Financial Literacy Center
WORKING PAPER

Do Consumers Know How to Value Annuities?

Complexity as a Barrier to Annuitization

JEFFREY R. BROWN, ARIE KAPTEYN,
ERZO F.P. LUTTMER AND OLIVIA S. MITCHELL

FLC | **Financial Literacy Center**

A joint center of the RAND Corporation, Dartmouth College,
and the Wharton School

Do Consumers Know How to Value Annuities?
Complexity as a Barrier to Annuitization

Jeffrey R. Brown, Arie Kapteyn, Erzo F.P Luttmer, and Olivia S. Mitchell

September 21, 2011

Do Consumers Know How to Value Annuities?
Complexity as a Barrier to Annuitization

Jeffrey R. Brown, Arie Kapteyn, Erzo F.P. Luttmer, and Olivia S. Mitchell

Abstract

This paper suggests that many people find the annuitization decision complex, and that this complexity, rather than a taste for lump sums, may explain observed low levels of annuity purchases. Specifically, we find that consumers tend to value annuities less when given the opportunity to buy more, but value them more highly when given the opportunity to sell annuities in exchange for a lump sum. We further show that more financially literate consumers are better able to offer responses that are consistent across alternative ways of eliciting preferences for annuitization. These results raise doubts about whether consumers can make utility-maximizing choices when confronted with the decision about whether to buy an annuity in real-world situations. In addition, we suggest that observers should be very careful in drawing conclusions about individual welfare based on observed behavior when it comes to annuities and, possibly, other complex financial products such as long-term care insurance.

Jeffrey R. Brown
Department of Finance
University of Illinois
515 E. Gregory Drive
Champaign, IL 61820
and NBER
brownjr@illinois.edu

Arie Kapteyn
RAND, Labor and Population
1776 Main Street
P.O. Box 2138
Santa Monica, CA 90407-2138
and NBER
Kapteyn@rand.org

Erzo F. P. Luttmer
Department of Economics
6106 Rockefeller Center
Dartmouth College
Hanover, NH 03755
and NBER
Erzo.FP.Luttmer@Dartmouth.edu

Olivia S. Mitchell
The Wharton School
University of Pennsylvania
3620 Locust Walk, 3000 SH-DH
Philadelphia, PA 19104
and NBER
mitchelo@wharton.upenn.edu

Do Consumers Know How to Value Annuities?
Complexity as a Barrier to Annuitization

1. Introduction

One of the enduring empirical puzzles in the economics literature is why individuals so rarely purchase annuities to insure against length-of-life uncertainty, despite the substantial value that annuities can be shown to provide in standard life cycle models. Following Yaari's (1965) seminal paper establishing conditions under which full annuitization of resources is optimal, many subsequent studies have sought to solve what has been dubbed the "annuity puzzle," which refers to the question of why few 'real world' consumers fail to annuitize their retirement wealth. This research, discussed in more detail below, explores several plausible explanations ranging from supply-side market imperfections (e.g., adverse selection, aggregate risk, or incomplete annuity markets) to rational demand-side limitations (e.g., bequest motives, the availability of formal or informal substitutes, or the presence of insured expenditure shocks). In general, however, it appears that no single factor can explain the limited demand for payout annuities; moreover, while combining many factors into one model can generate limited annuity demand, such an approach typically comes at the cost of creating new puzzles.

Of late, several research teams have begun to explore psychological barriers to annuitization in both theoretical and experimental studies.[1] In the present paper, we contribute to this nascent literature by providing evidence consistent with the hypothesis that the observed reluctance of individuals to annuitize may be the result of their difficulty making complex decisions about annuitization, rather than due to a strong preference for non-annuitized wealth. That is, the decision to annuitize is complex for two reasons. First, as discussed by Beshears et

---

[1] For a recent survey, see Benartzi et al., (2011).

al. (2008), choices with distant consequences are especially complex. Determining the optimal mix of annuitized and non-annuitized resources requires that one forecast mortality, capital market returns, inflation, future expenditures, income uncertainty, and other factors, and appropriately weigh all of this relative to one's current assessment of future preferences. Second, as also noted by Beshears et al. (2008), limited personal experience can create a wedge between revealed preferences (i.e., those that might be inferred from our action) and our true underlying preferences. Bernheim (2002) makes a related point, noting that individuals who fail to save adequately for retirement are unable to learn from experience because by the time one retires with inadequate resources it is not possible to go back to younger ages and save more. At the time of retirement, most individuals have little or no experience making annuitization decisions, let alone being able to learn from the experience of having or not having an annuity later in their own lives. While it might be possible to learn from observing the experience of others, Choi et al. (2005) show that this does not always happen: when Enron, WorldCom, and Global Crossing employees' 401(k) balances were devastated due to over-investment in their employers' stock, there was virtually no reaction by workers at other U.S. firms to reduce their own investments in employer stock.

A large literature in psychology and behavioral economics suggests that, when faced with complex decisions, people are likely to exhibit bounded rationality: they may resort to simplified decision-making heuristics, are more likely to accept default options rather than make an active choice, and are more likely to be sensitive to how decisions are framed.[2] Our central hypothesis is that many people do not fully understand the lifetime utility implications of the annuitization

---

[2] Bounded rationality is generally attributed to Simon (1947) and research on framing is linked to Kahnemann and Tversky (1981); in the annuity context, see Agnew et al. (2008); Brown et al. (2008b); and Brown et al. (2010).

decision, and therefore they have difficulty forming an appropriate assessment of the value of annuities.

This hypothesis has several implications, including that individuals (i) will be reluctant to voluntarily annuitize their saving, and (ii) will value the annuity more highly when they are already "endowed" with an annuity (such as Social Security or a defined benefit (DB) pension plan). We also hypothesize that (iii) individuals' stated preferences for annuitization will be sensitive to changes in how an annuity offer is presented, and (iv) the strength of these effects will vary with individuals' level of financial sophistication (with less sophisticated individuals exhibiting less stable valuations across a range of offers). To test these inferences, we offer evidence from a randomized experiment we have conducted using the RAND American Life Panel (ALP) wherein individuals were given hypothetical choices between various lump-sum or annuity increments (or decrements) to their Social Security benefits (which are provided in the form of an inflation-indexed annuity). For example, respondents were asked if they would prefer to keep their expected Social Security benefit streams or, instead, accept a monthly benefit permanently reduced by $100/month in exchange for a lump-sum payment. By experimentally varying the sets of choices offered, the size of the increments, the order of questions, and so on, we can trace the subjective values individuals placed on the Social Security benefit stream.

Our results indicate, first, that peoples' *average* valuations of the annuity stream are quite reasonable, as measured by their proximity to the value that would be actuarially fair (based on average population characteristics). Nevertheless, these averages still hide substantial variation. In fact, we find that a substantial minority of individuals reports objectively irrational values for the annuity; that is, their values appear to be inconsistent with any plausible set of preference parameters.

Second, we show that average annuity valuations are reduced when, instead of offering a lump sum in return for a reduced annuity, people are instead offered an opportunity to pay a lump sum to purchase additional annuity income. After ruling out liquidity constraints as the reason for this finding, we demonstrate that this decline in valuation is not due to a proportional downward shift of the full distribution of valuations. Rather, answers to these questions are *negatively* correlated at the individual level, and this pattern arises because individuals who suggest they would need to be compensated the most (i.e., receive the highest lump sum) to *reduce* their monthly annuity payment are also those who are willing to pay *the least* to receive an *increase* in their benefit. This pattern is consistent with the interpretation that such unsophisticated individuals stick with what they know (i.e., the status quo) when faced with a complex choice, unless the payoff for deviating from the status quo is extremely favorable. Moreover, we show that this within-person variance in subjective valuations is substantially smaller for people who are better-equipped to make an informed choice. For example, individuals who score higher on measures of financial literacy are far more likely to report valuations that are consistent across measures. Moreover, such individuals are most likely to be male, better-educated and from higher income households. Conversely, women, blacks, and Hispanics are least likely to score well on the consistency checks.

We next turn to an assessment of what factors are correlated with higher versus lower annuity valuations. We find that even when we focus on subsets of individuals for whom the responses to our valuation questions are most informative – people who are most financially literate, and those that give consistent answers across questions – it is difficult to explain a large share of the variation in the annuitization values. Our models account for only about 6-9% of the variance in annuity valuations, even among the most financially sophisticated, and there are few

4

systematic patterns permitting us to predict who would be most likely to value the Social Security annuity highly.

In addition to advancing our academic understanding of consumer behavior in this area, our results also have considerable practical policy relevance. Particularly in the aftermath of the financial crisis, there is an ongoing discussion of what role payout annuities should play in defined contribution (DC) or 401(k) pension plans, with active discussions about whether and how life annuities ought to be encouraged in such settings (Gale et al., 2008; Brown 2009). Numerous countries including the U.S. are grappling with fiscally unsustainable pay-as-you-go public pension systems. To the extent that households are poorly-equipped to value the annuities they have been promised from their public pensions, this can have implications for the political feasibility of reforms that change the benefit structure. The same, of course, is true with state and local public defined benefit plans in the U.S., which also face substantial underfunding problems (Novy-Marx and Rauh, 2011).

