Eliciting Individuals' Financial Decision-Making Approaches with Verbal Protocols

Thomas Post[a,b,c], Wändi Bruine de Bruin[d]

[a] Maastricht University, School of Business and Economics, Department of Finance, Tongersestraat 53, 6211 LM Maastricht, The Netherlands, e-mail: t.post@maastrichtuniversity.nl (contact author)

[b] Open University, Faculty of Management, Science, and Technology, Department of Accounting & Finance, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

[c] Network for Studies on Pensions, Aging and Retirement (Netspar), P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[d] University of Southern California, Sol Price School for Public Policy, Dornsife Department of Psychology, Schaeffer Center for Health Policy and Economics, Center for Economic and Social Research

This version: September 28, 2020

Abstract

How exactly do individuals approach financial decisions? Do they apply sophisticated formulas, back of the envelope math, do they guess or something else? We show that literally asking individuals to explain their approaches generates new insights about those questions that are reliable in sample and have predictive power out of sample. Among other things, we find that when assessing the value of an annuity 40% of individuals use some math with some using formulas similar to actuaries; the other 60% seem to apply guessing strategies. Valuation approaches predict valuation results, valuation precision, puzzling results in earlier literature, and behavioral reaction to a priming intervention designed to match one approach discovered. Our method opens new pathways to understanding financial decision making as well as designing effective policy interventions.

JEL Classification: D01, D11, D12; D14; D91

For their valuable comments and feedback on earlier versions of this paper we thank seminar participants at Maastricht University, VU University Amsterdam, ZEW – Leibniz Centre for European Economic Research, Maastricht University Center for Neuroeconomics, University of Münster well as Erzo Luttmer, Thomas Langer, Moshe Milevsky, Peiran Jiao.

1 Introduction

A few years ago, Barbara Summers from Leeds University gave a seminar at one of the authors' institutions. Amongst other things, Barbara reported on interviews she performed with people in the UK about their financial decision-making. When asked about stocks and the stock market, some interviewees responded: "What do you mean by the word stock, what is a stock market, I never heard about it" To us, this response was striking – economists have studied many reasons why people do not participate in the stock market (e.g., high risk aversion, bad experiences with stocks, low expectations for stock returns …). But maybe there are alternative, much simpler reasons? Some people may have never heard about stocks? We realized that maybe some very important factors outside the imagination of a trained economist might have been completely left out of the equation when it comes to understanding households' financial decisions. Thus, the overarching research question we start to address in this paper is: How exactly do individuals approach financial decisions, such as valuation of a financial product? Do they apply, for example, sophisticated formulas, back of the envelope math, do they guess or something else?

Based on established methods in psychology, that is, verbal protocols, we show in two studies with in total more than 450 participants that directly asking individuals to write down and explain their approaches right after they solved a financial decision-making task, that is, valuing an annuity, generates new insights about judgement and decision-making approaches. Our results are reliable within sample (i.e., it is likely what people say they do is what they do) as well as have predictive power for behavior out of sample. Moreover, we demonstrate that by letting people self-classify what they do is as good as having a researcher analyzing their written texts. That is, the method we use has the potential to be extended to large samples and does not rely on tedious reading and coding of texts. Our results show also that new insights generated can be used to design nudges effectively, for example, for policy interventions. Finally, our results link to earlier literature on annuity valuations in that they uncover a mechanism underlying the in part puzzling results found Brown et al. (2017) (correlation between valuation results and numeracy) and Chalmers and Reuter (2012) (pension plan members not reacting on interest rate changes).

In his 2006 presidential address to the American Finance Association John Campbell describes household finance as "challenging [field] because household behavior is difficult to measure, and households face constraints not captured by textbook models" (Campbell, 2006). On both aspects – measurement as well as outside of textbook factors substantial progress has been made since then. While traditionally, often individual heterogeneity was rather controlled for to focus on studying textbook components such as risk aversion and time preferences, the literature moved forward to focus as well on what we might learn from looking into the fixed effect itself, analyzing individual differences and non-standard decision making models.

To progress in uncovering what might be going on when making a decision a typical method is to assume a range of candidate models (that might be inspired by economic or psychological theory) and then test which model is superior in explaining individuals' behavioral data (e.g., Goldstein et al., 2008; Ericson et al., 2015) or to include new factors in regression models. The literature studied the role of individual factors and characteristics such as financial literacy (e.g., Van Rooij et al., 2011), numeracy (e.g., Brown et. al, 2017), optimism (Puri and Robinson, 2007), happiness (Kaplanski et al., 2015), self-efficacy (Kuhnen and Melzer, 2018), or subjective age identity (Ye and Post, 2019). Moreover, cognitive and emotional biases, non-standard decision-making strategies such as shortcuts and heuristics, and non-standard preferences have been analyzed (e.g., Binswanger and Carman, 2010b, 2012; Ericson et al. 2015).

There is nothing wrong with such methods; we learn which model is more likely to underlie judgement and decisions are which factors relate to decision outcomes. Still, the way people really judge and make decisions might be outside the imagination of the investigator and the chosen the model candidates and factors. For example, when considering the decision to buy an annuity (the product we study as an example in this paper), do most people first evaluate the annuity by calculating its value, do they make a guess at it instead, or something completely different?

In this paper, we go one step further in uncovering how individuals approach financial decisions, using annuity valuation as an example. We apply methods established in psychology (e.g., Johnson et al., 2007; Weber et al., 2007) or mathematics education (e.g., LeFevre and Morris 1999; Hogan and Brezinski 2003) developed for eliciting thought and decision-making processes, the so-called the think-aloud or type-aloud verbal protocol method. That is, first we let individuals solve a financial task. Then, we ask them to describe and explain what they did in order to solve the task. Finally, we classify, analyze, and cross-validate those responses.

In particular, we address the following research questions: Research Question 1 (RQ1) is exploratory and can be answered by analyzing the written texts:

RQ1: What do individuals say they do when they evaluate annuities?

A second block of questions relates to scrutinizing carefully what individuals said, that is, Is what individuals say they do reliable in that it is really what they do? This research question needs more operationalization. In turn, we break it down into the following sub questions:

RQ2: Do individual characteristics reflecting sophistication (financial literacy, numeracy) correlate positively with the sophistication of the approach used?

RQ3: Does more sophistication of the approach used predict a higher quality (i.e., direction and precision) of the task's result?

RQ4: Does an intervention designed based on the most prevalent approach uncovered causally change in behavior out of sample?

We investigate those research questions in two studies with more than in total 450 U.S. respondents that had to value an annuity, a financial product whose limited use has stimulated numerous discussions in the literature (see Section 2). In Study 1 we elicit what individuals say they do, personal characteristics capturing sophistication (financial literacy, subjective numeracy, objective numeracy), and annuity valuation parameters (estimates of life expectancy and the interest rate). In Study 2 we design and out of sample test a priming intervention based on the knowledge on approaches generated in Study 1.

We find that roughly 40% of participants in Study 1 say they used calculations. For a subset of those, the formula they use turns out to be a simplified but correct version of the actuarial annuity valuation formula – except that no discounting is involved and not all participants include correct parameter values (e.g., life expectancy) in the formula. The other 60% of participants say they guessed.

Based on a series of tests, it seems likely that what participants say they do approximates what they do. First, we show that reporting a more sophisticated approach (calculation vs. guessing) is more likely among participants' that have higher objective and subjective numeracy as well as financial literacy. Second, we show that reports of having used calculation (and the corresponding formula) predict the direction and precision of the valuation result: Parameter values used in the formula uncovered predict the valuation results in the right direction. That is, only for those who reported making calculations, higher estimated life expectancy is correlated with lower annuity rates. For those who guessed this is not the case. Moreover, for those who said having calculated, annuity rates estimated are more precise. That is, estimated values are less dispersed around the mean as well as closer to an objective benchmark (i.e., current market annuity rates). Third, our test of causal changes in behavior is successful. In Study 2, we use the insights from Study 1 to design and test an intervention by priming with a higher life expectancy (while still avoiding deception) which was a component of the calculation approach participants said they used in Study 1. In this out of sample test we find that participants' annuity valuations become lower and behavior (annuity demand) changes in line with our predictions. That is, if we provide participants with feedback on market annuity rates, the lower valuations caused by the prime make annuities based on markets rates look more attractive and demand increases.

Robustness checks show that is unlikely that our results are driven by experimental demand effects (i.e., respondents telling us sophisticated approaches as they assume we want to hear it) or that reporting having guessed reflects lack of motivation or effort.

Our design and test of an intervention in Study 2 demonstrates why moving forward in studying individual judgement and decision-making approaches is important beyond scientific curiosity and theory building but also for policy-makers. There is an ongoing discussion about which intervention (e.g., raising financial literacy, debiasing, nudging) is effective for improving households' financial decision-making. Our results demonstrate, that investigating how people evaluate financial decision options helps to improving and deepening our understanding of how individuals approach such decisions and therefore provides an informed basis for choosing and designing effective interventions.

The verbal protocol method we use in this paper brings the promise of augmenting existing research strategies. While in our sample, we clearly identify some approaches towards evaluating annuities, more evidence is needed to investigate how the verbal protocol method performs in different samples and for different decisions. In addition, while we got a better understanding what some 40% of participants in our study did, for the other 60% it is still not entirely clear what they do other than classifying it as guessing.

We structure this paper as follows: In Section 2 we motivate our choice to study annuities. Section 3 contains Study 1, where we analyze valuation approaches. In Section 4 we develop and test an intervention (Study 2) that builds on the knowledge generated in Study 1 in order to cross-validate our results out of sample. We discuss our results and conclude in Section 5.

2 Why Study Annuities?

As an example to study judgment and financial decision-making approaches we chose the decision to buy an annuity, and in particular how individuals evaluate its financial value. Annuities are financial products where the buyer pays a premium (for example a lump-sum) and in return receives a regular, guaranteed payout stream as long as he or she lives. Doing so insures against the risk of outliving one's financial resources as well as allows to smooth consumption.

One reason to focus on annuities is that an annuity valuation is likely not to be a straightforward and easy problem. Just when thinking about only the financial value (next to the insurance value) such a product might generate (or whether the value offered by an insurer is good value for money) individuals that would approach the problem from a purely mathematical point of view is already confronted with considering several challenges: First, they have to consider a long stream of cash flows – so additions, multiplications or divisions might need to be performed. Second, the decision to exchange a lump-sum into a cash flow stream involves an intertemporal trade-off – so discounting or compounding might need to be performed. Third, the duration of the cash flow stream depends on life expectancy – so expectation formation might be required.

