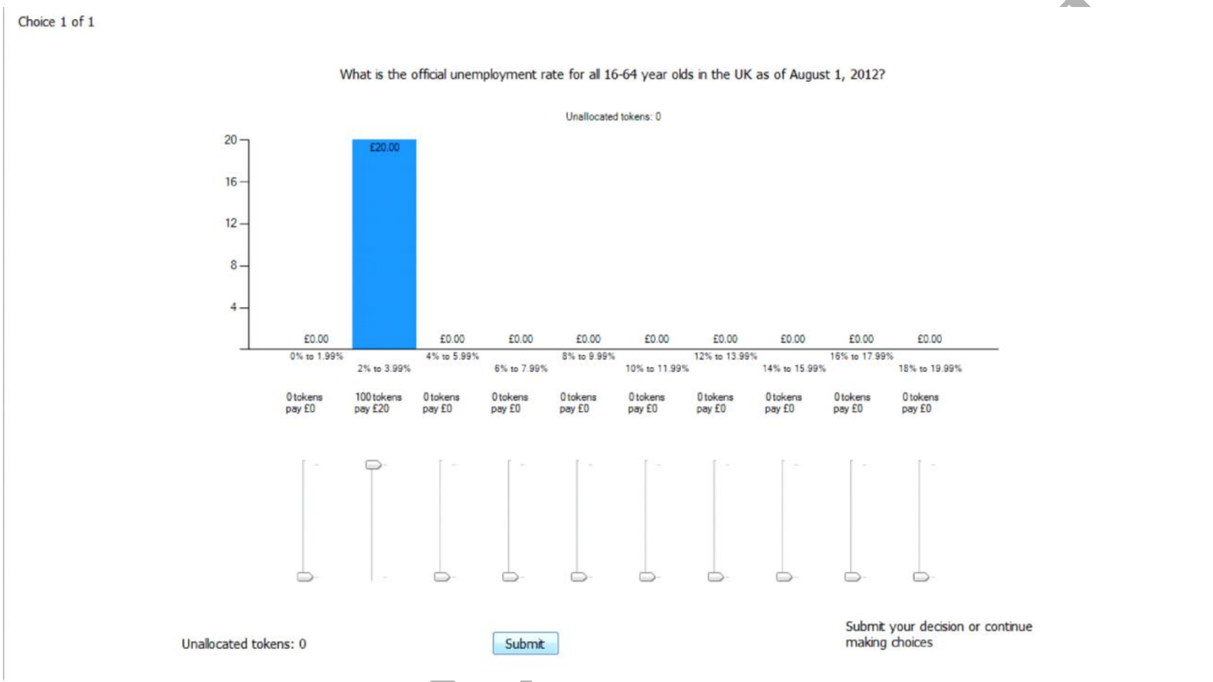

Figure 3: Pooled Subjective Beliefs

Labels in left (right) panel show lower bound (midpoint) of elicited interval

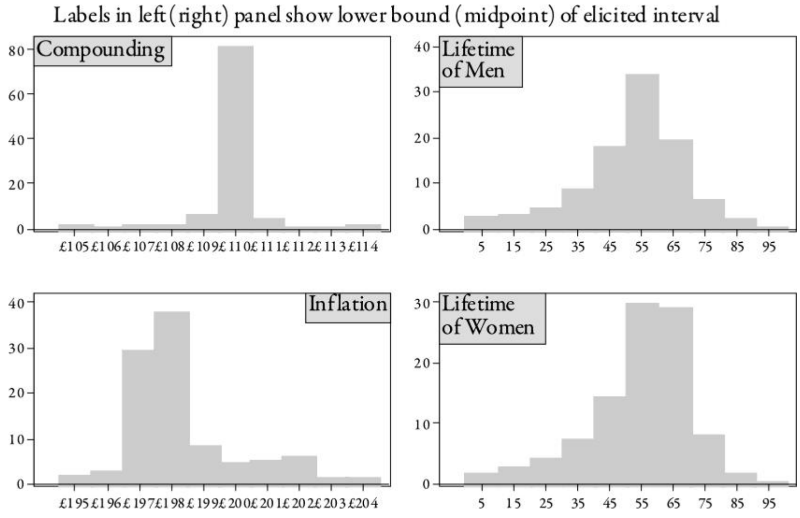

Figure 4: Subjective Beliefs of Three Subjects to Interest Rate Compounding Question