In what follows, we first summarize prior studies on the demand for annuities, focusing both on neoclassical and the behavioral economics literatures. Next we describe the American Life Panel internet survey, a roughly representative sample of the US population, and we outline our technique of eliciting lump sum versus annuity preferences. Using a randomization approach, we probe the reliability of responses and link them to key socio-demographic characteristics. After describing our experimentally-elicited annuity valuations, we relate these to complexity and show how respondent answers are shaped by anchoring and starting values, as well as two financial literacy measures which prove to be highly significant. The paper concludes with a discussion of possible policy implications and future research questions.

2. What We Know About the Annuity Puzzle

2.1 Prior Analysis of the Demand for Payout Annuities[3]

The modern economics literature on a nnuities was initiated by Yaari (1965) who developed a set of conditions under which it would be optimal for an individual to annuitize 100% of wealth.[4] This theory was extended by Davidoff et al. (2005), who showed that full annuitization will be optimal under a much more general set of conditions.[5] Recent studies have also measured how consumers value payout annuities using extended life-cycle models to compute how optimal annuitization varies with other factors, including pricing (Mitchell et al., 1999); pre-existing annuitization (Brown, 2001; Dushi and Webb, 2006); risk-sharing within families (Kotlikoff and Spivak, 1981; Brown and Poterba, 2000); uncertain health expenses (Turra and Mitchell, 2008; Sinclair and Smetters 2004; Peijnenburg et al., 2010a, 2010b), bequests (Brown 2001; Lockwood 2011); inflation (Brown et al., 2001, 2002); the option value of learning about mortality (Milevsky and Young 2007); and broader portfolio choice issues including labor income and the types of assets on offer (Inkmann et al., 2007; Koijen et al., 2007; Chai et al., 2011; Horneff et al., 2009, 2010).

Our overall assessment of the theoretical and simulation literature is that it has not been fully successful in fully resolving the annuity puzzle, even for marginal annuitization decisions (Shepard, 2011). W hile some papers have been able to simulate low overall demand for annuities (e.g., Dushi and Webb 2006; Inkmann, et al. 2007; Horneff et al. 2009, 2010), the

---

[3] Rather than providing a comprehensive review here, we instead highlight those studies most germane to the research that follows. Readers interested in the broader literature on payout annuities may consult Benartzi et al. (2011); Poterba et al. (2011); Brown (2008); Horneff et al. (2007); and Mitchell et al. (1999).

[4] His conditions included no bequest motives, time-separable utility, exponential discounting, and actuarially fair annuities (among others).

[5] Peijnenburg et al. (2010a; 2010b) also show that if agents save optimally out of annuity income, full annuitization can be optimal even in the presence of liquidity needs and precautionary motives. They further show that full annuitization is suboptimal only if agents risk substantial liquidity shocks early after annuitization and do not have liquid wealth to cover these expenses. This result is shown to be robust to the presence of significant loads.

proposed annuity puzzle "solutions" often create new puzzles. For example, studies that rely on risk-sharing within families are unable to fully explain why the demand for annuities does not rise after people transition from married life to widowhood. Studies that emphasize the lack of inflation protection or actuarially unfair pricing are unable to explain why it is so common for people to forego the opportunity to purchase higher Social Security benefits (which are inflation-indexed and priced based on average population mortality) by delaying the date of claiming.[6] Studies that emphasize the inability to access equity returns in an annuitized form are unable to explain why individuals appear reluctant to annuitize even when they can do so in the form of a variable payout annuity. As such, nearly five decades after Yaari's contribution, and nearly 25 years after Franco Modigliani (1988) noted in his Nobel acceptance speech that the absence of annuities was "ill-understood," the annuity puzzle continues to be of interest.

2.2 Empirical Evidence on Annuity Demand

Compared to the large size of the theoretical literature, the empirical literature on annuities is relatively small, mainly because the market for voluntary annuities in most countries is so small that household datasets contain too few observations on annuity purchasers. There are, however, a few notable exceptions. Using the 1992 wave of the US Health and Retirement Survey (HRS), Brown (2001) focused on respondents age 51-61 who had substantial assets in their defined contribution accounts. He examined the answer to a prospective question: "In what form do you expect to receive benefits?" and correlated their annuitization intentions with the annuity valuation predicted by a life-cycle model based on each individual's demographic characteristics. That study confirmed that, on the margin, intended annuitization was higher for those for whom the life-cycle model suggested higher valuations. But the analysis also found

---

[6] See, for instance, Brown et al. (2010) and Shepard (2011)

that it was difficult to explain more than a small fraction of the overall variation in the annuity decision.

In an interesting analysis of individuals leaving the U.S. military during the 1990s when 'separatees' were offered a ch oice between a (non-life contingent) annuity and a lump-sum payment, Warner and Pleeter (2001) found that most of the soldiers (90 percent) as well as half of the officers opted for the lump sums. Given the implicit pricing of the annuities, their actions implied that the soldiers had extraordinarily high discount rates – in excess of 17 percent (computed assuming these were fully-informed and rational decisions). A few other studies have documented high annuitization rates where most people had defined benefit (DB) plans as the status quo. For example, Hurd and Panis (2006) used five waves of the HRS (1992-2000) to explore how people made payout decisions from their defined benefit (DB) pension plans. Consistent with the hypothesis that individuals stuck with the status quo when faced with a complex decision, the authors concluded that two-thirds of retirees said they anticipated taking an annuity when given a choice to take a lump-sum distribution instead of the standard DB annuity. B enartzi et al. (2011) analyze two datasets where they had access to administrative records on retiree elections of annuities versus lump sums. In the first, they found that 88 percent of employees who retired from IBM during 2000-08 chose full annuitization, and another eight percent selected a combination of annuitization plus a lump sum. Even when they limit their sample to those age 65 and older at retirement (to ensure that the results were not driven by an overly-generous annuity to younger workers to incentivize early retirement), they found a 61 percent annuitization rate. They also examined payout patterns in 112 DB plans over the 2002-08 period, in a context where it was more difficult to measure whether a lump sum was offered. Roughly half the participants (49 percent) selected an annuity over the lump sum.

A related study by Bütler and Teppa (2007) used Swiss administrative data to track choices made by employees in 10 pension plans. When the annuity was the default option, the authors found substantial annuitization - 73 percent selected a pure annuity, with another 17 percent having partial annuitization. But in one firm providing a lump-sum option as the default, the annuitization rate was only about 10 percent. Although it is not possible to completely rule out the possibility that the firms set their default payouts to match employee preferences, the evidence is highly suggestive that the default payout option has considerable power in influencing behavior.

One of the only studies to examine plausibly exogenous variation in the price of annuities focused on Oregon public sector workers who were allowed to select between a pension life annuity versus a combination lump sum/lower "partial" monthly benefit payable for life (Chalmers and Reuter, 2009). Unexpectedly, the authors found that worker demand for partial lump-sum payouts rose, rather than fell, as the value of the forgone life annuity payments increased. When they controlled for the annuity's money's worth (measuring how close the annuity was to being actuarially fair), the demand for lump-sum payouts rose when the lump-sum payout was "large" or the incremental life annuity payment "small." The authors concluded that the decisions made in this plan were unsophisticated, so that retirees valued incremental life annuity payments at less than their expected present value, because they could not accurately value the life annuities, or perhaps because they strongly favored large lump-sum payments.

2.3 Behavioral Annuitization Studies

As noted above, our central hypothesis that the observed reluctance of individuals to annuitize may be the result of their difficulty with making complex decisions about annuitization, rather than due to a strong preference for non-annuitized wealth. There is limited

evidence on this or other behavioral explanations because the behavioral literature on annuities to date is quite small, with only a few recent papers providing evidence useful for evaluating our hypothesis. Agnew et al. (2008) and Brown et al. (2008b) show that annuity demand is sensitive to "framing." Specifically, Agnew et al. (2008) showed that men and women in an experimental setting could be 'steered' toward or away from purchasing annuities, depending on how the product was described. In the "unbiased" control, women chose the annuity 38 percent of the time relative to a 29 percent rate for men, and these gender differences persist even after controlling for financial literacy and risk aversion. When exposed to biased frames (either pro-annuity or pro-investment), men were more easily swayed than women. Specifically, men were 14 percent *less* likely to choose an annuity after the pro-investment presentation and 21 percent *more* likely to choose the annuity after the pro-annuity presentation, relative to the unbiased presentation. Women were comparably affected by the pro-investment presentation, leading to a nearly 16 percent decline in annuitization relative to the unbiased case, but the effect of the pro-annuity bias was less pronounced (about half the size of the male's response and not significantly different from zero). Brown et al. (2008b) used an internet survey that showed respondents age 50+ either a "consumption" or an "investment" frame, where the former stressed the ability to consume for life, while the latter emphasized guaranteed returns for life. In the consumption frame, the majority (70 percent) elected the annuity, whereas only 21 percent did so when shown the investment presentation. The fact that individuals were so easily swayed by relatively minor framing changes suggests that their preferences about annuities were not well defined.

Overall, we draw three lessons from the existing literature. First, the annuity puzzle cannot be easily explained within a standard neoclassical fully rational optimizing framework. In short, it is difficult to explain much of the variation in the annuity decision, suggesting that a

number of factors outside standard models may be at play. Finally, while global voluntary demand for annuities is rather limited (Mitchell et al., 2011b), annuitization rates are much higher when annuities are the default payout option Beyond this, any conclusions from the existing literature on how consumers think about this complex decision are speculative. In this paper, we use results from our own experimental study to provide further evidence regarding the complexity of decision-making when consumers contemplate protection against longevity risk.