Another reason to focus on annuities is that especially in annuity markets there is a gap between what classic economic theory predicts and actual behavior. Within the framework of a classical life-cycle model without bequest motives a risk-averse individual should annuitize all its wealth (Yaari 1965; Davidoff et al. 2005). In reality, however, annuities are rarely bought – the so-called annuity puzzle.

Various works addressed the annuity puzzle from the perspective of a rational decision-maker, that is, adding more factors into the classical model of Yaari (1996). Rational reasons for underannuitization include, for example, too high annuity prices due to adverse-selection (e.g., Mitchell et al. 1999), existence of bequest motives (e.g., Inkmann et al. 2011; Lockwood 2012), inflexibility of an annuity to deal with (e.g., health) expense shocks (e.g., Poterba 2006), interaction with long-term care risk and insurance (Davidoff 2009) or insolvency risk of the annuity provider (Schulze and Post 2010).

More recently, a strand of literature emerged looking at annuities from a behavioral perspective. This literature finds, for example, that part of the low annuity demand is related to high product complexity and cognitive constraints (e.g., Brown et al. 2017), framing effects that make the product look unattractive (Brown et al. 2008; Shu et al. 2016), default effects (Bateman et al. 2017), or evoking fear of death and mortality salience (Salisbury and Nenkov 2016). Within this stream of literature, several works find results that are consistent with that in tendency, individuals wrongly estimate the financial value of an annuity and that those incorrect estimates are related to individual's low financial literacy and numeracy (Brown et al. 2017; Bateman et al. 2018; McGowan et al. 2018) or difficulties to correctly interpret the numbers (Goldstein et al. 2016). Brown et al. (2017) speculate that valuations might be driven by simple break even heuristics, that is, individuals think about an annuity in terms of: "How long will it take me to break even."

So overall, studying a complex product like annuities has the potential to uncover unexpected judgement and decision-making approaches as well as explaining a mechanism driving findings in the literature.

3 Using the Verbal Protocol Method to Assess How Annuities Are Evaluated – Study 1

3.1 Purpose

Study 1 serves the following purposes: First, we use the verbal protocol method from psychology to gather evidence on what individuals do when they value annuities (RQ1). Second, we test if what individuals say they did is likely to be what they did. That is, we analyze how individual measures of sophistication (financial literacy, subjective and objective numeracy) relate to using a particular approach (RQ2). And, we analyze how particular approaches relate to the quality of annuity valuation results, that is, mean and precision (RQ3).

We also test if letting participants self-classify their own approach used (using a multiple-choice list) is a valid survey instrument that can be used in further studies as an alternative to typing and classifying text responses. That is, we compare how the text evaluation by a coder correlates with participants' self-classifying their approach.

Finally, we perform robustness checks. That is, we test if asking individuals about their approach creates experimental demand effects in that more individuals are triggered to use or report a more sophisticated approach (e.g., calculus) in the experiment than they naturally would do so, and if less sophisticated approaches reported reflect a lack of effort and motivation.

3.2 Method and procedure

We designed an online survey that presented participants first with an annuity valuation task and then subsequently elicited approaches used, measured participants numeracy skills, variables that are standard components in actuarial valuation formulas for annuities, as well as demographic characteristics.

We used a survey design with two different conditions to which participants were randomly assigned to. In both conditions participants were first presented with a scenario stating that a person is aged 65 today (full instructions are given in the Appendix). Then conditions differed by how the valuations task was presented: in a lump-sum or payout format.

In the lump-sum condition, participants were informed that the person had saved $500,000 and asked to fill in the monthly payout they think a person would get from a nominally constant life annuity. This scenario resembled the decision situation of an individual who approaches retirement planning from the perspective of already accumulated savings. In the payout condition, participants were asked about what they think how much the person needs to having saved in order to receive $2,800 as a nominally fixed annuity payout per month. This scenario resembled the decision situation of an individual who approaches retirement planning in terms of a desired or target payout.

Two different valuation tasks were included to understand if different approaches are used depending on the type of annuity valuation problem given, that is, the perspective the decision-maker takes. The valuation tasks were not financially incentivized to avoid that participants were triggered to perceive it like a mathematical test and in the survey would use different strategies to answer it than they would do naturally.

Both conditions avoided technical language as much as possible and informed participants that they can take as much time as they want to come up with their answer. The valuation task referred to a third person to minimize the impact of heterogeneity in participants' personal circumstances (e.g., ability to afford an annuity, time to expected retirement age, life expectancy, length of the annuity cash flow stream …) on results. The presented numerical values of the monthly payout and lump-sum corresponded to each other in that $2,800 is roughly the monthly amount an annuity pays out for a lump-sum of $500,000 based on current market rates in the U.S.[1] Moreover, we presented rounded numbers to avoid triggering a too high need for approximation (as non-round numbers like $2,817.67 might do, see, e.g., Campbell, 1995). At the same time, the we chosen numbers to avoid to easily triggering candidate approaches found in a pretest with university staff, that is, numbers than can be very easily be divided or multiplied by 12 (e.g., a lump-sum of $1,200,000).

After the valuation task, participants were requested in a multiple-line text field to describe what they did in order to come up with their numerical estimate. The text field was visible for a random half of the participants on the same screen as the task and for the other half on the next screen, that is, only after they submitted their result for the valuation task. These two versions of the experimental flow were included to test later if the survey potentially created experimental demand effects, that is, participants were more likely to report having used sophisticated methods because of feeling to be in an experimental situation.

The method to make participants tell what and how they performed the valuation is based on the think-aloud or type-aloud verbal protocol method (see, e.g., Johnson et al., 2007; Weber et al., 2007) with the precise wording being based on LeFevre et al. (1993) and Kuusela and Paul (2000). The verbal protocol prompt included the statement to "tell us how you came up with your answer" and to report "all of your thoughts that emerged when coming up with your answer." We did not use wordings like, for example, "how you computed," in order to avoid that participants felt it would be expected from them to come up with a response that included a mathematical approach (see the Appendix for the complete instructions).

On the next screen of the survey, participants were requested to self-classify (e.g., as in Johnson et al., 2007; Weber et al., 2007) their response. Classification options given to participants were chosen based on the results of a pretest[2] as well as on research on typical simplification

[1] We took annuity quotes from www.immediateannuities.com two weeks before the survey was distributed. The quotes for a lump-sum of $500,000 were $2,850 for men and $2,717 for women, so on average $2,784. Instead of this precise amount we used a value of $2,800 in order to avoid that the lump-sum scenario might appear to be an easier mathematical problem for subjects because being confronted with a round number there (Campbell, 1995).
[2] We performed a pretest with ten university staff and faculty members. Participants saw either the lump-sum or a payout task on a sheet of paper while the experimenter was present. After each participant filled in an estimate of a payout or lump-sum the experimenter handed another sheet of paper over to the participant with the question: How did you come up with this number? There are no right or wrong answers. Just tell me how you did it. From those responses we selected the answer categories for Study 1 if they occurred more than once.

strategies for mathematical problems (Campbell, 1995). In particular, we asked them to select one of the following categories: the number just popped up, I guessed, I did a calculation, I used the internet to find the answer (e.g., an annuity calculator), None of the above options is fitting. If a participant selected “I did a calculation” then he or she in addition was requested to self-classify the approach by which the calculation was made. In the lump-sum condition the options given were: I divided 500,000 by the number of years to live and then by 10, I divided 500,000 by the number of years to live and then by 10 and made some adjustments, I divided 500,000 by the number of years to live and then by 12, I divided 500,000 by the number of years to live and then by 12 and made some adjustments, None of the above is fitting. In the payout condition the options were: I multiplied 2,800 by 10 and then by the number of years to live, I multiplied 2,800 by 10 and then by the number of years to live and made some adjustments, I multiplied 2,800 by 12 and then by the number of years to live, I multiplied 2,800 by 12 and then by the number of years to live and made some adjustments, None of the above is fitting.

Next, we collected several measures capturing participants’ sophistication. Using established scales, we elicited subjective numeracy, that is, what they think how good their mathematical skills are (4-item scale from Fagerlin et al., 2007); financial literacy (5-item basic literacy scale from van Rooij et al., 2011); objective numeracy, that is, how good their mathematical and reasoning skills are based on solving a test (8-item scale from Weller et al., 2013).

After that, we collected participants’ estimates of typical annuity valuation parameters (life expectancy at age 65 for men and women, interest rate on a 10-year U.S. Treasury bond) to examine their potential role in the valuation approach. Finally, we collected standard control variables, that is, demographic information (gender, age, education, income, and savings) as well as risk tolerance (1-item scale from Dohmen et al., 2011), time preference (1-item scale from Falk et al., 2016).

3.3 Survey participants and characteristics

The online survey was distributed by the survey provider Qualtrics in February 2019 to a sample with the following selection criteria: being a U.S. resident and aged 45 to 60. We selected the lower age bound to make sure the annuity scenario is potentially relevant for participants. We excluded young individuals as our pretest indicated that young individuals were confused with the task.[3] Moreover, for young individuals thinking about and or buying an annuity is not a relevant decision problem at the current point in time. We selected the upper age bound to be below a typical retirement age to avoid that participants instead of valuing the annuity simply would report numbers known to them in case they already owned an annuity that paid them regularly.

We received in total complete 239 responses. From those, ten responses were removed from the data, as respondents entered a "0" to the valuation questions.[4] The average time to fill in the survey was 15 minutes (median: 11.5).

Before statistically analyzing the data, first the text input regarding the valuation approach was coded. Following standard procedures (e.g., LeFevre et al., 1993; LeFevre and Morris, 1999), two independent coders (not part of the author team) were instructed to first classify whether a participant had used guessing, any sort of calculation, looked up the values on the internet or neither of those options fitted. In addition, coders were instructed to report and classify any approaches not included in this list if they appeared more than five times. That is, they coded using the same categories shown to participants in the survey's self-coding part. In a second step, coders were instructed to classify the method for those calculated (again using the self-classification categories). Again, coders were instructed to report and classify any approaches not included in this list if they appeared more than five times. Overall, the results of coding were reliable in that both coders agreed in their categorization in 91% of all cases (Cohen's Kappa = 0.77). In consequence, throughout the analyses in Study 1 we rely for simplicity on the coding of one of them (all results hold if we used the other coder's classification instead).

Table 1 shows descriptive statistics for the sample:

- Table 1 here -

Based on comparison of means and median (for skewed variables) statistical tests we find no significant differences between participants in the two conditions for all variables included in Table 1, thus the random assignment of participants to the conditions worked.

[3] Younger participants in the pretest, for example, did not take the valuation problem as it was given but started to think about when they will start to earn money, how much time they have to save until age 65 or how likely it is to survive to age 65 in the first place.