Correct Answer is £110.40

Average Literacy index L = 0.81 and average Concordance index $\rangle_c$ = 0.879

Labels show lower bound: e.g., £105 refers to the interval £105 to £105.99

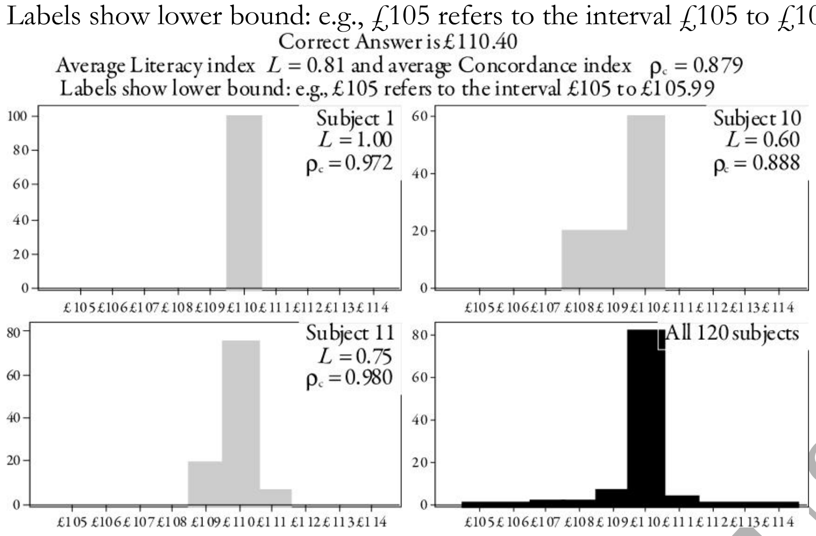

Figure 5: Subjective Beliefs of Three Subjects to Inflation and the Value of Money Question

Correct Answer is £198

Average Literacy index L = 0.38 and average Concordance index $\rangle_c$ = 0.434

Labels show lower bound: e.g., £195 refers to the interval £195 to £195.99

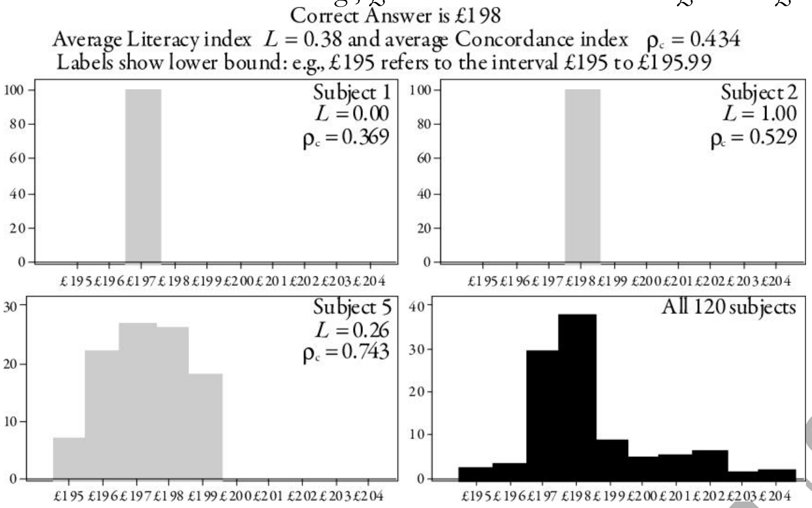

Figure 6: Beliefs on the Remaining Lifetime of Men

Correct Answer is 59.1

Average Literacy index L = 0.34 and average Concordance index $\rangle_c$ = 0.523