3. Methodology and Data

3.1 The Social Security Context

In the nearly eight decades since President Franklin D. Roosevelt introduced legislation to establish the Social Security Administration, this program has become by far the largest source of lifetime income benefits for U.S. retirees and the only meaningful source of inflation-indexed annuity payments (Scheiber and Shoven, 1999). Using the Social Security benefit amount as a context for our study has several advantages. First, because of the nearly universal nature of Social Security benefits in the U.S., most workers have at least some understanding that the program pays benefits to retirees that last for as long as they live.[7] This allows us to ensure that respondents understand the nature of our "offer" (to trade off annuities and lump sums), without concern that individuals may avoid annuities simply because they do not understand how the product works. Second, this context provides a simple way to control for possible concerns about the private annuity market that might influence results, such as the lack of inflation protection (our question makes it clear that Social Security is adjusted for inflation) or concerns about counter-party risk of the insurance company providing the annuity (of course, concerns about the fiscal sustainability of Social Security means that it will be important for us to control

---

[7] See Greenwald et al. (2010), and Liebman and Luttmer (2011).

for and/or test for any effect of political risk on the valuation of the Social Security annuity, which we do below). Third, given the ongoing debate about the U.S. long-term fiscal situation, this setting is policy-relevant. For example, past discussions of possible pension reforms around the world have included proposals to partially "buy-out" benefits by issuing government bonds to workers in exchange for a reduction in their annuitized benefits.

3.2 The American Life Panel

From June to August 2011, we fielded a survey using the American Life Panel (ALP), which is a panel of U.S. households that regularly take surveys over the Internet. If at the recruiting stage, households lack internet access, they are provided this by RAND.[8] By not requiring Internet access in the recruiting stage, the ALP has the advantage relative to most other Internet panels.[9]

The American Life Panel includes about 4000 active panel members at present. Our survey was conducted over two waves of the ALP to keep the length of each questionnaire within manageable bounds. ALP participants age 18 or older were invited to take our survey. If participants indicated they did not think they would be eligible to receive Social Security benefits either on their own earnings record or that of a current, late, or former spouse, they were asked to assume for the purposes of the survey that they would receive Social Security benefits equal to the average received by people with their average age/education/sex characteristics (see Appendix A.) In all, 2210 complete responses were obtained for both wave 1 and wave 2 respondents, which comprise the sample we analyze here.

---

[8] Previously these households would receive a WebTV allowing them to access the Internet. More recently households lacking Internet access at the recruiting stage have received a laptop and broadband Internet access.
[9] We present a more detailed explanation of the ALP in the data appendix, along with a brief description of how we estimated Social Security benefits for respondents.

Table 1 compares our sample characteristics with those of the same age group in the Current Population Survey (CPS). Results indicate that our sample is not precisely nationally representative: the ALP sample is somewhat older, has slightly higher incomes, and somewhat smaller household sizes. The regional distribution is close to that of CPS, while ethnic minorities are somewhat underrepresented. The most striking difference lies in the education distribution. ALP members are clearly better educated. This means that our sample was, if anything, in a better position than a truly representative sample to provide meaningful responses to complex annuity valuation questions. Despite the statistically significant differences between the demographic characteristics between the ALP and the CPS, we note that the ALP sample contains respondents from a wide variety of backgrounds. In that sense, we think of the ALP as broadly representative of the U.S. population.

3.3 Eliciting Lump-Sum versus Annuity Preferences

To elicit preferences over annuitization, we ask respondents a number of questions of the following sort:

> *"In this question, we are going to ask you to make a choice between two money amounts.*
> *Please click on the option that you would prefer*
> *Suppose Social Security gave you a choice between:*
> *(1) Receiving your expected [current] Social Security benefit of $SSB per month.*
> *or*
> *(2) Receiving a Social Security benefit of $(SSB-X) per month <u>and</u> receiving a <u>one-time payment</u> of $LS at age Z [one year from now]."*

The variable SSB is an estimate of the individual's estimated monthly Social Security benefit; the variable LS refers to the lump-sum amount and Z is the individual's self-reported expected claiming age. Note that for those currently receiving Social Security benefits, the questions compare a change in monthly benefits to the receipt of a lump sum in one year. For those not currently receiving benefits, the trade-off is posed as a change in future monthly Social Security

benefits, in exchange for a lump sum at that person's expected claiming age. We introduced the one-year delay in the receipt of the lump-sum benefit to avoid contaminating the answers of current retirees with features of hyperbolic discounting. Before asking the annuity trade-off question, we instruct all respondents: "please assume that all amounts shown are after tax (i.e., you don't owe any tax on any of the amounts we will show you)" and "please think of any dollar amount mentioned in this survey in terms of what a dollar buys you today (because Social Security will adjust future dollar amounts for inflation)." In the trade-off question, we tell married respondents "Benefits paid to your [SPOUSE] will stay the same for either choice." Thus, so we are asking for the valuation of a single-life real annuity that has no special tax treatment.

In order to probe the reliability of the valuations provided by respondents, we varied the question in a systematic way along two dimensions. First, one can consider whether the individual is being asked to *give up* Social Security income in exchange for receiving a lump sum (which we will call "plus"), or whether the individual is asked whether he would be willing to *receive more* Social Security income in exchange for paying a lump sum (which we will call "minus"). The second dimension is whether to measure a compensating variation (CV) – the annuity / lump-sum trade that would keep them at their existing utility level – or whether to measure an equivalent variation (EV) – the lump-sum amount that would be equivalent in utility terms to the change in the monthly annuity amount.

We designate these four measures as **CV-plus** (the initial example given above), **CV-minus**, **EV-plus** and **EV-minus**. The chart below illustrates the essential differences across these four scenarios. Let SSB equal the amount of monthly Social Security benefit that the individual is currently receiving (if retired) or is expected to receive in the future (if not yet

retired). Let X be the increment (or decrement if subtracted) to this monthly Social Security benefit. Finally, let LS equal the amount of the lump sum being offered in exchange for the change in monthly benefits. In essence, this paper is about how individuals trade-off X for LS.

Four versions of the annuity valuation tradeoff question

| | "Plus"-version | | "Minus"-version | |
	Choice A	Choice B	Choice A	Choice B
Compensating Variation (CV)	[SSB-X] + LS	[SSB]	[SSB+X] - LS	[SSB]
Equivalent Variation (EV)	[SSB]+ LS	[SSB+X]	[SSB] - LS	[SSB-X]

Note: SSB stands for current/expected monthly Social Security benefits, X is the amount by which monthly Social Security benefits would change, and LS is a one-time lump-sum payment. Positive amounts are received by the individual while negative amounts indicate a payment by the individuals. Amounts between square brackets are paid monthly for as long as the individual lives, whereas LS is a one-time payment. The individual is asked to choose between Choice A and Choice B.

The CV-Plus scenario presents individuals with a choice between their current or expected Social Security benefit (SSB), and a benefit reduced by $X per month in exchange for receiving a lump sum of $LS. The CV-Minus scenario provides a choice between SSB and a benefit increased by $X in exchange for paying $LS to Social Security. The EV-Plus scenario provides a choice between receiving a higher monthly benefit (SSB+X) or receiving $SSB plus a lump sum of $LS. EV-Minus provides a choice between receiving a lower monthly benefit (SSB-X) or paying a lump sum to maintain the existing benefit. We note that one obtains the EV version from the CV version if one increases SSB in the EV version by X. In other words, the EV version is simply the CV version asked with a Social Security benefit amount that is X higher than the respondents' actual current or expected benefits level. Given that we set X at $100/month for almost all of our tradeoff questions, one would expect the difference between the EV and CV version to be small for fully-rational decision makers.

Based on the individual's answer to the first trade-off with which he is presented, we then either increase or decrease the amount of the lump-sum payment. B y walking individuals through such a "branching" process, we are able to converge on a sm all range of lump-sum values that approximate the value the individual places on the annuity stream.

Our approach to eliciting annuity values using an annuity vs. lump-sum choice in a survey setting has been recently implemented by Cappelletti et al. (2011), who used a national survey of Italian households in 2008 to ask people whether they would give up half their monthly pension income (assumed to be €1000) in exchange for a lump sum of €60,000 to be paid immediately.[10] Those who said they would not were then offered a lump sum of 80,000 euros (estimated to be roughly actuarially fair for a married male individual assuming a 3% real interest rate); those who still favored the annuity had the lump sum raised to 100,000 euros. Their results indicated that, at a price of €80,000, over two-thirds of the respondents (69 percent) favored the annuity; when the price was reduced to €60,000, annuity-favorers rise to 82 percent, and when the price of the annuity is raised to €100,000 if falls to 40 pe rcent. T hey also concluded that low-income households were less likely to find annuities attractive. While their approach is similar to ours, they did not examine how the responses varied with the specific elicitation approach, nor did they test for some of the experimental structures we do below. Two previous attempts to ask questions of this nature have also been attempted in experimental modules in the U.S. Health and Retirement Survey, but in both cases, errors in the questions or

---

[10] In the spirit of our analysis, Liebman and Luttmer (2011) report results from a 2008 survey they conducted on perceived labor supply incentives from the Social Security benefit rules. T hey also include in their survey a question asking people for the equivalent variation of a $100/month increase in their Social Security benefits (so "EV-plus" version). They find that the median 50 to 70 year-old individual values a $100/month Social Security annuity the same as a $17,500 lump-sum payment. They do not examine determinants of this valuation and moreover they do not investigate whether the valuation of a monthly benefit increase is symmetric with the valuation of a benefit reduction.

the coding of the responses prevented a full examination of the results. [11] Thus, the present study

represents the most comprehensive and in-depth attempt to elicit annuity preferences in this way.