[4] All results where it is technically possible to include also those responses hold (available on request). When analyzing annuity rates (Section 3.4.2.2) those responses cannot be included as performing a transformation to an annuity rate in this case involves a division by zero.

The three measures of participants' sophistication (subjective numeracy, objective numeracy, financial literacy) are positively and strongly significantly correlated at the 1% level. The correlation coefficients between subjective and objective numeracy is 0.37; between subjective numeracy and financial literacy 0.39; and between objective numeracy and financial literacy 0.52. Thus, for multivariate analyses performed later including all three measures jointly may create multicollinearity issues.[5]

Participants' estimates for the number of years a male of female person aged 65 can expect to live (17.9 and 21.6) are close to their empirical counterparts based on publicly available official estimates, that is, 18-19.2 years for men and 20.6-21.6 years for women.[6] While these estimates might indicate that participants looked up such information on the internet their estimate for the interest rate on a 10-year T-Bond (elicited on exactly the same survey screen as the life expectancy) shows that this scenario is unlikely. The mean estimate of participants is 1.6% while on the day the survey was in the field the interest rate was 2.7% according to publicly available sources.[7]

3.4 Results

3.4.1 What do participants say they do (RQ1)?

Participants provided a wide variety of responses in the text field asking them to explain how they derived their valuation estimate. Answers ranged from responses where it was rather clear that guessing was applied (e.g., "educated guess", "I'm not sure"), clear that a calculation was used (e.g., "2800 x12=33600 so I figure you will live 15 years which equals 504000", "Estimated living to age 85, I roughed it to $3000 a month and multiplied it by 12 months. And then roughed my answer by multiplying 20 years and put in a little more to my answer") but also contained more ambiguous or less clear responses (e.g., "I used a calculator", "I

[5] We also performed a factor analysis on the three sophistication measures. The analysis resulted in one factor with an Eigenvalue greater than one explaining 98.9% of the variance. In addition, we used the resulting Bartlett factor scores as weights to construct a composite measure of sophistication. Using this composite measure in our models produces virtually the same results.

[6] At the time of the survey, the latest (end of 2016) OECD statistics reported estimates of 18 (male) and 20.6 years (female) (https://data.oecd.org/healthstat/life-expectancy-at-65.htm) while the Social Security Administration reported estimates of 19.2 years (male) and 21.6 years (female) (https://www.ssa.gov/OACT/population/longevity.html).

[7] https://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/textview.aspx?data=yield

divided the amount by the age" or "500000 ÷ 20 years"). A world cloud illustration of the variety of responses is given in Figure 1.

- Figure 1 here -

Eventually, based on the classification by the coders we find that 36% of participants have used a mathematical approach / calculation. According to participants' self-classification the percentage who calculated is 42%. Overall, coders' and participants' classifications agree in 79% of all cases. Potentially, in some cases the text was too vague to being classified by coders as a clear calculation or clear guess or something else (e.g., unclear).

Looking at the predominant approach used, the Division Rule (which we describe in detail below), coder and self-classification are remarkably close: 26.2% and 26.6%. Interestingly, no new methods (i.e., other than those already used for the self-classification categories) were identified by the coders. Overall, we conclude from these findings, that participants' self-classification of their approach is a valid instrument in terms of agreeing with the coders' classification. Therefore, we use it in all subsequent analyses (results hold if we use the coder classification). Moreover, for Study 2, when we again elicited valuation approaches we can solely rely on self-classification and do not need to ask participants to write text.

Table 2 gives an overview of the distribution of the different approaches used by participants.

- Table 2 here -

According to Table 2, 49% of all participants guessed, 42% calculated, 2% used the internet to look up values, and 7% could not or were not willing to classify their approach. Based on the follow-up question in the survey we find that slightly different calculation approaches were used. In case a lump-sum was given a participant started by dividing the lump-sum first by an estimate for the expected number of years to live to arrive at annual payouts, then by 10 ("Rule 10") or 12 ("Rule 12") to arrive at monthly values. Finally, some participants made adjustments in this procedure. The approaches Rule 10 or 12 (without and with adjustments) we label as "Division Rule"). In case a payout was given, the approach was conceptually identical but in reversed order, thus, participants multiplied. For the sake of simplicity, we also call those approaches division rule. Typical "adjustments" participants made were – based on manually cross-checking with participants' text entries – rounding before a calculation (e.g., from 2.800 to 3.000) and / or adjusting the final result up or down in order to correct for suspected directional mistakes their simple method could have resulted in. Doing so is in line with the evidence found in the literature on mathematics education, for example, Campbell (1995). No participant, however (again based on reading their texts) did take an interest rate into account for discounting or compounding future payments – although based on our survey question we know that they are aware that current interest rates are positive.

Next, we compare participants' calculation approaches with the actuarial valuation of an annuity. Ignoring (for simplicity) within years discounting as well as assuming a flat interest rate term structure, the actuarial valuation formula for an annuity paying at the beginning of a month is given by:

(1a) $P = (A \cdot 12) \sum_{t=0}^{T-x-1} \frac{{}_tp_x}{(1+r)^t}$,

with P being the lump-sum premium, A the monthly payout, ${}_tp_x$, the probability of an individual aged x today to survive to age $x + t$, T the maximum possible age assumed, and r the annual interest rate. In case the interest rate is zero, the formula simplifies to:

(1b) $P = (A \cdot 12) \sum_{t=0}^{T-x-1} {}_tp_x = (A \cdot 12) \cdot life\ expectancy\ at\ age\ x$.

From (1b) it follows that participants who used the Rule 12 method used a valuation rationale which is remarkably close to the way actuaries value annuities in case the interest rate was zero. Note, however, the formula alone is not sufficient to value an annuity like an actuary – for doing that a participant needs to insert into the formula a correct estimate for life expectancy.

Table 2 contains as well statistics on the frequency of valuation approaches used within the two valuation conditions. Differences between conditions are not statistically significant. Regardless of whether participants had to estimate payouts or lump-sums, they were on average equality likely to calculate and if so using the Division Rule.

3.4.2 Do participants do what they say?

3.4.2.1 Relation of participants' sophistication with approaches reported (RQ2)

Next, we analyze how participants' sophistication (and other characteristics) relate to using calculation versus guessing in the annuity valuation task. We run logistic regressions of a dummy variable indicating having calculated or not on different sets of variables. Table 3 displays the regression results.

- Table 3 here -

Consistently across models higher sophistication (subjective and objective numeracy, and financial literacy) predict having applied a calculation, that is, a more sophisticated approach. All three sophistication measures are positively related to the likelihood to having applied a calculation when included separately in a model (see Table 3, Models 2-4) which is in line with Sinayev and Peters (2015), for example, in that numeracy predicts using more sophisticated strategies. When we include the three variables at the same time in the regression (Model 5) only subjective numeracy remains significant because of the high correlation among those three variables. Among the control variables, higher age is in some models negatively related with having applied a calculation consistent with an age related decline of crystalized intelligence (e.g., Bruine de Bruin et al. 2007, 2012, 2014).

We also analyze who used one of the four variants of the Division Rule conditional on having reported having done a calculation (Table 3, Model 6). The results are remarkably similar, in that again, especially subjective numeracy is a strong predictor of having used one of the variants of the Division Rule (and for the other numeracy measures it depends on what other covariates are included).[8] That result implies potentially, calculation rules not specified by a participant ("no option fits") most likely are less sophisticated approaches as they go along with less sophistication of the participant.

3.4.2.2 Relation of approaches used and participant sophistication with valuation quality (RQ3)

We next analyze the relation of the sophistication of an approach used as well as participants' sophistication with the annuity payouts and lump-sums estimated in the task in terms of their quality, that is, mean and dispersion. For making the numbers that participants entered in the

[8] Note, in this model objective numeracy predicts negatively having applied the division rule. This effect is caused by the high correlation between the sophistication measures. When each one is included separately, subjective numeracy and financial literacy positively and significantly predict the use of the division rule, while the coefficient for objective numeracy is not significant.

two conditions comparable on the same scale, we first transform, both payouts and lump-sums to implied annual annuity rates according to the following formula:

$$(2) \qquad Annuity\ rate = \frac{A \cdot 12}{P}\,.$$

Table 4 gives descriptive statistics for annuity rates that participants estimated in the two valuation conditions; Figure 2 plots their distribution.

- Table 4 here -

- Figure 2 here -

First, participants' responses include large outliers – mean and median annuity rates differ strongly (similar as in Brown et al., 2017). Some outliers we identified to be typos based on cross-checking with their written text (e.g., in the text input a participant said $4,166 but entered then $41,166 in the valuation field); others as potential typos (e.g., a participant stated having used the Division Rule (i.e., Rule 12), but entered an annual amount), others seem to be clear guesses. For all subsequent analyses we do not correct potential typos to avoid researcher driven bias, especially in case where the response could have been a typo (monthly vs. annual) or a true valuation mistake. We do however take outliers into account by evaluating in subsequent analyses the median of the annuity rate (univariate analyses), the logarithm of the annuity rate or discrete indicator variables (regression models).

Second, median annuity rates of 11% (payout given condition) and 5% (lump-sum given condition) estimated by participants are economically reasonable. Both values are close to the value an annuity calculator (www.immediateannuities.com) gave at the time of the survey, that is, 6.7% (= average of male and female rates).

Third, annuity rates differ between the conditions. The difference in means is significant at the 5% level, while the difference in medians is significant at the 1% level. Participants in the condition with payouts given estimate higher annuity rates (at the median) than the market offers while participants in the lump-sum conditions estimate lower rates. The differences from each condition's median annuity rate to the value of 6.7% are statistically significant at the 1% level (we discuss this effect in Section 5).

Forth, we analyze how the valuation result is related to participant's sophistication as well the approach used. Given the outliers and skewness of the annuity rate variable we first transform it by taking the natural logarithm. Then, we run several OLS regression models of the log annuity rate on different sets of covariates. Results are shown in Table 5.