Labels show midpoint: e.g., 5 refers to the interval 0 to 9

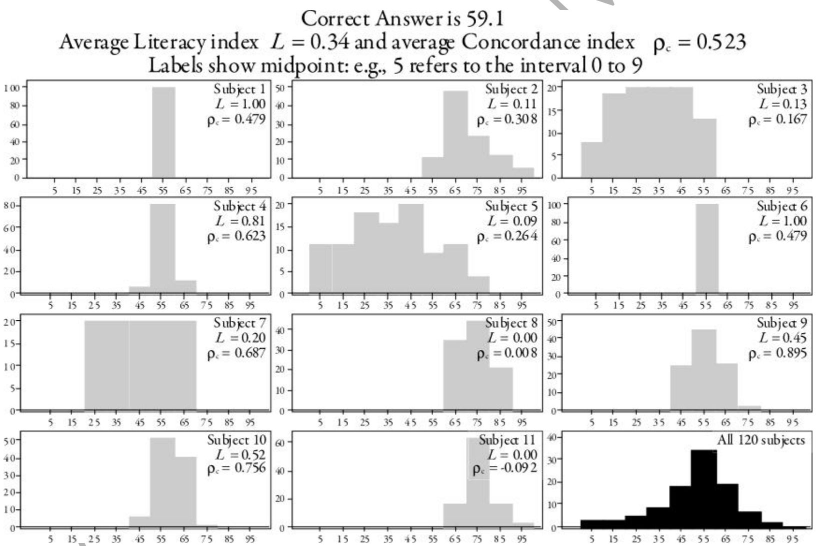

Figure 7: Elicited Beliefs For Remaining Lifetime

Correct answer is 59.1 years for men and 62.9 years for women
Average Literacy index L = .34 for men and .29 for women