3.4 Other Sources of Experimental Variation

In addition to presenting the trade off in the four different scenarios above, we also varied

the trade-off question along other dimensions. In particular, we randomized:

(i)   The order in which individuals were exposed to the alternative scenarios.

(ii) Whether we asked the option that mentioned the lump sum (i.e., Choice A in the chart above) as the first or the second choice.

(iii) The size of the change in the monthly benefit (X in the chart above). Specifically, we asked the CV-Plus version multiple times to each respondent: for X=$100, X=$500, for X=$SSB (so the entire amount of the respondent's Social Security benefits), and for a random X that is a multiple of $100, less than min($SSB-100, 2000), and not equal to 100 or 500. All the other versions (CV-Minus, CV-Plus, and CV-Minus) are asked once for X=$100.

(iv) The starting value of the lump-sum amount (LS in the chart above) over whether the first lump-sum amount provided was roughly actuarially fair, or whether it was lower (by 50%) or higher (by 50%) than this amount.

---

[11] Brown et al. (2008a) fielded an experimental module in the 2004 HRS asking individuals their willingness to trade $500 of a hypothetical $1000 monthly Social Security benefit for a lump sum. While the lump-sum amount offered to unmarried individuals was approximately actuarially fair, the amount offered to married couples (a majority of the sample) was far too low. We concluded that results were too noisy when singles and married individuals were separately analyzed, due to small sample sizes. A second experimental module was then designed by Brown and Mitchell and fielded in the 2008 HRS. Here, however, coding instructions provided by the HRS to field interviewers led to an inability to distinguish answers at the two extremes, i.e., those who place zero value on an annuity and those who place a very high value on annuities. Moreover, empirical results were not robust to alternative assumptions about how to treat these observations. These concerns, paired with the lack of robustness in results, lead us to view any analyses of the earlier data with suspicion, which is a key reason we undertook the present study.

In addition to the above sources of variation, which allow us to test for the reliability and stability of the elicited valuations, we also test explicitly for whether responses are sensitive to what we call "political risk," i.e., a specific form of counter-party risk that is specific to the possibility that individual valuations may be influenced by t heir subjective assessment of whether or not their future Social Security benefits would be reduced to help address the long-term financial shortfalls facing the system.

4. Initial Results: The Distribution of Annuity Valuations

4.1 CV-Plus

We begin by reporting in Figure 1 the cumulative distribution function (CDF) of the sample responses to the CV-Plus question shown above. Given our bracketing of responses, what we observe is both an upper and a lower bound on the annuity value for each respondent, and the figure plots both the upper and lower bound. The median lower bound represents a valuation of $17,500 (s.e.: $1211) for a $100-per-month reduction in Social Security benefits, while the median upper bound i s $20,000 (s.e.: $1211). Taking the midpoint, the median valuation is $18,750. Interestingly, this median valuation is remarkably close to the "actuarially fair" value of the annuity at age 62 calculated using Social Security Trustees' assumptions, which we estimate to be approximately $18,860 for the average individual in our sample.[12]

While the median response is therefore quite sensible compared to the income stream's actuarial value, the CDF shown in Figure 1 a lso reveals quite substantial heterogeneity in

---

[12] This lump-sum value roughly corresponds to the expected discounted value as of age 62 of a $100 real annuity calculated using unisex mortality rates from the 2010 OASDI Trustees' Report for the 1961 birth cohort (i.e., the cohort turning age 50 in 2011, which roughly corresponds to the median age of our sample), and a real interest rate of 2.9% (which is the long-term rate used by SSA in the 2010 report). We approximated the present value of the $100 monthly income stream by averaging the present value of an annual $1200 stream of payments starting at the beginning of the year, and the present value of annual $1200 payments received at the end of the year.

respondent's valuations. For example, about 6 percent of the sample reports a valuation of $1,500 or lower – a level so low that it is difficult to explain using any "rational" economic model (unless the individual is virtually certain he will not live more than another year or two; below we examine the impact of self-reported health status and survival probabilities on the results).

At the other extreme, over one-quarter of the sample reports annuity values of $60,000 or higher. Moreover, some 12 percent of the sample says that it would not accept the lump sum for less than $200,000. It is hard to imagine this being a fully-informed, rational response to a question eliciting the minimum amount they would accept for a reduction in Social Security benefits of $100 per month, or $1,200 per year: even if one earned only a 1% nominal return on the $200,000 lump sum, they could generate $2000 per year in interest (more than the foregone $1,200) and still maintain the value of the lump sum.

As we will discuss in more detail below, these results cannot be explained away by reference to standard concerns about subjective life expectancy, or numerous other possibly "rational" explanations. Nor can concerns about political risk to Social Security explain these results.[13] Thus, these responses suggest that at least some of the respondents are having difficulty providing economically meaningful values for the Social Security annuity, at least in the tails of the CDF.

---

[13] We controlled for political risk in two ways in this study. First, we ask a question assessing individuals' perceptions about the probability that Social Security benefits will be reduced in the future. Including responses to this question as a control variable in various analyses is consistently insignificant. Second, we have a version of our annuity valuation question in which we explicitly instruct individuals not to consider political risk by stating: *"From now on, please assume that you are absolutely certain that Social Security will make payments as promised, and that there is no chance at all of any benefit changes in the future other than the trade-offs discussed in the question below."* Using the most unbiased comparison available (i.e., comparing the response to the no-political-risk question to the baseline CV-Plus question for those for whom the two questions were adjacent, we find that the response to the no-political-risk question is a statistically significant 7 percent *lower* that the response to the baseline CV-Plus question. Taken literally, this implies a negative risk premium. We believe, however, that a more likely explanation is that our question may have had the unintended effect of making political risk more salient, rather than less. Overall, our analysis suggests that the incorporation of political risk does not alter our main findings.

4.2 CV-Plus versus CV-Minus

In Figure 2, we show the CDF of the CV-Minus question along with the CV-Plus. Recall that the CV-Minus question is one in which the individual is asked how much of a lump sum he would be willing to pay in order to gain access to an additional $100 per month of Social Security benefits for life. This figure shows that the distribution of annuity valuations from the CV-Minus solicitation is significantly below that of the CV-Plus. For example, the median midpoint response drops from $18,750 (s.e.: $1211) to $3,000 (s.e.: 247). The response at other points on the distribution similarly drop; the decline at the 25th percentile is from $9,250 (s.e.: 322) to $1,000 (s.e.: 200), and at the 75th percentile from $55,000 (s.e.: 3,803) to $8,500 (s.e.: 318). Taken at face value, these results indicate that people place a higher valuation on the Social Security annuity when asked about their willingness to *give up* some of it in return for a lump sum, but a lower value when asked how much they would be *willing to pay* to access a higher monthly benefit. This pattern is consistent with the idea that when individuals are faced with a complex decision, they respond in a manner consistent with status quo bias. In other words, they are only willing to shift away from the status quo when the "deal" they are offered is clearly a good one, such as buying an annuity very cheaply, or selling it a very high price.[14]

To rule out the possibility that these answers could be driven by consumers experiencing liquidity constraints, we also asked respondents about their ability to come up with the money needed for the lump sum if they had to. Interestingly, the results indicate that the vast majority (91 percent) indicated that their choice was not due to liquidity constraints.[15]

---

[14] For more on the status quo bias concept see Samuelson and Zeckhauser (1988) and Kahneman, et al. (1991).
[15] Specifically we asked whether the respondent could come up with $5,000 "if he had to", and separately whether he could come up with the lump sum needed to purchase the higher annuity. The time frame for coming up with the money was the same time frame as in the annuity valuation question, namely one year from now or the respondent's expected claim date, whichever is later. About two-thirds of the respondents answered that they were certain they could come up with $5,000, and over 90 percent respond that they could come up with the amount probably or certainly. About 82 percent of respondents indicate that they could come up with the lowest lump-sum amount that

4.3 Equivalent Variation Measures

Our main CV-Plus question above provided respondents with a scenario in which they could either remain with the status quo – what they believed to be their Social Security benefit under the current rules (which we call here the 'current Social Security benefit') – or move away from it by trading a portion of the annuity for a lump sum. An alternative way to ask the question is the EV-Plus approach, in which people could choose between receiving their current benefit plus an additional supplemental payment of $100 per month, or receiving their current benefit plus a one-time lump sum. In the EV-Minus version of the question, people were told their current benefit would be reduced, but they had to choose between receiving this reduction as a $100 per month reduction, or as a lump-sum reduction.