- Table 5 here -

Models 1 and 2 show that variables capturing to participants' sophistication (subjective numeracy, objective numeracy, and financial literacy) and potentially related to sophistication (higher education, higher income) predict smaller annuity rates. As the overall sample mean of the annuity rate is higher than the market rate of 6.7% (compare Table 4), this finding implies that higher sophistication is related to more realistic results in the valuation task. In Model 3, where include all demographic characteristics and sophistication variables jointly, no coefficient is statistically significant due to the high correlation between those variables.[9]

Models 4 and 5 demonstrate that whether a participant had applied a calculation or not, as well as valuation inputs (life expectancy at age 65, interest rate) do not explain annuity rates.[10] But, when interacting the valuation inputs with the dummy variable indicating having used a calculation (Model 6) the main effect for life expectancy as well as the interaction term become significant. For those who guessed valuation inputs play no statistically significant role in the valuation process. But, for those who calculated, the interaction with life expectancy term's negative sign shows that such participants not only use a formula that is similar to an actuarial valuation but integrate valuation inputs in an economical meaningful way. The higher a participant estimated life expectancy to be the lower his or her estimate for annuity rate becomes as monthly payments have to last for longer periods on average. This result provides further evidence that participants did not only claim using a particular method but most likely also applied it. Unsurprisingly, the interaction term with the interest rate is not significant as already our analyses of valuation approaches found that participants did not take the interest rate (i.e., compounding or discounting) into account.

Finally, we analyze how valuation results' dispersion around an objective benchmark, that is, their precision, relates to the valuation approach used (guessing, calculating, or division rule). Based on univariate analyzes we find that the more sophisticated the approach used by the participant was the more precise estimates becomes (lower standard deviation) and the closer

[9] When including those three variables separately in Model 3 each coefficient is statistically significant at the 5% (subjective and objective numeracy) or 1% level (financial literacy).
[10] In Models 4 – 6 we use as the measure of life expectancy participants' estimate for females. Results hold as well if we include male life expectancy (available on request).

estimates become to market rates (at the subsample median) (see Table 4). Likewise, the fraction of participants that reports an estimate within a range of 2 percentage points around the market rate of 6.7% increases with sophistication of the valuation method.[11]

Next, using a logit model we regress a measure of precision, that is, a dummy variable of having estimated an annuity rate within a range 2 percentage points around the market rate (=1) or not (=0) on participant characteristics and the method reported. Results in Table 6 show that demographic characteristics do not explain precision. The three sophistication measures individually (Columns 1-3) positively predict precision, while when included jointly because of their correlation (Column 4) they are not significant. Having calculated or not is in no model significantly related to precision (Columns 1-4).

It seems that having applied calculation does explain the valuation results directionally (see Table 5) but not necessarily the quality of the outcome in terms of precision. These results hint at that it is not mathematical reasoning per se and using the right formula (e.g., as in Foltice and Langer, 2017) but also the right ingredients in a formula to arrive at a precise result. So does financial sophistication in combination with calculating lead to more precision? To investigate this possibility, we interact sophistication measures with having calculated or not. For doing that we select financial literacy, as among the three measures its coefficient's *p*-value is the largest (p=0.025, Column 3) and interact it with having calculated or not. Moreover, to facilitate calculation and interpretation of interaction terms we estimate the interaction models with OLS.

First, Column 5 reproduces the corresponding logit model of Column 3 with results being consistent. Column 6 then shows that based on the significant interaction term that the combination of calculating and having high financial literacy increases precision. In other words, being smart enough to use a formula and to understand how to use it is required to derive precise valuation results.

Columns 7 and 8 show corresponding results when we use instead of having calculated per se the indicator variable for having used a formula that is close to the actuarial formula, the division rule. In this specification, the dummy variable for having used such a formula is itself already significant – consistent with the results in section 3.4.2.1 where we found that that given having used calculation the use of such a formula is already an indicator of a higher level

[11] This result holds if we use instead of the average male-female annuity rate male or female annuity rates as well as an interval range of 1 percentage point.

of sophistication and thus potentially leading to better valuation results. Indeed, the interaction (Column 8) of having used the advanced formula and financial literacy is now not significant anymore.

- Table 6 here -

3.4.2.3 Robustness checks

3.4.2.3.1 Experimental demand effects

Based on the coder classification as well as the self-classification of participants we check if our survey created experimental demand effects in that if participants who saw immediately a requirement to report on an approach were triggered to report or use more sophisticated approaches than they actually used or intended to use. For doing that, we compare means and medians of a number variables between those participants who saw the text entry field on the same screen as the task and those who only saw it at the following screen. The latter group thus did not know when solving the task that they would be asked to indicate how they did it. Within each valuation task condition (lump-sum or payout) we compare the following variables: valuation task duration in seconds, text length in characters, percentage being coder-classified having used a calculation, percentage self-classified having used a calculation, and annuity payouts and lump-sums estimated. We do not find significant differences in those variables. These results demonstrate that regardless of whether participants were aware that they need to report their approach or not they used (on average) the same approach.

3.4.2.4.2 Guessing or a lack of motivation?

There are multiple reasons why roughly 60% of participants might have indicated that they guessed the answer. Those participants might have guessed because they truly had no clue; they might have had a good intuition and thus not felt a need to perform a calculation in the first place or they simply were not motivated to exert effort when filling in the survey and just opted for the quickest option to complete the survey.

Having a good intuition is not supported by our results. Annuity rates of those who guess are further away from realistic values than of those who calculate as well annuity rates are more dispersed (Table 4). To test for a lack of motivation we analyze the time spent on the task, as well as on the total survey. Participants who guessed spent on average 2.4 minutes on the task (that is providing a valuation result and explaining their approach) which is 2.5 minutes less than other participants. This difference is statistically significant (p=0.011). The total time to fill in the survey (excluding the time worked on the task) was for those who guessed 11.0 minutes, which is 1.3 minutes less than other participants. This difference is not statistically significant (p=0.205). That is, while guessing resulted in a shorter time worked on the task, those participants did not spend significantly less time on filling in the complete survey (which included several tedious sections, like five financial literacy questions and eight in part difficult objective numeracy questions). So overall, indicating to having guessed does not seem to reflect a lack of motivation, having a good intuition but rather having truly guessed. In addition, the text responses of those who self-classified their approach as guessing point in the same direction – they are only very rarely filled with non-meaningful text (e.g., “no thoughts”, “Common sense”) but on the contrary typically openly admit guessing (e.g., “I was just guessing”, “I guessed honestly”, “Truthfully all I did was guess”).

4 Developing and Testing an Intervention – Study 2

4.1 Purpose

Study 1 reveals that a subset of participants uses calculation to value annuities. Those participants are characterized by having higher financial literacy and numeracy. They do not only report to use a calculation – their annuity valuation results are significantly and in the right direction related to valuation inputs used in the formula (life expectancy) and to higher valuation precision.

The question we address in Study 2 is: Based on the knowledge of a typical valuation approach (the Division Rule) – can we design an intervention that matches the formula and by doing that first influencing valuation outcomes and in consequence increasing participants’ interest in buying an annuity, that is, change their behavior (RQ4)? While Study 1 provided statistically significant evidence that the self-reported formula is used to value annuities, the evidence is not causal. Potentially unobserved variables could have driven our results. That is, we lack causal evidence on whether participants use calculation to value annuities.

4.2 Development of the intervention and hypotheses

The most widely used approach we identified in Study 1 is the Division Rule. That is, when starting from a lump-sum, participants first divide that number by an estimate of life expectancy at age 65 and then divide that number by 12 to arrive at monthly payouts. In case of a payout given, the rule is reversed, in that first monthly payouts are multiplied by 12 and then by the estimate of life expectancy.

As participants ignored the interest rate in their valuations the parameter in their formula we should target to influence is life expectancy at age 65 (compare Formula (1b)). That is, a potential intervention to change annuity valuations and thereby influence demand could be to prime participants with information they use in the formula – life expectancy at age 65.

Doing so poses, however, several challenges. First, a natural candidate for a life expectancy prime could be to make this number salient by including it in the text prompting for an annuity valuation. Study 1, however, found that participants were on average in their assessment of life expectancy already close to realistic estimates of life expectancy. By using such a prime we do not expect an impact on annuity valuations on average. At best, we expect a reduction in the dispersion of estimates.

Second, in both the lump-sum as well the payout conditions participants' valuations were unrealistic in that they expected annuities to deliver too low payouts (lump-sum treatment) or to be too cheap (payout treatment) compared to what the annuity market does provide. In both conditions, an intervention that would make annuities look more attractive would be priming participants with a lower number than actual life expectancy. We expect that in a lump-sum condition, expected payouts will increase, while in a payout condition, expected annuity prices will decrease. Thus, in both cases an annuity would look more attractive. But, while such an intervention may yield greater interest in annuities, it is not likely to be a feasible and successful intervention in practice. Unrealistic valuations generated may increase initial interest in annuities. However, as soon as participants receive feedback on actual annuity payouts and prices (e.g., when looking up an annuity rates on a website) they will experience a negative surprise in that actual annuity rates are worse than they expected. In consequence, such a prime is not very likely to increase annuity demand either.

Alternatively, we could prime participants with a higher number than actual life expectancy. Doing so, we expect will result in more realistic valuations, that is, bring expected payouts and lump-sums closer to market values. But, in that case participants are likely to view annuities as even more expensive (high lump-sums needed) or as yielding even lower returns (low payouts expected), with potentially negative effects for demand.[12]

As a solution to the challenge of designing a priming intervention we propose to combine priming with immediate feedback on market rates. That is, first we prime a participant with a life expectancy number to improve the realism of the valuation and then in a second step give immediate feedback on market rates – thereby reducing a potentially negative surprise when confronted with market rates or even generating a positive surprise. Thus, in turn we expect this combination to increase annuity demand.

To test the impact of priming and providing feedback we use a two-by-two design. That is, we randomly assign participants to one of the following four conditions: control – no feedback given; control – feedback given; number prime – no feedback given; number prime – feedback given. We expect:

H1: Compared to a control condition without a prime, priming subjects with in an annuity valuation task with a higher number than actual life expectancy will result in lower estimated annuity rates.

H2: Giving feedback on market annuity rates increases annuity demand irrespective of a prime.

H3: Compared to a control group with no prime given, the effect of giving feedback on annuity demand is stronger when subjects are primed with a higher number than actual life expectancy.

4.3 Method and procedure

Overall, we keep the design of Study 2 as close as possible to Study 1 to isolate the effects of priming and feedback. That is, we again create a survey confronting participants with an annuity valuation task and keep the survey flow as well as the variables elicited identical to Study 1 except with the following changes to variables and flow:

[12] It is empirically well documented that people that tend to live longer purchase more annuities. It is not clear, however, whether those people buy annuities because they expect to live longer (active selection) or other variables (correlated with life expectancy) like higher wealth or income increase annuity demand (passive selection) (see, e.g., Bucher-Koenen and Kluth 2012; Cannon and Tonks 2016).