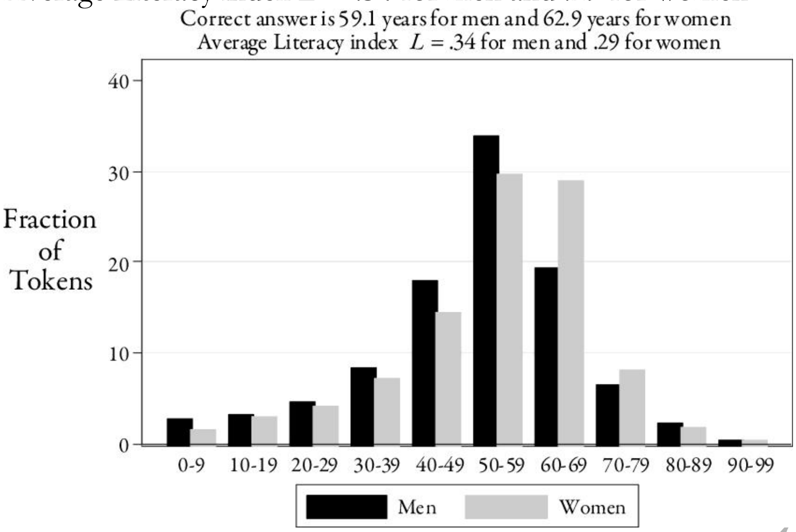

Figure 8: Literacy Indices

Dashed line is the average

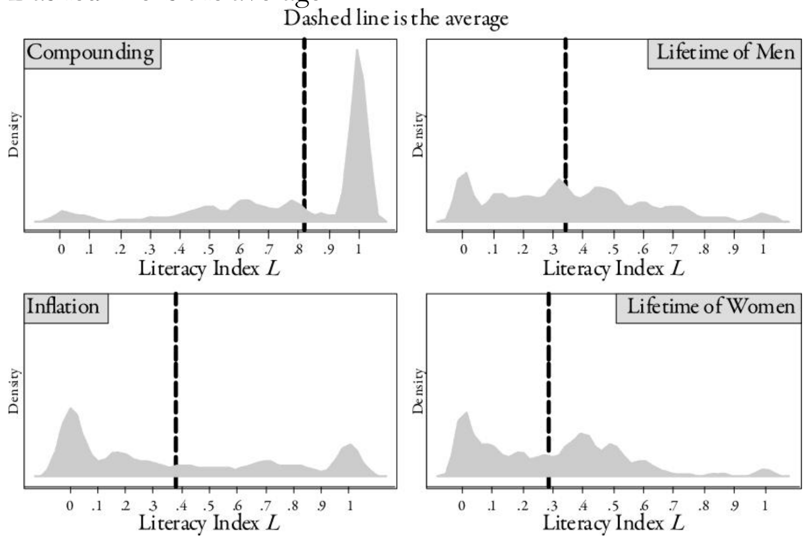

Figure 9: Concordance Indices
Dashed line is the average

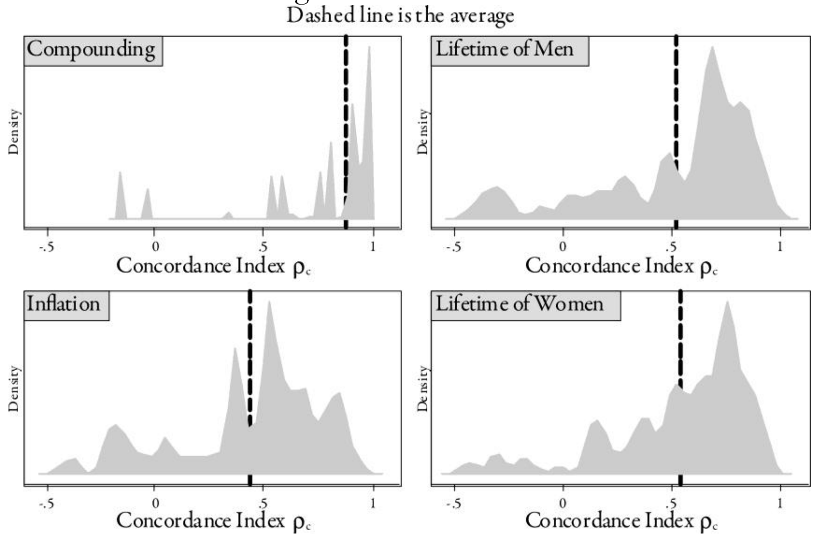

Table 1: Estimates for Interest Compounding Question

True response was £ 110.4

Variable	Estimate	Robust Standard Error	*p*-value	95% Confidence Interval		
Female	0.12	0.14	0.38	-0.15	↔	0.39
Age	0.40	0.19	0.04	0.02	↔	0.78
Single	0.64	0.45	0.15	-0.24	↔	1.51
White	0.23	0.16	0.15	-0.08	↔	0.54
Finance major	0.11	0.16	0.50	-0.20	↔	0.42
Non EU citizen	0.002	0.13	0.99	-0.25	↔	0.25
Current smoker	-0.32	0.31	0.31	-0.93	↔	0.30
Cognitive Reflection Test	-0.04	0.08	0.58	-0.19	↔	0.11
Berlin Numeracy Test	0.05	0.08	0.54	-0.11	↔	0.21
Constant	109.71	0.46	<0.001	108.81	↔	110.61
σ	0.95	0.12		0.74	↔	1.22

Table 2: Estimates for Inflation Question

True response was £198

Variable	Estimate	Robust Standard Error	*p*-value	95% Confidence Interval		
Female	0.66	0.30	0.03	0.07	↔	1.25
Age	-0.05	0.21	0.81	-0.45	↔	0.35
Single	-0.03	0.49	0.94	-1.00	↔	0.93
White	-0.15	0.42	0.73	-0.98	↔	0.68
Finance major	-0.33	0.25	0.19	-0.83	↔	0.16
Non-EU citizen	-0.78	0.47	0.10	-1.70	↔	0.14
Current smoker	-0.10	0.35	0.78	-0.77	↔	0.58
Cognitive Reflection Test	0.08	0.15	0.59	-0.21	↔	0.37
Berlin Numeracy Test	0.07	0.15	0.66	-0.23	↔	0.37
Constant	199.17	0.60	<0.001	197.99	↔	200.34
σ	1.70	0.10		1.51	↔	1.91