The distribution of responses is shown in Figure 3. As with the CV versions of the questions, valuations implied by the responses suggest a higher average valuation for the "Plus" than the "Minus" version. It is worth noting that the CV-Plus version of the question is equivalent to adding a $100 monthly payment to both options in the EV-Plus version of the question (and the CV-Minus and EV-Minus are similarly shifted by $100 per month). Relative to lifetime income and net worth, any income effect difference should be minimal for most of the population.

4.4 Correlation across Alternative Measures of Annuity Values

While Figures 1–3 showed the overall distribution of responses for the four valuation measures, they do not show how the *within-person* responses to these alternative valuation measures are correlated. For instance, we seek to learn whether those who give a high annuity

---

they declined to pay. Of the 18 percent that indicated that they could not come up with this amount, half said that even if they had had the money, they would have decline the pay the lump sum. Thus, for 91 percent of the respondents, liquidity constraints are not the reason for the low reported annuity valuation in the CV-minus trade-off question.

value in one measure also provide high valuations in the other three. To address this question, Table 2 provides correlations across our various outcome measures. Column 1, for example, confirms that CV-Plus and EV-Plus are positively correlated, a conclusion obtained by regressing the log of the midpoint value of the response of the EV-Plus question on the log of the midpoint value of the response to the CV-Plus question. The highly statistically significant coefficient of +0.35 clearly indicates that these responses are correlated with one another. This is particularly notable inasmuch as we asked the CV-Plus and the EV-Plus questions of all respondents but in different waves of the survey; thus every individual answered these two questions at least two weeks apart. Given this lag, it is unlikely that this correlation is driven by anchoring or memory effects that could arise if the questions were asked within the same questionnaire.

It is also important to rule out the possibility that this positive correlation is due to the fact that when we randomized the starting values for the lump-sum amounts, we randomized across individuals (rather than within individuals and across questions). This might raise a concern that correlated responses could simply be driven by different individuals facing starting values that are the same across waves, but different across individuals. Column 2 of Table 2 shows this is not a concern. Even after controlling for the starting values, the coefficient remains very similar (+0.34 versus +0.35).

In recognition of the fact that responses will be subject to some random measurement error, we also average across different CV-Plus measures (e.g., our standard CV-Plus with a $100 change, CV-Plus with a $500 change, etc.) to increase power by allowing some of the random measurement error within individuals to cancel out. When we do this in column 3, while still controlling for starting values, we find the correlation is even higher, with the coefficient on

the average CV-Plus measure now coming in at +0.48. Next we shift to examining the correlation between the "Plus" and "Minus" versions of our questions. In column 4, we display the correlation of CV-Minus with CV-Plus, again controlling for the starting value, where we see that the correlation is now *negative* (-0.15) and quite statistically significant.[16] In column 5, we report the correlation of the average Plus value with the average Minus value (with averages taken across CV and EV to reduce measurement error), again conditioning on the starting value. Again, we find a strongly significant negative coefficient of -0.27.

These patterns are consistent with the hypothesis that there is heterogeneity in the population, with some individuals give similar responses to the Plus and Minus questions, while others provide answers consistent with them demanding a "better deal" in order to move away from the status quo. That is, when faced with a complex decision that they may not fully understand, they require a higher lump sum to give up an annuity than they are willing to pay to obtain an annuity. To further assess this, in column 6 we interact the correlations with an index of financial literacy. This is measured as the sum of correct answers to the three questions devised for the Health and Retirement Study (Lusardi and Mitchell, 2007), and they are used in the ALP as well to rate respondents' financial literacy. Consistent with our hypothesis that the discrepancy between Plus and Minus is driven by heterogeneous responses to complexity, we find that that the wedge between the responses is much greater for those with lower levels of financial literacy. Specifically, for those with the lowest level of financial sophistication, the conditional correlation is -0.59. The interaction term is +0.16, suggesting that for the most literate individuals (for whom the financial literacy index equals 3), the correlation is a much lower and only marginally significant -0.10.

---

[16] While not reported in this table, we have also confirmed that other combinations of Plus and Minus are also negatively correlated (e.g., EV-Plus and EV-Minus, or CV-Plus and EV-Minus).

5. Further Evidence that Complexity Matters

Thus far, we have documented that (i) individuals are able to give responses to valuation questions that are sensible on average, and (ii) their responses to similarly constructed offers (e.g., CV-Plus and EV-Plus) are positively correlated, despite being asked in different waves. However, we have also shown that (iii) there are a non-trivial number of individuals who give "nonsensical" answers in the extreme tails of the distribution for each question, (iv) there is a negative correlation between responses to "Plus" and "Minus" question, and (v) this negative correlation is strongest for the least financially sophisticated. In this section, we dig deeper into the question of whether our results are consistent with financially unsophisticated individuals having more difficulty making decisions in this complex environment.

5.1 Sensitivity to Anchoring and Starting Values

Our experimental design allowed us to randomize along numerous dimensions. Among the margins along which we randomized were (i) the order in which the various versions of the annuity valuation question were presented (e.g., CV-Plus in wave 1 and [CV-Minus, EV-Plus, EV-Minus] in wave 2, [EV-Plus, CV-Minus, EV-Minus] in wave 1 and CV-Plus in wave 2),[17] (ii) the starting value for the size of the lump sum, (iii) the order in which the variations in increment sizes for CV-Plus were presented (i.e., small-to-large or large-to-small), and (iv) whether or not we showed the option with the lump-sum amount first. These randomizations were conducted independently, and a simple correlation analysis (not reported here) confirms

---

[17] We first randomized at the individual level whether CV-Plus was asked in the first or second wave of our survey. CV-Minus, EV-Plus, and EV-Minus were asked in the other wave of the survey. Within the wave where they were asked, we randomized the order in which we asked CV-Minus, EV-Plus, and EV-Minus over each of the six possible orderings.

that this randomization was indeed done correctly, such that variation along each dimension is orthogonal from the variation along other dimensions.

If our hypothesis is correct, i.e., if respondents found the annuity valuation problem a complex one, then we expect to find that they were more sensitive to irrelevant cues such as the impact of the starting value and the ordering of the variation size. In contrast, we do not necessarily expect that the order of the options (i.e., "Lump sum shown last") would matter for complex decisions as long as the respondent tries to answer the question. If, on the other hand, respondents were simply not taking the time to read the questions carefully (such as, for example, by always choosing the first choice), then we might expect the option order to matter.

These hypotheses are analyzed in the first column of Table 3, where we regress the log midpoint of our baseline CV-Plus variable (using a $100 variation in Social Security benefits) against the four variables capturing all sources of randomization.[18] The results are consistent with our hypothesis. First, there is no evidence that individuals were simply taking the first option shown (i.e., there is no effect of "Lump sum shown last"). This provides us with some comfort that the respondents are taking care answering the questions. Relatedly, it does not matter whether the question was asked in the first or second wave (i.e., "Asked in wave 1" is small and insignificant"). Second, there is bias with respect to both of the other measures, as would be expected if individuals were having difficulty making a complex decision. Specifically, the impact of the starting value is a statistically significant +0.29. Because both the annuity valuation and the starting value are measured in logs, this means, for example, that increasing the first lump-sum amount shown by 10% increases the average valuation reported by respondents by approximately 3%. Furthermore, if the CV-Plus question was shown after a CV-

---

[18] We do this analysis on the CV-Plus version because only the CV-Plus version is asked for different increment sizes of the Social Security amount. This means that we can randomize the order in which the increment sizes were shown only for the CV-Plus version.

Plus question with a larger change in Social Security (so the order is large-to-small), the respondents report on average a 70 log point higher valuation of the annuity than if the baseline CV-Plus question is shown first.

In columns 2 and 3, we divide the sample into groups based on financial literacy. Specifically, column 2 reports results for the most financially literate respondents (i.e., those scoring a 3 on the financial literacy index), and column 3 reports results for those with a lower score on financial literacy. Results show that the most financially literate were much less likely to be influenced by the irrelevant cues of the starting value and the ordering of the variation size, whereas the less-literate were much more sensitive. In column 4, we revert to the full sample but now interact financial literacy with our randomization measures. The findings confirm that less financially literate respondents are substantially more sensitive to the randomly selected parameters in the questions, particularly the starting value used to begin the lump-sum question series.

5.2 The Role of Financial Literacy

We have confirmed above that financial literacy is strongly correlated with the consistency of the annuity values people provide across alternative formulations of the annuity versus lump-sum tradeoff. To further show the importance of financial literacy, we follow Liebman and Luttmer (2011) by randomizing our questions over three possible starting values for the lump sum: $10,000, $20,000 and $30,000, and then we branch subsequent responses from there. Given this, one can engage in the following "thought experiment." If individuals were truly randomizing their responses, then we can calculate the expected annuity value for each of the three starting values as the average of the log midpoint of the full set of categories offered. This relationship is approximately linear. We can then calculate the slope to find that if an

individual is totally randomized, the log midpoint response should increase about 0.4 for each $10,000 increase in the starting value. To test this, we run a regression of the log of the midpoint valuation on the starting value (measured in units of $10,000), the coefficient β of which tells us how people actually responded to changes in the starting value. If we assume (for illustrative purposes only) that every individual was *either* a "total randomizer" or someone with perfect understanding of the task who expressed a consistent underlying annuity valuation, then we can interpret (β / 0.4) as being the proportion of the sample behaving as if they completely randomized, and $1 - (\beta / 0.4)$ as the fraction expressing a true, underlying valuation (following Luttmer and Samwick, 2011). Of course we recognize that people may not be strict randomizers or strict reporters of an immutable underlying value. But the calculation is offered as a way to illustrate and scale the effect of the starting value on respondents' expressed valuations. For the sample as a whole, the results are consistent with 41% of the sample randomizing their responses (this effect is statistically significant, with a standard error of 11%).