- We include in Study 2 only conditions where a lump-sum is given. Our discussion in the last section highlighted, we expect the same effects regardless of a scenario with lump-sums or payouts given.
- We remove the text elicitation question for describing the valuation approach as self-classification in Study 1 proved to be a reliable technique. In addition, we remove the "10 rules" from the self-classification section as almost no participant indicated having used them.
- After a participant has provided his or her estimate for the annuity payout in two out of four groups feedback on market rates is given (see the details below). Right after that, annuity demand is elicited by asking "In general, how likely is it that you will be buying an annuity? (select 7 if you already own an annuity)" measured on a 7-point scale from 1 (Extremely unlikely) to 7 (Extremely likely).

In the valuation task, we implement priming with a number higher than actual life expectancy through adding the following sentence to the text all participants saw: "Note, 25 percent of the U.S. population live up to age 90 (that is another 25 years after age 65). The numbers used in the prime were based on estimates of the distribution of number of years to live after age 65 (average of male and female) using U.S. mortality data (Milevsky, 2019). Thus, participants were not deceived.

Feedback is implemented for 50% of participants in each of the conditions (control and prime) right before the annuity demand question in the following way: "You estimated that a person aged 65 will receive a lifetime payout of $ <*own estimate shown here*> per month if s/he has saved $500,000. Currently, U.S. insurance companies offer lifetime payouts of about $2,800 per month for a person that has saved $500,000."

Finally, we implement three more additions to test in robustness checks if the prime might affect annuity demand though alternative channels (other than through the formula). First, we elicit participants' confidence in their numerical estimate right after the valuation task based on the scale of Gamble et al. (2015). Providing life expectancy as a valuation input might potentially increase confidence in the estimate derived. Being more confident could change participants' annuity demand (see, e.g., Ben-David et al., 2018). Second, after the annuity demand question we elicit mortality salience with the scale used in Salisbury and Nenkov (2016). Providing information on how long people live might increase mortality salience and thus in turn reduce annuity demand (Salisbury and Nenkov, 2016). Third, we include two additional auxiliary priming conditions (with or without feedback) where we provided a number prime that resembled just average life expectancy (that is, the value participants in Study 1 on average reported). Providing a number in a prime per se (irrespective of the number) might change the approach participants use to value an annuity.

4.4 Survey participants and characteristics

The online survey was distributed by Qualtrics in August 2019 to a sample with the following criteria: U.S. resident, aged 45-60 (as in Study 1) and not having participated in Study 1. We received in total 232 (+116 for the two auxiliary conditions) complete responses. The average time to fill in the survey was 12.4 minutes (median: 10.3). In terms of demographic characteristics, the sample is very similar to Study 1 with the difference that 54% were female (77% in Study 1). Again, there are no statistically significant differences in demographic characteristics between conditions. Table 7 shows descriptive statistics for the sample:

- Table 7 here -

4.5 Results

First, we test whether the priming was effective, that is, perform a manipulation check. Table 7 shows that life expectancy estimated by participants differs conditional on whether they received a prime or not. The prime increased the estimated life expectancy by 3.4 years (male) and 4.1 years (females) respectively. These increases are statistically significant (p=0.003 for male and p=0.001 for female). Thus, the manipulation was effective.

Table 7 also shows, that estimated annuity rates (again calculated according to Formula 2), differ between the control and the priming conditions. Note, in Table 7, we merge the feedback and no feedback conditions, as feedback on payouts was only given after the valuation result was entered by a participant and thus could not impact the valuation. In particular, annuity rates decreased due to the priming and became more realistic. Like in Study 1, mean annuity payouts and the corresponding rates are impacted by outliers. Therefore, the difference in mean annuity rates (18.0%) is not statistically significant (p=0.450), but the difference in medians (0.8%), which is a more robust and reliable measure in this case, is significant (p=0.009). We conclude that we find support for H1: priming subjects with a high life expectancy number reduces participants' expected annuity payouts.

Table 8 shows participants' annuity demand within the different experimental conditions.

- Table 8 here -

Comparing participants that received no feedback on market rates with participants that received feedback irrespective of the priming condition (Table 8, line: full sample) shows that giving feedback increases annuity demand by 0.64 on the 7 point scale. This difference between the conditions is statistically significant (p=0.011) and thus supports H2: Giving feedback on market rates increases annuity demand.

Table 8 also shows how giving feedback impacts annuity demand within the conditions control and prime. We find an increase in annuity demand in both conditions, with the larger increase in the prime condition (0.45 vs. 0.82). The increase is statistically significant only in the prime condition (p-value for control = 0.241 and for prime = 0.012). These results imply that we find for H2 ultimately only partial support, as in the control group the increase in annuity demand found is not statistically significant. For H3 however we find full support: giving feedback has a larger positive impact on annuity demand when participants are primed with a higher number than actual life expectancy.

4.6 Robustness checks

Presenting participants with a number prime related to life expectancy at age 65 might change annuity valuations and/or demand not only through feeding into a valuation approach. To check for alternative channels, first we analyze mortality salience as well as confidence in the value estimated (see details in Section 4.3). For both measures, we do not find significant differences between conditions.

Second, just confronting participants with numerical information about life expectancy could have triggered different approaches for estimating an annuity value. A first test on the percentage of participants that report having used a calculation shows that this is not the case. There are no significant differences between the conditions. When analyzing annuity rates between the control condition and the auxiliary condition (where we provided participants an on average uninformative prime, i.e., the average life expectancy from Study 1), we also find neither at the mean or median significant differences.

Third, we argued when developing the priming intervention, that the mechanism through which it is supposed to work is in essence to provide those who use calculation the an ingredient for their formula – the value for life expectancy. As a further test on whether the priming intervention worked through that mechanism, we analyze the dispersion in estimated annuity rates. Like in Study 1 we use the log of the annuity rate as variable of interest. Then, we calculate group specific measures of dispersion. Groups are defined along the two dimensions priming condition and having used calculation or not. In particular, we first calculate the mean of the log of the annuity rate within each group. Then we calculate for each participant the absolute distance to the group-specific mean of the log annuity rate. Table 9 shows the results.

- Table 9 here -

Results in Table 9 show, that while for those who do not calculate, the prime did not change the dispersion of estimated annuity rates, for those who calculate the dispersion decreases. The decrease is significant (p=0.021). These results are consistent with our expectation that the priming intervention provided not only information about which number to use when calculating but that the number was (on average) indeed used resulting in more similar estimates for annuity rates for those who calculate.

5 Discussion and Conclusion

We demonstrate that by using established methods from psychology, that is, the verbal protocol method, provides new and fresh insights about individuals' judgment and decision-making approaches that are reliable within sample and out of sample, and help to design effective interventions.

For the particular problem we study as an example, that is, valuing an annuity, we find that roughly 40% of participants use back of the envelop calculations, while others see to guess. Based on a number of tests, we conclude what participants tell us is a reasonable description of what they do. Using a more sophisticated approach, that is, calculation is positively related participants' sophistication (financial literacy and numeracy), predicts valuation results directionally and their precision, and a behavioral response to a priming intervention matching the valuation formula uncovered. Further tests show that experimental demand effects or lack of motivation are unlikely to explain our results.

In addition, our results uncover mechanisms that can explain puzzling results in earlier literature. Brown et al. (2017) find that valuations were closer to actuarial rates, the higher a participant's numeracy is. Our results show a mechanism behind this finding: Higher numeracy is related to a higher likelihood to apply calculation, and using calculation is related to arriving valuations that are on average closer to market rates. Likewise, Brown et al. (2017)'s conjecture that valuations results might derive from applying the heuristic: "How long will it take me to break even." Our results uncover that a central element in the valuation formula that participants use is the expected number of years an annuity is going to pay. Our finding that participants ignore the interest rate in their reasoning, that is, neither compound or discount cash flows, can explain why the variation in interest rates does not relate pension plan members' annuitization choices in Chalmers and Reuter (2012).

The difference in annuity valuations between our conditions (higher annuity rates in the payout than in the lump-sum condition) mimics the results of Binswanger and Carman (2010a) in their analysis of backward versus forward valuations of consumption streams as well Goldstein et al. (2016) analysis of annuity adequacy evaluations in lump-sum versus payout formats. Like Binswanger and Carman (2010a) we have no clear evidence to explain this result. They found some suggestive evidence that particular mistakes or loss aversion might explain some of the differences. For our data, however, these explanations even less likely to hold. A potential mistake participants could have made that Binswanger and Carman (2010a) document is to use the payout (in their case consumption) as an upper bound for the premium (in their case savings). In our data, less than 9% of respondents might have made such a mistake while in their data it is more prevalent (20%). Likewise, loss aversion is an unlikely candidate as our task referred to think about a third person, which makes it less likely to trigger strong feelings of gains and losses. The ignorance of discounting and compounding found for our respondents as well cannot explain the difference, as ignoring interest has the same effect in both conditions (mathematically) on annuity rates estimated. Potential other explanations, that we cannot test with our data are, for example, that both conditions differ in terms of dealing with rather small (when starting from payouts) or large numbers (as also Goldstein et al. 2016 conjecture), and require multiplication (when starting from payouts) or division. There is evidence, that generally, small versus large numbers, and multiplication versus division create a different level of difficulty and are processed differently in the brain (e.g., Campbell, 1995; LeFevre and Morris, 1999; Dehaene et al., 2008; Hyde and Spelke, 2009).

However, the difference in valuations between the conditions helps to understand widely documented low annuity demand: Participants who approach retirement planning in terms of a desired payout envision too high annuity rates than the market does deliver. Thus, they might initially view annuities positively (cheap) when thinking about them but will be negatively surprised (e.g., when checking rates on a website) when confronted with market lump-sums needed to reach the desired payout. Thus, they shy away from annuities. Participants who approach retirement planning in terms of how much an accumulated nest egg (savings) will deliver might view already initially annuities as poor products that deliver low payouts. For that reason, they might not even check out markets rates and as well shy away from annuities.

Our test of priming intervention reveals why moving methodologically forward in studying individual decision-making approaches is important beyond scientific curiosity and theory building but also for policy-makers. There is an ongoing discussion about which methods are effective for improving household financial decision-making. Initiatives to improving financial literacy seem to often have limited success (Fernandes et al. 2014). The effectiveness of debiasing and providing decision support seem to crucially depend decision-makers capabilities and (potentially biased) awareness of their skills (Foltice and Langer 2018; Cordes et al. 2019). And, while there are many examples of successful nudging interventions, numerous nudges have been shown to be ineffective or even resulting in worse outcomes (see, e.g., Sunstein, 2017). Our results demonstrate, that investigating judgement and decision-making at a very basic level helps to improving and deepening our understanding of how individuals approach such decisions and therefore provides a sound basis for designing effective interventions or to opt for other strategies.