Table 3: Estimates for Remaining Lifetime of Men Question

True response was 59.1 years

Variable	Estimate	Robust Standard Error	*p*-value	95% Confidence Interval		
Female	2.26	2.50	0.36	-2.63	↔	7.16
Age	2.45	2.31	0.29	-2.09	↔	6.98
Single	-3.49	7.02	0.62	-17.25	↔	10.27
White	7.08	3.45	0.04	0.33	↔	13.84
Finance major	3.84	2.69	0.15	-1.43	↔	9.11
Non-EU citizen	1.31	3.62	0.72	-5.79	↔	8.41
Current smoker	0.40	2.32	0.86	-4.15	↔	4.95
Cognitive Reflection Test	-0.92	1.34	0.50	-3.54	↔	1.71
Berlin Numeracy Test	1.79	1.46	0.22	-1.06	↔	4.64
Constant	48.28	7.50	<0.001	33.59	↔	62.98
σ	15.84	0.90		14.17	↔	17.70

Table 4: Estimates for Remaining Lifetime of Women Question

True response was 62.9 years

Variable	Estimate	Robust Standard Error	*p*-value	95% Confidence Interval		
Female	0.18	2.75	0.95	-5.21	↔	5.58
Age	4.21	2.00	0.03	0.29	↔	8.13
Single	4.53	5.63	0.42	-6.50	↔	15.57
White	4.04	3.21	0.21	-2.25	↔	10.33
Finance major	2.80	2.42	0.25	-1.94	↔	7.54
Non-EU citizen	-0.06	3.63	0.99	-7.17	↔	7.05
Current smoker	1.30	2.75	0.63	-4.08	↔	6.69
Cognitive Reflection Test	-0.36	1.25	0.78	-2.81	↔	2.10
Berlin Numeracy Test	0.42	1.43	0.77	-2.38	↔	3.21
Constant	47.63	7.04	<0.001	33.83	↔	61.43
σ	15.29	0.77		13.85	↔	16.88

Table 5: Pooled Estimates Characterizing Literacy

Variable	Average Marginal Effects	Robust Standard Error	*p*-value	95% Confidence Interval		
Inflation Question						
No	0.42	0.016	<0.001	0.39	↔	0.46
Yes	0.34	0.031	<0.001	0.28	↔	0.41
Difference	-0.08	0.036	0.029	-0.15	↔	-0.008
Expected Lifetime for Men Question						
No	0.43	0.016	<0.001	0.40	↔	0.46
Yes	0.31	0.023	<0.001	0.27	↔	0.36
Difference	-0.11	0.026	<0.001	-0.17	↔	-0.06
Expected Lifetime for Women Question						
No	0.43	0.016	<0.001	0.40	↔	0.47
Yes	0.27	0.023	<0.001	0.23	↔	0.32
Difference	-0.16	0.028	<0.001	-0.21	↔	-0.11
Female						
No	0.45	0.031	<0.001	0.39	↔	0.51
Yes	0.40	0.018	<0.001	0.36	↔	0.43
Difference	-0.05	0.038	0.158	-0.13	↔	0.02
Older than 24						
No	0.40	0.018	<0.001	0.37	↔	0.44
Yes	0.44	0.026	<0.001	0.39	↔	0.49
Difference	0.04	0.033	0.269	-0.03	↔	0.10
Single						
No	0.35	0.040	<0.001	0.27	↔	0.43
Yes	0.42	0.016	<0.001	0.39	↔	0.46
Difference	0.07	0.043	0.087	-0.01	↔	0.16
White						
No	0.43	0.021	<0.001	0.39	↔	0.47
Yes	0.38	0.027	<0.001	0.32	↔	0.43
Difference	-0.05	0.038	0.148	-0.13	↔	0.02
Finance major						
No	0.38	0.019	<0.001	0.35	↔	0.42
Yes	0.47	0.025	<0.001	0.42	↔	0.51
Difference	0.08	0.033	0.014	0.02	↔	0.14
Non-EU citizen						
No	0.49	0.038	<0.001	0.42	↔	0.57
Yes	0.38	0.017	<0.001	0.35	↔	0.42
Difference	-0.11	0.043	0.011	-0.20	↔	-0.03
Current smoker						
No	0.42	0.016	<0.001	0.39	↔	0.45
Yes	0.36	0.034	<0.001	0.29	↔	0.43
Difference	-0.06	0.040	0.123	-0.14	↔	0.02
Cognitive Reflection Test *						
No	0.43	0.023	<0.001	0.38	↔	0.47
Yes	0.40	0.021	<0.001	0.36	↔	0.44
Difference	-0.03	0.033	0.386	-0.09	↔	0.04
Berlin Numeracy Test *						
No	0.33	0.019	<0.001	0.29	↔	0.37
Yes	0.50	0.026	<0.001	0.45	↔	0.55
Difference	0.18	0.035	<0.001	0.11	↔	0.24