When we decompose the sample into three groups based on our index of financial literacy, rather dramatic differences emerge. For the most sophisticated, highest scorers on the financial literacy index, the proportion of "randomizers" falls to a statistically insignificant 20%. Those in the middle of the financial literacy index behave in a manner consistent with about 30% of the sample randomizing. Finally, for the least sophisticated individuals, the results are consistent with *all* of them being randomizers – indeed, the point estimate is 115%, with a standard error of 32%.

Having established that more financially sophisticated individuals provide more consistent responses to annuity valuation questions than do those who score more poorly on the financial literacy index, we take this discussion of financial literacy further in three ways. First,

we construct a new definition of decision-making ability based on the dissimilarity in responses to our Plus and Minus versions of the annuity valuation questions. Second, we empirically examine the determinants of financial literacy. Third, we restrict our attention to the most financially literate subset of the population and examine what factors are correlated with the reported annuity values.

We have noted above that our financial literacy index has good explanatory power for determining the extent of variation in the Plus versus Minus versions of our question. Next we leverage this insight by constructing a measure of decision-making ability based on how closely a respondent's EV-Plus and EV-Minus responses correspond. We then employ this EV-based measure as an *explanatory variable* in our CV-Plus regressions, as another proxy for financial literacy. The fact that the CV-Plus and EV questions were asked in different waves of the survey and elicit the information in slightly different ways means that we are not simply picking up a mechanical effect. Rather, we view the similarity of the EV-Plus and EV-Minus responses as a proxy for how informative one's CV-Plus response might be.

Results appear in Table 4. The first three columns use as a dependent variable the financial literacy question, where we coding someone as "sophisticated" if he scored a 2 or 3 on our scale, and unsophisticated otherwise. Columns 4 and 5 use our EV similarity measure, such that a respondent is counted as sophisticated if the log of his difference in EV-Plus and EV-Minus was less than one.[19] These definitions are admittedly arbitrary, but they have the advantage of counting approximately a third of respondents as "sophisticated" under either definition. Not surprisingly, our two measures of financial sophistication are significantly correlated, as is evident in Column 1 which offers a simple regression of financial literacy on our

---

[19] We further require that the small difference in EV-Plus and EV-Minus was not a result of the respondent always choosing the lump-sum option or the respondent never choosing the lump-sum option.

measure of within-EV similarity: the coefficient of +0.13 is highly statistically significant. Of course, the R-squared in such a simple correlation regression is small (0.02), suggesting that while correlated, these two measures are capturing somewhat different phenomena. In the next two columns of Table 4, we regress the financial literacy index measure against various demographic characteristics. Columns 2 and 3 show that sophistication increases with age, is higher for men than for women, and higher for whites than for blacks or Hispanics. We also find that financial sophistication is higher for better-educated and higher-income respondents. These findings are consistent with prior studies of financial literacy (e.g., Lusardi and Mitchell, 2007). Column 3 adds additional covariates; while we continue to find significant and quantitatively similar effects of age, sex, race, education and income, the additional variables beyond those add little explanatory power.[20]

In columns 4 and 5 of Table 4, we repeat the regressions from columns 2 and 3, but this time we use the degree of within-EV similarity as our dependent variable. The pattern of responses is, for the most part, consistent with what was generated using the financial literacy index (that is, sex, race, education and income all matter). Interestingly, the coefficient on age, while significant in all columns, changes signs from columns 2-3 to columns 4-5. One plausible explanation for this may be that as individuals age, they learn more "facts" about financial matters increasing their financial literacy score, but they become less able to think through novel complex decisions decreasing the similarity between their two EV answers.

5.3 Annuity Valuation among "Financially Sophisticated" Individuals

Finally we turn to an exploration of how the most "financially sophisticated" respondents to our survey value the annuity versus lump-sum tradeoffs to which they were exposed.

---

[20] For instance we also controlled on – but did not find to be significant - whether respondents indicating having children (to account for a possible bequest motive), whether they were in good or fair health, whether they were self-reported as risk-averse, whether they trusted financial institutions, and whether they owned their own home.

Columns 1-3 of Table 5 focus on the subset of almost 700 respondents who received a top score on the financial literacy questions, while columns 4-6 concentrates on approximately 650 respondents who were ranked in the prior table as most sophisticated based on the similarity of their two EV answers. The median annuity valuations among members of either subset are extremely similar to the median valuations in the entire sample, so there is no effect of financial sophistication on median valuations.

It is of interest that even for these two select groups, there are few, if any, robust findings across the models and dependent variables. Many coefficient estimates on a priori predictors of annuity valuations are not significant at conventional levels, and those that are statistically significant do not show a robust pattern across the various specifications. Moreover, the explanatory power of the individual predictors of annuity valuation is very low, with R-squareds ranging from 0.06 to 0.09. The lesson we take away from this analysis is that it appears rather difficult to predict annuitization decisions, even among the most financially sophisticated subset of the population.

6. Discussion and Conclusions

This paper provides evidence supportive of the hypothesis that many people find the annuitization decision quite complex, and that this complexity, rather than a taste for lump sums *per se*, may explain the observed low levels of annuity purchase. Specifically, we find that consumers tend to value annuities less when given the opportunity to buy more, but value them more highly when given the opportunity to sell annuities in exchange for a lump sum. Such behavior is consistent with people deciding to stick to the status quo, a pattern also detected in other, similar, settings including being more likely to take an annuity when offered it through a

DB pension - which traditionally pays benefits as annuities - than when offered one through a DC plan - which traditionally pays lump sums (Benartzi et al., 2011). It is also consistent with recently emerging evidence on the sensitivity of annuity choice to framing (Brown et al. 2008b; Brown et al., 2010). Moreover, we have demonstrated that consumers who are more financially literate are also much more likely to provide informative answers, and their responses are rather consistent across alternative ways of eliciting preferences.

If our conclusion - that complexity contributes to the lack of annuity demand - proves through future research to be robust, this will have a number of important implications for the annuities literature and for public policy. First, such a finding may raise doubts about whether consumers will be able to make utility-maximizing choices when confronted with the decision about whether to buy longevity protection in real-world situations. To the extent that individuals find these decisions complex, this might be important for assessing various policy interventions ranging from providing better information, to changing the default option in the typical DC plan from one of no annuitization to one of partial annuitization, to mandating some measure of compulsory annuitization. Naturally, the degree of compulsory annuitization that is deemed optimal is also a first-order consideration in determining the appropriate level of Social Security benefits in the U.S. and elsewhere.

In addition, our findings suggest that observers must be very careful when drawing conclusions about individual welfare based on observed behavior (i.e., "revealed preference") when it comes to annuities, and quite possibly other complex financial products such as long-term care insurance. For example, the fact that so few people annuitize their defined contribution pension balances when given the opportunity to do so, should not be interpreted as clear evidence that people do not value annuities.

Despite these caveats about the difficulty in eliciting annuity valuations from surveys or observed behavior, it is worth reiterating that there is a subset of financially more sophisticated individuals who can provide consistent, and presumably more informative, responses. Better-educated and more highly paid people are overrepresented among the group of financially more sophisticated respondents, while women and minorities are underrepresented in this group. This suggests that financial literacy efforts to enhance subgroups' retirement security might be most fruitfully targeted on ethnic/racial minorities and women. It should be noted however that consistency does not imply the absence of bias. Even sophisticated individuals may misjudge the annuity value of a lump sum (e.g. Stango and Zinman 2011).

The evidence provided in this paper, though experimental in nature, is also somewhat indirect. To further test whether complexity of the lump-sum versus annuity decision is in fact a driving force behind the reluctance to voluntarily annuitize, we suggest at least two possible avenues for future research. First, it may be possible to alter the degree of complexity in the lump-sum versus annuity choice, and find out if the dispersion in valuations is indeed a function of choice complexity. There are several dimensions along which the complexity could be varied, but two interesting ones would be to truncate the time horizon to simplify the intertemporal choice, and to reduce the dimensionality of the uncertainty that individuals face. Second, one could maintain the level of complexity and test whether individuals can be "taught" to make more informed and consistent decisions by experimentally providing them with task-relevant financial literacy training. We view these as two fruitful areas for future research.

In addition to advancing our academic understanding of consumer behavior in this area, our results also have considerable practical policy relevance. The U.S. Social Security system is on a fiscally unsustainable path that will require increasing revenue or curtailing benefit growth

in the not-too-distant future (Cogan and Mitchell, 2003). As policymakers evaluate alternative approaches to reform, it is important to understand how consumers actually value the system's mandatory old-age annuity payments, and how this perceived value is affected by the nature and the framing of the trade-off presented. In particular, our findings do not offer any particular road map as to how much people of different demographic characteristics might be willing to pay to maintain the current system, nor are they able to pinpoint people's willingness to give up some portion of their annuity benefits in exchange for a lump sum. Our findings are also relevant to state and local pension plans in the U.S. which are now grappling with how to reform their defined benefit (DB) pensions to address their substantial underfunding problems (e.g., Novy-Marx and Rauh, 2011). Additionally, there is an ongoing discussion of what role annuities ought to play in defined contribution (DC) or 401(k) pension plans, with increasing discussion of whether life annuities could and should be encouraged in such settings (c.f. Gale et al. 2008). In the US and the rest of the world, it is critical to explain why people continue to be ill-protected against outliving their retirement assets and to find ways to enhance markets for payout annuities.