In relation to established methods in household finance used, we think the verbal protocol method complements but does not replace them, and surely has its limitations. In our analyses of text responses (Study 1) we study a small sample of 229 participants. Surely, we are not envisioning to roll out human coder based text analysis to large samples. But, we demonstrated that a combination of a small sample initial study (a pretest in our case) to generate approach classification candidates with a larger study where participants self-classify their approaches can give reliable results.

For future research, we envisioning testing different financial decisions or samples, for example, to understand if for different decisions it is as well possible to identify new approaches easily. Moreover, while for 40% of our participants our method delivers insights that can be interpreted right away, for the other 60% more work and methodological progress is needed to understand what guessing really means. In addition, based on existing evidence on processing of numbers (e.g., LeFevre et al. 1993; Campbell, 1995; Hyde and Spelke, 2009) more tests are needed to understand how the numbers presented in decision-making scenario per se (e.g., small or large ones, round or non-round, …) are related to triggering certain decision-making approaches. Likewise, differences in culture or mathematics education at school might result in different approaches used by individuals (e.g., Foltice and Langer 2017, 2018) and in consequence require different interventions.

References

Bateman, H., Eckert C., Iskhakov, F., Louviere, J., Satchell, S., Thorp, S. (2017), Default and naive diversification heuristics in annuity choice, *Australian Journal of Management*, 42: 32-57.

Bateman, H. Stevens, R., Alonso Garcia, J. and Ponds, E. (2018), Learning to Value Annuities: The Role of Information and Engagement, Working Paper.

Ben-David, I., Fermand, E., Kuhnen, C. M., Li, G. (2018), Expectations Uncertainty and Household Economic Behavior, NBER Working Paper No. 25336.

Binswanger, J., Carman, K. G. (2010a), The Miracle of Compound Interest: Does Our Intuition Fail?. CentER Discussion Paper Series No. 2010-137.

Binswanger, J., Carman, K. G. (2010b), The Role of Decision Making Processes in the Correlation Between Wealth and Health. Netspar Discussion Paper No. 12/2010-081.

Binswanger, J., Carman, K. G. (2012), How real people make long-term decisions: The case of retirement preparation, *Journal of Economic Behavior & Organization*, 81: 39-60.

Brown, J. R., Kapteyn, A., Luttmer, E. F., Mitchell, O. S. (2017), Cognitive constraints on valuing annuities. *Journal of the European Economic Association*, 15, 429-462.

Brown, J. R., Kling, J. R., Mullainathan, S, Wrobel, M. V. (2008), Why Don't People Insure Late Life Consumption: A Framing Explanation of the Under-Annuitization Puzzle. *American Economic Review*. 98: 304-309.

Bruine de Bruin, W., Parker, A. M., Fischhoff, B. (2007), Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, 92: 938-956.

Bruine de Bruin, W., Parker, A. M., Fischhoff, B. (2012), Explaining adult age differences in decision-making competence. *Journal of Behavioral Decision-making*, 25: 352-360.

Bruine de Bruin., W., Strough, J., Parker, A. M. (2014), Getting older isn't all that bad: Better decisions and coping when facing sunk costs. *Psychology and Aging*, 29: 642-647.

Bucher-Koenen, T., Kluth, S. (2012), Subjective Life Expectancy and Private Pensions, *MEA Discussion Paper 14-2012*.

Campbell, J. I. D. (1995), Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, 1: 121-164.

Campbell, J. Y. (2006), Household Finance, *Journal of Finance*, 61: 1553-1604.

Cannon, E., Tonks, I. (2016), Cohort mortality risk or adverse selection in annuity markets?, *Journal of Public Economics*, 141: 68-81.

Chalmers, J., Reuter, J. (2012), How do retirees value life annuities? Evidence from public employees. *Review of Financial Studies*, 25: 2601-2634.

Cordes, H., Foltice, B., Langer, T. (2019), Misperception of Exponential Growth: Are People Aware of Their Errors?. *Decision Analysis*, 16:261-280.

Davidoff, T. (2009), Housing, Health, and Annuities. *Journal of Risk and Insurance*, 76: 31-52.

Davidoff, T., Brown, J. R., Diamond, P. A. (2005), Annuities and Individual Welfare, *American Economic Review*, 95: 1573-1590.

Dehaene, S., Izard, V., Spelke, E., Pica, P. (2008), Log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures, *Science*, 320: 1217-1220

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G. (2011), Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9: 522-550.

Ericson, K. M. M., White, J. M., Laibson, D., Cohen, J. D. (2015), Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better Than Delay Discounting Does, *Psychological Science*, 26: 826-833.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., Smith, D. M. (2007), Measuring numeracy without a math test: Development of the Subjective Numeracy Scale (SNS). *Medical Decision Making*, 27: 672-680.

Falk, A., Becker, A., Dohmen, T., Huffman, D., Sunde, U. (2016), The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. IZA Discussion Paper No. 9674.

Fernandes, D., Lynch Jr., J. G., Netemeyer, R. G. (2014), Financial Literacy, Financial Education, and Downstream Financial Behaviors. *Management Science*, 60, 1861-1883.

Foltice, B., Langer, T. (2017), In Equations We Trust? Formula Knowledge Effects on the Exponential Growth Bias in Household Finance Decisions. *Decision Analysis*, 14:170-186.

Foltice, B., Langer, T. (2018), Exponential growth bias matters: Evidence and implications for financial decision making of college students in the U.S.A. *Journal of Behavioral and Experimental Finance*, 19: 56-63.

Gamble, K. J., Boyle, P. A., Yu, L., Bennett, D. A. (2015), Aging and financial decision making, *Management Science*, 61: 2603-2610.

Goldstein, D. G., Hershfield, H. E., Benartzi, S. (2016), The Illusion of Wealth and Its Reversal. *Journal of Marketing Research*, 53, 804-813.

Goldstein, D. G., Johnson, E. J., Sharpe, W. F. (2008), Choosing outcomes versus choosing products: Consumer-focused retirement investment advice, *Journal of Consumer Research*, 35: 440-456.

Hogan, T. P., Brezinski, K. L. (2003), Quantitative Estimation: One, Two, or Three Abilities?, *Mathematical Thinking and Learning*, 5: 259-280.

Hyde, D. C., Spelke, E. S. (2009), All numbers are not equal: an electrophysiological investigation of small and large number representations, *Journal of Cognitive Neuroscience*, 21: 1039-1053.

Inkmann, J., Lopes, P., Michaelides, A. (2011), How Deep Is the Annuity Market Participation Puzzle?, *Review of Financial Studies*, 24: 279-319.

Johnson, E. J., Häubl, G., Keinan, A. (2007), Aspects of endowment: a query theory of value construction. *Journal of Experimental Social Psychology: Learning, Memory and Cognition*, 33: 461–474.

Kaplanski, G., Levy, H., Veld, C., Veld-Merkoulova, Y. (2015), Do happy people make optimistic investors?. *Journal of Financial and Quantitative Analysis*, 50: 145-168.

Kuhnen, C. M., and Melzer, B. (2018), Noncognitive abilities and financial delinquency: the role of self-efficacy in avoiding financial distress, *Journal of Finance*, 73: 2837-2869.

Kuusela H, Paul P. (2000), A comparison of concurrent and retrospective verbal protocol analysis, *American Journal of Psychology*, 113: 387-404.

LeFevre, J., Greenham, S., Waheed, N. (1993), The Development of Procedural and Conceptual Knowledge in Computational Estimation. *Cognition and Instruction*, 11: 95-132.

LeFevre, J., Morris, J. (1999), More on the relation between division and multiplication in simple arithmetic: evidence for mediation of division solutions via multiplication. *Memory & Cognition*, 27: 803-812.

Lockwood, L. M. (2012), Bequest motives and the annuity puzzle, *Review of Economic Dynamics*, 15: 226-243.

McGowan, F., Lunn, P. D., Robertson, D. A. (2018), The Framing of Options for Retirement: Experimental Tests for Policy, working paper.

Milevsky, M. A. (2019), Calibrating Gompertz in Reverse: Mortality-adjusted Biological Ages around the World, Working Paper.

Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., Brown, J. R. (1999), New Evidence on the Money’s Worth of Individual Annuities, *American Economic Review*, 89: 1299–1318.

Poterba, J. M. (2006), Annuity markets. In G. L. Clark, A. H. Munnell, & J. M. Orszag (Eds.), *The Oxford handbook of pensions and retirement income*, (pp. 562-583). Oxford: Oxford University Press.

Puri, M., Robinson, D. T. (2007), Optimism and economic choice, *Journal of Financial Economics*, 86: 71-99.

Salisbury, L. C., Nenkov, G. Y. (2016), Solving the annuity puzzle: The role of mortality salience in retirement savings decumulation decisions, *Journal of Consumer Psychology*, 26: 417-425.

Schulze, R. N., Post, T. (2010), Individual Annuity Demand under Aggregate Mortality Risk, *Journal of Risk and Insurance*, 77: 423-449.

Shu, S. B., Zeithammer, R., Payne, J. W. (2016), Consumer Preferences for Annuity Attributes: Beyond Net Present Value, *Journal of Marketing Research*, 53: 240-262.

Siegler, R., Opfer, J. (2003), The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science*, 14: 237-243.

Sinayev, A. and Peters, E. (2015), Cognitive reflection vs. calculation in decision making, *Frontiers in Psychology*, 6: 1-16.

Sunstein, C. R. (2017), Nudges that fail, *Behavioural Public Policy*, 1, 4-25.

Van Rooij, M., Lusardi, A., Alessie, R. (2011), Financial literacy and stock market participation, *Journal of Financial Economics*, 101: 449-472.

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., Goldstein, D. G. (2007), Asymmetric discounting in intertemporal choice: a query theory account, *Psychological Science*, 18, 516–523.

Weller, J. A., Dieckmann, N. F., Tusler, M. , Mertz, C. K., Burns, W. J., Peters, E. (2013), Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach. *Journal of Behavioral Decision Making*, 26: 198-212.

Yaari, M. E. (1965), Uncertain Lifetime, Life Insurance, and the Theory of the Consumer, *Review of Economic Studies*, 32: 137-150.

Ye, Z., Post, T. (2019), What age do you feel? –Subjective age identity and economic behaviors, *Journal of Economic Behavior and Organization* (forthcoming).