* Binary indicator that the Cognitive Reflection Test score was 2 or more and that the Berlin Numeracy Test score was 3 or more. In each case this indicates "smarter" individuals by these metrics.

References

Allen, James, *Natural Language Understanding* (Redwood City, CA: Benjamin/Cummings Publishing, Second Edition, 1995).

Andersen, Steffen; Fountain, John; Harrison, Glenn W., and Rutström, E. Elisabet, "Estimating Subjective Probabilities," *Journal of Risk & Uncertainty*, 48, 2014, 207-229.

Basu, Laushik, and Foster, James E., "On Measuring Literacy," *Economic Journal*, 108, November 1998, 1733-1749.

Bateman, Hazel; Eckert, Christine; Geweke, John; Louviere, Jordan; Thorp, Susan, and Satchell, Stephen E., "Financial Competence and Expectations Formation: Evidence from Australia," *Economic Record*, 88, March 2012, 39-63.

Becker, Gordon; DeGroot, Morris, and Marschak, Jacob Marschak, "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 1964, 226-232.

Bissonnette, Luc, and de Bresser, Jochem, "Eliciting Subjective Survival Curves: Lessons from Partial Identification," *Cahiers De Recherche 15-03*, Faculty of Social Sciences, University of Laval, 2015.

Borghans, Lex; Meijers, Huub, and ter Weel, Bas, "The Role of Noncognitive Skills in Explaining Cognitive Test Scores," *Economic Enquiry*, 46(1), 2008, 2-12

Cason, Timothy N., and Plott, Charles R., "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 122, 2014, 1235-1270.

Clark, Herbert H., *Arenas of Language Use* (Chicago: University of Chicago Press, 1992).

Cokely, Edward T.; Galesic, Mirta; Schulz, Eric; Ghazal, Saima and Garcia-Retamero, Rocio, "Measuring Risk Literacy: The Berlin Numeracy Test," *Judgment and Decision Making*, 7(1), January 2012, 25-47.

Delavande, Adeline; Rohwedder, Susann and Willis, Robert J., "Preparation for Retirement, Financial Literacy and Cognitive Resources," *Working Paper WP 2008-190*, Michigan Retirement Research Center, University of Michigan, 2008.

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

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19(4), 2005, 25-42.

Gigerenzer, Gerd., and Hoffrage, Ulrich, "How to Improve Bayesian Reasoning Without Instructions: Frequency Formats," *Psychological Review*, 102(4), 1995, 684-704.

Grice, Paul, *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1989).

Harrison, Glenn W, Martínez-Correa, Jimmy, and Swarthout, J. Todd, "Eliciting Subjective Probabilities with Binary Lotteries," *Journal of Economic Behavior & Organization*, 101, 2014, 128-140.

Harrison, Glenn W, Martínez-Correa, Jimmy; Swarthout, J. Todd, and Ulm, Eric "Scoring Rules for Subjective Probability Distributions," *Working Paper 2012-10*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2012.

Harrison, Glenn W, Martínez-Correa, Jimmy; Swarthout, J. Todd, and Ulm, Eric R., "Eliciting Subjective Probability Distributions with Binary Lotteries," *Economics Letters*, 127, 2015, 68-71.