References

Agnew, Julie R., Lisa R. Anderson, Jeffrey R. Gerlach and Lisa R. Szykman. 2008. "Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing and Defaults." *American Economic Review*. May: 418-422.

Benartzi, Shlomo, Alessandro Previtero and Richard H. Thaler. 2011. "Annuitization Puzzles." *Journal of Economics Perspectives*. Forthcoming.

Bernheim, B. Douglas. 2002. "Taxation and saving". In Alan J. Auerbach and Martin Feldstein, Editors. *Handbook of Public Economics*, Elsevier, 3: 1173-1249.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian. 2008. "How are Preferences Revealed?" *Journal of Public Economics*. 92(8-9) August: 1787-1794.

Brown, Jeffrey R. 2001. "Private Pensions, Mortality Risk, and the Decision to Annuitize." *Journal of Public Economics* 82(1): 29–62.

Brown, Jeffrey R. 2008. "Understanding the Role of Annuities in Retirement Planning." In Annamaria Lusardi, ed., *Overcoming the Savings Slump: How to Increase the Effectiveness of Financial Education and Saving Programs*. Chicago: University of Chicago Press. 178 – 206.

Brown, Jeffrey R. 2009. "Automatic Lifetime Income as a Path to Retirement Security." White paper written for the American Council of Life Insurers. http://www.acli.com/Events/Documents/3f3ac12d2d004472906283563c87ee55Automatic_Lifetime_IncomePaper.pdf Last accessed 9/15/2011.

Brown, Jeffrey R., Marcus Casey, and Olivia S. Mitchell. 2008a. "Who Values the Social Security Annuity? Evidence from the Health and Retirement Study." Unpublished manuscript.

Brown, Jeffrey R., Arie Kapteyn, and Olivia S. Mitchell 2010. "Framing Effects and Expected Social Security Claiming Behavior." NBER Working Paper 17018.

Brown, Jeffrey R., Jeffrey R. Kling, Sendhil Mullainathan and Marian Wrobel. 2008b. "Why Don't People Insure Late Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle." *American Economic Review*. May: 304-309.

Brown, Jeffrey R., Olivia S. Mitchell, and James M. Poterba. 2001. "The Role of Real Annuities and Indexed Bonds in an Individual Accounts Retirement Program." In *Risk Aspects of Investment-Based Social Security Reform*, ed. J. Campbell and M. Feldstein. Chicago: University of Chicago Press: 321–360.

Brown, Jeffrey R., Olivia S. Mitchell, and James M. Poterba. 2002. "Mortality Risk, Inflation Risk, and Annuity Products," In *Innovations in Retirement Financing*, ed. O. Mitchell, Z. Bodie, B. Hammond, and S. Zeldes. Philadelphia: University of Pennsylvania Press: 175–197.

Brown, Jeffrey R., and James M. Poterba. 2000. "Joint Life Annuities and the Demand for Annuities by Married Couples." *The Journal of Risk and Insurance* 67(4): 527–553.

Bütler, Monika and Frederica Teppa. 2007. "The Choice between an Annuity and a Lump Sum: Results from Swiss Pension Funds." *Journal of Public Economics*. 91(10): 1944-1966.

Cappelletti, Giuseppe, Giovanni Guazzarotti, and Pietro Tommasino. 2011. What Determines Annuity Demand at Retirement?" Bank of Italy Temi di Discussione Working Paper No. 805.

Chalmers, John and Jonathan Reuter. 2009. "How Do Retirees Value Life Annuities? Evidence From Public Employees." NBER Working Paper 15608.

Chai, Jingjing, Wolfram Horneff, Raimond Maurer, and Olivia S. Mitchell. 2011. "Optimal Portfolio Choice over the Life Cycle with Flexible Work, Endogenous Retirement, and Lifetime Payouts." *Review of Finance* forthcoming.

Chang, L. and J.A. Krosnick. 2009. "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly*, 73: 641-648.

Choi, James, David Laibson and Brigitte C. Madrian. 2005. "Are Empowerment and Education Enough? Underdiversification in 401(k) Plans." *Brookings Papers on Economic Activity*, 2: 151–198.

Cogan, John F. and Olivia S. Mitchell. 2003. "Perspectives from the President's Commission on Social Security Reform." *Journal of Economic Perspectives*. 17(2). Spring: 149–172.

Couper, M.P., A. Kapteyn, M. Schonlau, and J. Winter. 2007. "Noncoverage and Nonresponse in an Internet Survey." *Social Science Research*, 36(1):131-148.

Davidoff, Thomas, Jeffrey R. Brown, and Peter A. Diamond. 2005. "Annuities and Individual Welfare." *American Economic Review* 95(5): 1573–1590.

Dillman, D.A., J.D Smyth, and L.M. Christian. 2008. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd edition. Hoboken, NJ: Wiley.

Dushi, Irena and Anthony Webb. 2006. "Rethinking the Sources of Adverse Selection in the Annuity Market." In Pierre Andre Chiappori and Christian Gollier (editors). *Competitive Failures in Insurance Markets: Theory and Policy Implications*. Cambridge: MIT Press: 185-212.

Duffy, B., K. Smith, G. Terhanian, and J. Bremer. 2005. "Comparing Data from Online and Face-to-face Surveys." *International Journal of Market Research,* 47: 615–639.

Gale, William G., J. Mark Iwry, David C. John and Lina Walker. 2008. *Increasing Annuitization of 401(k) Plans with Automatic Trial Income*. Retirement Security Project Report. Washington, D.C.: Brookings Institution.

Greenwald, Mathew, Arie Kapteyn, Olivia S. Mitchell, and Lisa Schneider. 2010. "What Do People Know about Social Security?" Financial Literacy Consortium Report to the SSA, September.

Horneff, Wolfram, Raimond Maurer, Olivia S. Mitchell, and Ivica Dus. 2007. "Following the Rules: Integrating Asset Allocation and Annuitization in Retirement Portfolios." *Insurance: Mathematics and Economics*. 42: 396-408.

Horneff, Wolfram J. Raimond H. Maurer, Olivia S. Mitchell, and Michael Z. Stamos. 2010. "Variable Payout Annuities and Dynamic Portfolio Choice in Retirement." *Journal of Pension Economics and Finance*. 9, April: 163-183.

Horneff, Wolfram, Raimond Maurer, Olivia S. Mitchell, and Michael Stamos. 2009. "Asset Allocation and Location over the Life Cycle with Survival-Contingent Payouts."*Journal of Banking and Finance*. (33) 9 September: 1688-1699.

Hurd, Michael and Stan Panis. 2006. "The Choice to Cash out Pension Rights at Job Change or Retirement," *Journal of Public Economics*, 90: 2213-2227.

Inkmann, Joachim, Paula Lopes, and A. Michaelides. 2007. "How Deep is the Annuity Market Participation Puzzle?" Netspar Discussion Paper 2007-011.

Kahneman, Daniel, and Amos Tversky. (1981). "The Framing of Decisions and the Psychology of Choice." *Science,* 211(4481), January 30: 453-458.

Kahneman, Daniel, Knetsch, J. L. and Thaler, R. H. 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5(1): 193-206

Koijen, Ralph S.J., Nijman, Theo E. and Werker, Bas J.M. 2007. "Optimal Annuity Risk Management." Review of Finance forthcoming. CentER Working Paper Series No. 2006-78.

Kotlikoff, Laurence J., and Avia Spivak. 1981. "The Family as an Incomplete Annuities Market." *Journal of Political Economy* 89(2): 372–391.

Liebman, Jeffrey B. and Erzo F.P. Luttmer. 2011. "The Perception of Social Security Incentives for Labor Supply and Retirement: The Median Voter Knows More Than You'd Think. Unpublished Manuscript, Dartmouth College.

Lockwood, Lee. 2011. "Bequest Motives and the Annuity Puzzle," *Review of Economic Dynamics*. Forthcoming.

Loosveldt, G. and N. Sonck. 2008. "An Evaluation of the Weighting procedures for an Online Access Panel Survey/" *Survey Research Methods*, 2: 93–105.

Lusardi, Annamaria and Olivia S. Mitchell. 2007. "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth." *Journal of Monetary Economics*. 54(1) January: 205-224.

Luttmer, Erzo F.P., and Andrew A. Samwick. 2011. "The Costs and Consequences of Perceived Political Uncertainty in Social Security." Unpublished Manuscript, Dartmouth College.

Malhotra, N. and J.A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Non-probability Samples." *Political Analysis*, 15: 286–323.

Milevsky, Moshe Arye, and Virginia R. Young. 2007. "Annuitization and Asset Allocation." *Journal of Economic Dynamics and Control.* 31(9): 3138-3177.

Mitchell, Olivia S., James M. Poterba, Mark J. Warshawsky, and Jeffrey R. Brown. 1999. "New Evidence on the Money's Worth of Individual Annuities." *American Economic Review* 89(5): 1299–1318.