Tables

Table 1: Descriptive Statistics Study 1

	Condition								
	Full sample (N=229)			Payout given (N=117)			Lump-Sum given (N=112)		
Variable	Mean	Median	Std.	Mean	Median	Std.	Mean	Median	Std.
Age (years)	52.03	52.00	4.38	52.21	53.00	4.48	51.85	52.00	4.28
Gender (1=female)	0.77	1.00	0.42	0.79	1.00	0.41	0.74	1.00	0.44
Education high (>median)	0.36	0.00	0.48	0.41	0.00	0.49	0.31	0.00	0.47
Savings in $'000	74.64	5.00	165.24	89.08	5.00	190.87	59.55	5.00	132.52
Income in $'000	46.14	35.00	38.55	46.54	35.00	41.41	45.71	35.00	35.51
Risk tolerance (1-10)	5.66	6.00	2.40	5.59	6.00	2.43	5.73	5.00	2.37
Time preference (1-10)	6.30	6.00	2.35	6.49	7.00	2.43	6.10	6.00	2.26
Survey duration (seconds)	897.70	695.00	688.34	875.19	717.00	538.10	921.21	669.00	818.25
Text lenght (characters)	68.63	40.00	75.36	73.25	49.00	77.78	63.80	34.50	72.79
Subj. Numeracy (1-6)	3.73	4.00	1.51	3.78	4.00	1.44	3.68	3.75	1.59
Obj. Numeracy (1-8)	2.52	2.00	1.69	2.58	2.00	1.71	2.46	2.00	1.68
Financial Literacy (1-5)	2.61	3.00	1.44	2.61	3.00	1.48	2.62	3.00	1.40
Exp. years to live 65+, male	17.93	15.00	8.85	17.60	17.00	7.68	18.27	15.00	9.94
Exp. years to live 65+, female	21.59	20.00	9.05	20.99	20.00	7.92	22.21	20.00	10.09
Interest rate on 10 year T-Bond (%)	1.55	1.00	1.64	1.47	1.00	1.73	1.63	1.25	1.53

Notes: This table shows descriptive statistics for the sample collected for Study 1 for the full sample as well as separately for the two valuation conditions (payout given and lump-sum given).). Education is a dummy variable where a 1 indicates a level above the sample median ("Some college but no degree") and 0 otherwise. Text length is the length in characters that a participant entered to describe his or her approach; subjective numeracy is the average score over the 4 construct items; objective numeracy is the number of correct answers over the 8 construct questions; financial literacy is the number of correct answers over the 5 construct question.

Table 2: Valuation Approaches in Study 1

General approach	Text entry example	Fraction %			Calculation approach	Fraction %		
		Full sample	Payout	Lump-sum		Full sample	Payout	Lump-sum
Calculation	"Took the 500000 and divided by 20 assuming the person lives 20 years then divided it by 12"	41.92	44.45	39.29	Rule 10	2.18	1.71	2.68
					Rule 10 and adjustments	2.62	5.13	0.00
					Rule 12	14.85	12.82	16.96
					Rule 12 and adjustments	6.99	8.55	5.36
					No option fits	15.28	16.24	14.29
Guessing	"A complete guess"	49.34	47.87	50.9				
Using the internet	"Google searched the question"	1.75	0.85	2.68				
No option fits	"People live longer"	6.99	6.84	7.14				

Notes: This table shows the distribution of valuation approaches according to participants' self-classification for the full sample and by conditions. The entries "Rule…" refer to participants' indicating that they divided the lump-sum by the number of years to live and then by 10 or 12 (with potential adjustments) in case of the lump-sum condition. In the payout condition the entries refer to methods that multiplied the payout by 10 or 12 and the number of years to live. Fraction is the percentage of participants using an approach relative to the full sample or participants in each condition.

Table 3: Characteristics of Participants that Calculate

	(1) Calculated	(2) Calculated	(3) Calculated	(4) Calculated	(5) Calculated	(6) Division
Age	-0.010	-0.014*	-0.009	-0.012	-0.013*	-0.023**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)
Gender	-0.028	0.003	0.005	-0.005	0.022	0.218**
	(0.076)	(0.073)	(0.076)	(0.077)	(0.074)	(0.104)
Education high	0.121*	0.039	0.102	0.106	0.036	0.147
	(0.068)	(0.068)	(0.068)	(0.068)	(0.067)	(0.093)
Income in $'000	0.002*	0.001	0.001	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Savings in $'000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Risk tolerance	-0.001	-0.008	-0.000	-0.003	-0.007	0.014
	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)	(0.019)
Time preference	0.012	0.001	0.007	0.011	-0.000	0.033*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.020)
Lump-sum condition	-0.045	-0.050	-0.043	-0.047	-0.048	-0.067
	(0.064)	(0.061)	(0.063)	(0.063)	(0.061)	(0.089)
Subj. Numeracy		0.104***			0.093***	0.073**
		(0.020)			(0.022)	(0.035)
Obj. Numeracy			0.051***		0.025	-0.084***
			(0.019)		(0.022)	(0.032)
Financial Literacy				0.052**	0.009	0.145***
				(0.022)	(0.026)	(0.039)
Observations	229	229	229	229	229	96
Pseudo R-squared	0.039	0.107	0.061	0.055	0.113	0.242

Notes: This table presents the marginal effects from logistic regressions of a having applied a calculation (0 = not applied, 1 = applied) in Models 1-5 and having applied the division rule (0 = not applied, 1 = applied) given that a calculation was used in Model 6 on different sets of explanatory variables. Gender is a dummy variable (0 = male, 1 = female). Education is a dummy variable where a 1 indicates a level above the sample median ("Some college but no degree") and 0 otherwise. Lump-sum condition is 0 for the payout condition and 1 for the lump-sum condition. Reported are marginal effects at means of independent continuous and discrete dummy variables. Standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

Table 4: Annuity Rates and Distance from Market Benchmark

		Condition: payout given			
Variable		All participants	No calculation	Any calculation	Division rule
Annuity rate	Mean	45.64	58.27	29.85	21.47
	Median	0.11	0.22	0.07	0.06
	Std	208.70	258.27	122.08	98.65
Correct ± 2%	Fraction	0.19	0.15	0.23	0.33
Observations		117	65	52	33
		Condition: lump-sum given			
		All participants	No calculation	Any calculation	Division rule
Annuity rate	Mean	1.16	1.84	0.11	0.09
	Median	0.05	0.04	0.05	0.05
	Std	11.33	14.54	0.20	0.18
Correct ± 2%	Fraction	0.35	0.29	0.43	0.46
Observations		112	68	44	28

Notes: This table shows descriptive statistics for the annuity rates that participants estimated in the two valuation conditions as well as the fraction of participants whose estimates were within a 2 percentage point range around the annuity market benchmark rate of 6.7%.

Table 5: Explaining the Valuation Result

	(1)	(2)	(3)	(4)	(5)	(6)
	Annuity rate	Annuity rate	Annuity rate	Annuity rate	Annuity rate	Annuity rate
Gender	-0.069		-0.240		-0.286	-0.171
	(0.339)		(0.339)		(0.340)	(0.343)
Age	-0.035		-0.027		-0.029	-0.024
	(0.033)		(0.033)		(0.034)	(0.034)
Education high	-0.633**		-0.450		-0.425	-0.413
	(0.314)		(0.319)		(0.322)	(0.322)
Income in $'000	-0.009**		-0.008*		-0.008*	-0.008*
	(0.005)		(0.005)		(0.005)	(0.005)
Savings in $'000	-0.000		0.000		0.000	0.000
	(0.001)		(0.001)		(0.001)	(0.001)
Risk tolerance	-0.003		0.013		0.001	-0.017
	(0.065)		(0.064)		(0.065)	(0.065)
Time preference	0.001		0.025		0.038	0.054
	(0.067)		(0.067)		(0.068)	(0.068)
Lump-sum condition	-1.736***	-1.674***	-1.730***	-1.626***	-1.680***	-1.574***
	(0.286)	(0.278)	(0.281)	(0.288)	(0.282)	(0.284)
Subj. Numeracy		-0.182*	-0.113		-0.143	-0.131
		(0.102)	(0.110)		(0.113)	(0.112)
Obj. Numeracy		-0.107	-0.089		-0.106	-0.082
		(0.098)	(0.102)		(0.103)	(0.103)
Financial Literacy		-0.206*	-0.200*		-0.198*	-0.223*
		(0.117)	(0.118)		(0.118)	(0.118)
Exp. years to live 65+, female				-0.001	-0.006	0.020
				(0.016)	(0.016)	(0.020)
Interest rate on 10 year T-Bond (%)				-0.137	-0.126	-0.116
				(0.090)	(0.089)	(0.107)
Calculated				-0.080	0.330	1.994**
				(0.292)	(0.304)	(0.793)
Calculated x exp. years to live 65+						-0.077**
						(0.035)
Calculated x interest rate						0.000
						(0.190)
Constant	1.191	0.064	1.620	-1.176***	2.111	1.088
	(1.836)	(0.420)	(1.831)	(0.414)	(1.905)	(1.951)
Observations	229	229	229	229	229	229
Adjusted R-squared	0.154	0.178	0.181	0.121	0.184	0.196

Notes: This table presents the results from OLS regressions of the log of the annuity rate on different sets of explanatory variables. Gender is a dummy variable (0 = male, 1 = female).). Education is a dummy variable where a 1 indicates a level above the sample median ("Some college but no degree") and 0 otherwise. Lump-sum condition is 0 for the payout condition and 1 for the lump-sum condition. Calculated is a dummy variable (0 = not applied, 1= applied). Standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