Harrison, Glenn W. and Ulm, Eric "Recovering Subjective Probability Distributions," *Working Paper 2015-01*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2015.

Holt, Charles A., and Smith, Angela M., "An Update on Bayesian Updating," *Journal of Economic Behavior & Organization*, 69, 2009, 125-134.

Kadane, J. B. and Winkler, Robert L., "Separating Probability Elicitation from Utilities," *Journal of the American Statistical Association*, 83(402), 1988, 357-363.

Karni, Edi, "A Mechanism for Eliciting Probabilities," *Econometrica*, 77(2), March 2009, 603-606.

Köszegi, Botond, and Rabin, Matthew, "Revealed Mistakes and Revealed Preferences," in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook* (New York: Oxford University Press, 2008).

Kruger, Justin, and Dunning, David, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 77(6) 1999, 1121-1134.

Lin, Lawrence I-Kuei, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, 45, 1989, 255–268.

Lin, Lawrence I-Kuei, "A note on the concordance correlation coefficient," *Biometrics*, 56, 2000, 324–325.

Lusardi, Annamaria, and Mitchell, Olivia S., "Baby Boomer Retirement Security: The Roles of

Planning, Financial Literacy, and Housing Wealth," *Journal of Monetary Economics*, 54, 2007, 205-224.

Lusardi, Annamaria, and Mitchell, Olivia S., "Planning and Financial Literacy: How do Women Fare?" *American Economic Review (Papers & Proceedings)*, 98(2), 2008, 413-417.

Lusardi, Annamaria, and Mitchell, Olivia S., *Financial Literacy: Implications for Retirement Security and the Financial Marketplace* (New York: Oxford University Press, 2012).

Lusardi, Annamaria, and Mitchell, Olivia S., "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52(1), 2014, 5-44.

Lusardi, Annamaria; Mitchell, Olivia S., and Curto, Vilsa, "Financial Literacy Among the Young," *Journal of Consumer Affairs*, 44(2), Summer 2010, 358-380.

Matheson, James E., and Winkler, Robert L, "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22(10), June 1976, 1087-1096.

Merkle, Christoph, and Weber, Martin, "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence," *Organizational Behavior and Human Decision Processes*, 116, 2011, 262-271.

Moore, Don A., and Healy, Paul J., "The Trouble With Overconfidence," *Psychological Review*, 115(2), 2008, 502-517.

Nickerson, Carol A. E., "A Note on 'A Concordance Correlation Coefficient to Evaluate Reproducibility'," *Biometrics*, 53, December 1997, 1503-1507.

Papke, Leslie E., and Wooldridge, Jeffrey M., "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 1996, 619-632.

Perozek, Maria, "Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something We Don't Know?" *Demography*, 45(1), February 2008, 95-113.

Remund, David L., "Financial Literacy Explicated: The Case for a Clearer Definition in an Increasingly Complex Economy," *Journal of Consumer Affairs*, 44(2), Summer 2010, 276-295.

Savage, Leonard J., "Elicitation of Personal Probabilities and Expectations," *Journal of American Statistical Association*, 66, December 1971, 783-801.

Savage, Leonard J., *The Foundations of Statistics* (New York: Dover Publications, 1972; Second Edition).

Shefrin, Hersh, *A Behavioral Approach to Asset Pricing* (New York: Academic Press, Second Edition, 2008).

Smith, V. Kerry; Taylor, Donald H., and Sloan, Frank A., "Longevity and Expectations: Can People Predict Their Own Demise?" *American Economic Review*, 91, September 2001, 1126-1134.

van Rooij, Maarten; Lusardi, Annamaria and Alessie, Rob, "Financial Literacy and Stock Market Participation," *Journal of Financial Economics*, 101(2), August 2011, 449–472.

Winkler, Robert L., and Murphy, Allan H., "Nonlinear Utility and the Probability Score," *Journal of Applied Meteorology*, 9, February 1970, 143-148.