Mitchell, Olivia S., John Piggott, and Noriyuke Takayama, eds. 2011b. *Revisiting Retirement Payouts: Market Developments and Policy Issues*. Oxford University Press. *Forthcoming*.

Modigliani, Franco. 1988. "The Role of Intergenerational Transfers and Life-Cycle Saving in the Accumulation of Wealth." *Journal of Economic Perspectives*, Spring 2(2): 15-40.

Novy-Marx, Robert and Joshua Rauh. 2011. The Liabilities and Risks of State-Sponsored Pension Plans. *Journal of Economic Perspectives*. 23(4): 191-210.

Peijnenburg, Kim, Theo Nijman and Bas Werker. 2010a. "Optimal Annuitization with Incomplete Annuity Markets and Background Risk during Retirement." Netspar Discussion Paper.

Peijnenburg, Kim, Theo Nijman, and Bas J.M. Werker. 2010b. "Health Cost Risk and Optimal Retirement Provision: A Simple Rule for Annuity Demand," Pension Research Council Working Paper No. WPS 2010-08.

Poterba, James, Steve Venti, and David Wise. 2011. "The Drawdown of Personal Retirement Assets." NBER Working Paper 16675.

Samuelson, W. & R. J. Zeckhauser. 1988. "Status-quo Bias in Decision Making." *Journal of Risk and Uncertainty*, 1: 7-59.

Schieber , Sylvester J. and John B. Shoven. 1999. *The Real Deal: The History and Future of* Social Security. New Haven: Yale University Press.

Schonlau, M., A. van Soest, A. Kapteyn, and M. Couper. 2009. Selection Bias in Web Surveys and the Use of Propensity Scores, *Sociological Methods and Research*, 37: 291-318.

Shepard, Mark. 2011. "Social Security Claiming and the Life Cycle Model." Unpublished Manuscript, Harvard University.

Simon, Herbert A. 1947. *Administrative Behavior, a Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.

Sinclair, Sven H.and Kent A. Smetters. 2004. "Health Shocks and the Demand for Annuities." Technical Paper Series; Congressional Budget Office. Washington, DC: GPO, July

Stango, Victor and Jonathan Zinman. 2011. "Exponential Growth Bias ad Household Finance". *Journal of Finance*, forthcoming.

Taylor, H. 2000. "Does Internet Research Work? Comparing Online Survey Results with Telephone Surveys." *International Journal of Market Research,* 42(1): 51–63.

Turra, Cassio and Mitchell, Olivia S. 2008. "The Impact of Health Status and Out-of-Pocket Medical Expenditures on Annuity Valuation." In John Ameriks and Olivia S. Mitchell, eds. *Recalibrating Retirement Spending and Saving*. Oxford University Press: 227-250.

Vehovar, V., Z. Batagelj, and K. Lozar Manfreda. 1999. "Web surveys: Can the weighting solve the problem?" *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 962–967.

Warner, John T., and Saul Pleeter. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review* 91(1): 33-53.

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

Yeager, D.S., J.A. Krosnick, L. Chang, H.S. Javitz, M.S. Levindusky, A. Simpser, and R. Wang. 2009. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." Working paper, Stanford University.

Data Appendix: The Rand American Life Panel

Sample Construction

Our survey was conducted in the RAND American Life Panel (ALP). The ALP consists of a panel of U.S. households that regularly takes surveys over the Internet. An advantage relative to most other Internet panels is that the respondents to the ALP need not have Internet when they get recruited (as is described in more detail below) and thus can be based on a probability sample of the US population.[21] This is in contrast with so-called convenience Internet samples, where respondents are volunteers who already have Internet and for example respond to banners placed on frequently visited web-sites, in which they are invited to do surveys and earn money doing it. The problem with convenience Internet samples is that their statistical properties are unknown. There is a fairly extensive literature comparing probability Internet samples like the ALP and convenience Internet samples or trying to establish if convenience samples can somehow be made population representative by reweighting.

For instance, Chang and Krosnick (2009) simultaneously administered the same questionnaire (on politics) to an RDD (random digit dialing) telephone sample, an Internet probability sample, and a non-probability sample of volunteers who do Internet surveys for money. They found that the telephone sample has most random measurement error, while the non-probability sample has the least. At the same time, the latter sample exhibits most bias (also after reweighting), so that it produces the most accurate self-reports from the most biased sample. The probability Internet sample exhibited more random measurement error than the non-probability sample (but less than the telephone sample) and less bias than the non-probability

---

[21] Other probability Internet surveys include the Knowledge Networks panel in the U.S. (http://www.knowledgenetworks.com/knpanel/index.html), and the CentERpanel and LISS panel in the Netherlands: (http://www.centerdata.nl/en/centerpanel and http://www.centerdata.nl/en/MESS). Of these the CentERpanel is the oldest (founded in 1991).

Internet sample. On balance, the probability Internet sample produced the most accurate results. Yeager et al. (2009) conducted a follow-up study comparing one probability Internet sample, one RDD telephone sample, and seven non-probability Internet samples and a wider array of outcomes. Their conclusions are the same: Both the telephone sample and the probability Internet sample show the least bias; reweighting the non-probability samples does not help (for some outcomes, the bias gets worse; for others, better). They also found that response rates do not appear critical for bias. Even with relatively low response rates, the probability samples yield unbiased estimates. It is not clear a priori why non-probability samples do so much worse. As they note, it appears that there are some fundamental differences between Internet users and non-Internet users that cannot be redressed by r eweighting. Indeed, Couper et al. (2007) and Schonlau et al. (2009) show weighting and matching do not eliminate differences between estimates based on samples of respondents with and without Internet access. Several other studies point at equally mixed results, including Vehovar et al. (1999); Duffy et al., (2005); Malhotra and Krosnick (2007), Taylor (2000), Loosveldt and Sonck (2008).

ALP respondents have been recruited in one of four ways. Most were recruited from respondents age 18+ to the Monthly Survey (MS) of the University of Michigan's Survey Research Center (SRC). The MS is the leading consumer sentiment survey that incorporates the long-standing Survey of Consumer Attitudes and produces, among others, the widely used Index of Consumer Expectations. Each month, the MS interviews approximately 500 hous eholds, of which 300 households are a random-digit-dial (RDD) sample and 200 a re re-interviewed from the RDD sample surveyed six months previously. Until August 2008, S RC screened MS respondents by a sking them if they would be willing to participate in a long-term research project (with approximate response categories "no, certainly not," "probably not," "maybe,"

"probably," "yes, definitely"). If the response category is not "no, certainly not," respondents were told that the University of Michigan is undertaking a joint project with RAND. They were asked if they would object to SRC sharing their information about them with RAND so that they could be contacted later and asked if they would be willing to actually participate in an Internet survey. Respondents who do not have Internet were told that RAND will provide them with free Internet. Many MS-respondents are interviewed twice. At the end of the second interview, an attempt was made to convert respondents who refused in the first round. This attempt includes the mention of the fact that participation in follow-up research carries a reward of $20 for each half-hour interview. Respondents from the Michigan monthly survey without Internet were provided with so-called WebTVs (http://www.webtv.com/pc/), which allows them to access the Internet using their television and a telephone line. The technology allows respondents who lacked Internet access to participate in the panel and furthermore use the WebTVs for browsing the Internet or email. The ALP has also recruited respondents through a snowball sample (respondents suggesting friends or acquaintances who might also want to participate), but we do not use any respondents recruited through the snowball sample in our paper. A new group of respondents (approximately 500) has been recruited after participating in the National Survey Project, created at Stanford University with SRBI. This sample was recruited in person, and at the end of their one-year participation, they were asked whether they were interested in joining the RAND American Life Panel. Most of these respondents were given a laptop and broadband Internet access. Recently, the American Life Panel has begun recruiting based on a random mail and telephone sample using the Dillman method (see e.g. Dillman et al., 2008) with the goal to achieve 5000 active panel members, including a 1000 Spanish language subsample. If these new

participants do not have Internet access yet, they are also provided with a laptop and broadband Internet access.

Calculation of Social Security Benefits

For most ALP respondents, we have previously estimated monthly Social Security benefits (described in Brown et al., 2010). To do s o, we took respondents through a fairly detailed set of questions asking about years in which they had labor earnings and an approximation of earnings in those years. We then fed these earnings through a benefit calculator provided by SSA to calculate the individual's "Primary Insurance Amount" (PIA) which is equivalent to the benefit the individual would receive if he were to retire at his normal retirement age. Next we applied SSA's actuarial adjustment for earlier or later claiming. We also asked respondents if the estimated benefit amount seemed reasonable to them, and we gave them an opportunity to change this estimate if they believed it was not a good approximation. All subsequent lump-sum and annuity questions then pivot off this estimated monthly Social Security benefit amount.

For the few respondents who indicated they did not expect to receive a benefit (nor did they expect one from a living or deceased spouse), we imputed 'standard monthly benefit amounts' based on age, sex, and educational levels. We then ask the respondent to assume, for the purposes of the questions to follow, that he or she would receive this benefit, as follows:

> *Even though we understand that you are not eligible to receive Social Security benefits, we would like to ask you to complete this survey assuming you would be eligible. In other words, please answer in this survey what you would have done or chosen if you would be eligible for Social Security benefits.*