Table 6: Explaining Precision

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Correct ± 2%	Correct ± 2%	Correct ± 2%	Correct ± 2%	Correct ± 2%	Correct ± 2%	Correct ± 2%	Correct ± 2%
Gender	0.057	0.079	0.078	0.089	0.067	0.055	0.054	0.056
	(0.070)	(0.071)	(0.071)	(0.071)	(0.070)	(0.070)	(0.070)	(0.071)
Age	-0.001	0.001	-0.002	-0.001	-0.001	-0.003	-0.001	-0.001
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Education high	-0.023	-0.007	-0.009	-0.023	-0.014	-0.011	-0.025	-0.026
	(0.065)	(0.063)	(0.063)	(0.065)	(0.066)	(0.065)	(0.066)	(0.066)
Income in $'000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Savings in $'000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Risk tolerance	0.010	0.012	0.009	0.010	0.008	0.009	0.007	0.007
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
Time preference	0.001	-0.000	0.004	0.000	0.003	0.003	0.003	0.004
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Lump-sum condition	0.161***	0.164***	0.161***	0.162***	0.161***	0.128**	0.160***	0.156***
	(0.055)	(0.055)	(0.055)	(0.055)	(0.059)	(0.059)	(0.058)	(0.060)
Subj. Numeracy	0.032			0.016				
	(0.021)			(0.023)				
Obj. Numeracy		0.039**		0.022				
		(0.017)		(0.020)				
Financial Literacy			0.050**	0.032	0.046**	-0.002	0.040*	0.033
			(0.021)	(0.025)	(0.021)	(0.027)	(0.021)	(0.023)
Exp. years to live 65+, female	-0.002	-0.001	-0.002	-0.001	-0.002	-0.005	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Interest rate on 10 year T-Bond (%)	0.009	0.009	0.009	0.008	0.009	-0.003	0.006	0.010
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.022)	(0.018)	(0.020)
Calculated	0.071	0.074	0.073	0.056	0.079	-0.489**		
	(0.060)	(0.058)	(0.058)	(0.060)	(0.061)	(0.199)		
Calculated x exp. years to live 65+						0.011		
						(0.007)		
Calculated x interest rate						0.018		
						(0.039)		
Calculated x fin. literacy						0.114***		
						(0.041)		
Division rule							0.143**	0.110
							(0.069)	(0.122)
Division x exp. years to live 65+								-0.002
								(0.008)
Division x interest rate								-0.022
								(0.049)
Division x fin. literacy								0.042
								(0.046)
Constant					-0.005	0.309	0.009	-0.006
					(0.392)	(0.401)	(0.387)	(0.389)
Observations	229	229	229	229	229	229	229	229
Pseudo R-squared	0.063	0.073	0.075	0.082				
Adjusted R-squared					0.031	0.064	0.042	0.034

Notes: This table presents in Models 1-4 the marginal effects from logistic regressions of having estimated an annuity rate within an interval of ± 2 percentage points around the market benchmark rate (=1) or not (0) on a having applied a calculation (0 = not applied, 1 = applied) on different sets of explanatory variables. Gender is a dummy variable (0 = male, 1 = female).). Education is a dummy variable where a 1 indicates a level above the sample median ("Some college but no degree") and 0 otherwise. Lump-sum condition is 0 for the payout condition and 1 for the lump-sum condition. Reported are marginal effects at means of independent continuous and discrete dummy variables. Models 5-8 show results from corresponding models estimated with OLS. Models 7-8 use as a measure of having calculated an indicator variable for the division rule (0 = not applied, 1 = applied). Standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

Table 7: Descriptive Statistics Study 2

	Condition								
	Full sample (N=232)			Control (N=113)			Prime (N=119)		
Variable	Mean	Median	Std.	Mean	Median	Std.	Mean	Median	Std.
Age (years)	52.20	53.00	4.76	52.10	53.00	4.80	52.29	53.00	4.74
Gender (1=female)	0.54	1.00	0.50	0.52	1.00	0.50	0.56	1.00	0.50
Education high (>median)	0.52	1.00	0.50	0.53	1.00	0.50	0.50	1.00	0.50
Savings in $'000	127.74	37.50	205.66	145.35	37.50	209.01	111.01	17.50	201.89
Income in $'000	64.31	55.00	51.30	68.85	55.00	51.65	60.00	45.00	50.82
Risk tolerance (1-10)	5.88	6.00	2.47	5.92	6.00	2.63	5.83	6.00	2.33
Time preference (1-10)	6.60	7.00	2.37	6.69	7.00	2.33	6.51	7.00	2.41
Survey duration (seconds)	747.11	626.00	681.21	755.47	611.00	892.45	739.18	636.00	390.30
Subj. Numeracy (1-6)	4.05	4.25	1.53	4.12	4.50	1.52	3.98	4.00	1.54
Obj. Numeracy (1-8)	2.73	3.00	1.81	2.69	3.00	1.78	2.76	3.00	1.85
Financial Literacy (1-5)	2.96	3.00	1.39	2.89	3.00	1.35	3.03	3.00	1.43
Exp. years to live 65+, male	18.36	20.00	8.95	16.60	15.00	8.83	20.03	20.00	8.78
Exp. years to live 65+, female	22.00	20.00	9.46	19.91	20.00	9.53	23.97	25.00	8.99
Interest rate on 10 year T-Bond (%)	1.69	1.50	1.43	1.86	1.50	1.49	1.54	1.25	1.36
Annuity rate	0.25	0.04	1.80	0.34	0.05	2.32	0.16	0.04	1.10

Notes: This table shows descriptive statistics for the sample collected for Study 2 for the full sample as well as separately for the control and priming conditions. Both conditions contain participants that received feedback or did not.). Education is a dummy variable where a 1 indicates a level above the sample median (“Some college but no degree”) and 0 otherwise. Subjective numeracy is the average score over the 4 construct items; objective numeracy is the number of correct answers over the 8 construct questions; financial literacy is the number of correct answers over the 5 construct question.

Table 8: Annuity Demand

		Feedback condition			
	Priming condition	No Feedback	Feedback	Difference	N
Annuity Demand	Full Sample	3.57	4.21	0.64	232
	Control	3.67	4.13	0.45	113
	Prime	3.47	4.29	0.82	119
N		118	114		

Notes: This table shows the mean annuity demand in Study 2 for different conditions. Conditions differ by whether participants received a life expectancy prime or not during the valuation task (control, prime) and whether participants received feedback on market annuity rates before the indicated their annuity demand. Annuity demand are responses to the question “In general, how likely is it that you will be buying an annuity? (select 7 if you already own an annuity)” measured on a 7-point Likert scale from 1 (Extremely unlikely) to 7 (Extremely likely).

Table 9: Dispersion in Annuity Rates

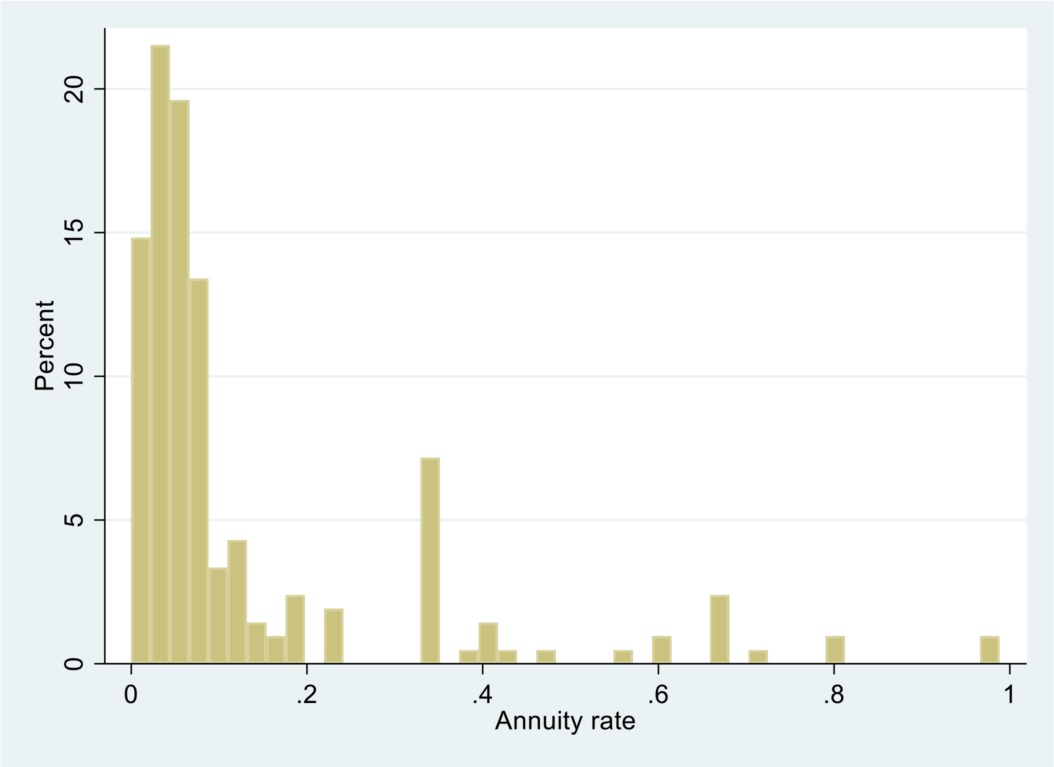
		Priming condition			
	Calculated	Control	Prime	Difference	N
Dispersion	No	0.88	0.85	-0.04	123
	Yes	1.05	0.50	-0.55	109
N		113	119		

Notes: This table shows the means of group-specific dispersion annuity rates in Study 2 for different conditions. Conditions differ by whether participants received a life expectancy prime or not during the valuation task (control, prime) and whether participants reported having calculated or not. Dispersion is calculated as the absolute distance of a participant's log of the annuity rate from the group-specific mean.

Figure 1: Word Cloud Representation of Text Responses

Notes: This figure shows a word cloud representation of text responses in Study 1. The figure was generated using Jason Davies word cloud generator (www.jasondavies.com/wordcloud/). Before processing the text to create the figure typos were corrected and abbreviations removed.

Figure 2: Distribution of Annuity Rates in Study 1



Notes: This figure shows the distribution of annuity rates in Study 1. The annuity rate is defined as the annual amount of annuity payouts (A·12) divided by the lump-sum premium P. For generating this histogram, 20 observations have been removed as annuity rates of those respondents were too large (>> 1) to allow creating a meaningful figure.

Appendix

Valuation task in Study 1

Lump-sum condition:

Please answer the following question to the best of your ability.

If someone is aged 65 and has saved $500,000,

how much of a lifetime payout per month do you think s/he will get from retirement at age 65 onward?

(That is, s/he will get a fixed payout every month for as long as s/he lives; such payout products are also called life annuities).

To come up with your answer, you may use as much time as you deem necessary.

Answer $______ per month:

Payout condition:

Please answer the following question to the best of your ability.

How much do you think someone needs to have saved at age 65 to get a lifetime payout of $2,800 per month from retirement at age 65 onward?

(That is, s/he will get a fixed payout every month for as long as s/he lives; such payout products are also called life annuities).

To come up with your answer, you may use as much time as you deem necessary.

Answer $______ of savings:

Free text heuristics elicitation prompt in Study 1

Please tell us how you came up with your answer.

It is very important to us that you report all of your thoughts that emerged when coming up with your answer.

We want to understand how you personally did it.

Valuation task in Study 2 (priming condition)

“Please answer the following question to the best of your ability.

If someone is aged 65 and has saved $500,000,

how much of a lifetime payout per month do you think s/he will get from retirement at age 65 onward?

That is, s/he will get a fixed payout every month for as long as s/he lives.

Such payout products are also called life annuities.

Note, 25 percent of the U.S. population live up to age 90 (that is another 25 years after age 65).

To come up with your answer, you may use as much time as you deem necessary.”

Answer $ per month: _________.”