Broad Framing in Retirement Income Decision Making

Hal E. Hershfield [a], Suzanne B. Shu [b], Stephen A. Spiller [a], and David Zimmerman [a]
[a] UCLA Anderson School of Management
[b] Cornell University and NBER

The research reported herein was pursuant to a grant from the U.S. Social Security Administration (SSA), funded as part of the Retirement and Disability Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, UCLA, Cornell University, or the NBER Retirement Research Center.

Abstract

Retirees often narrowly bracket retirement income decisions, myopically considering OASI claiming age, pension or 401(k) payouts, annuity purchases, long term care insurance, and use of home equity as independent and unrelated decisions. Prior research on narrow versus broad framing in financial decisions regularly finds that this type of narrow decision framing can cause individuals to accept lower risk, lower value outcomes, whereas a more broadly bracketed set of options can lead to more optimal aggregated choices. In this paper, we use a custom-built retirement decision aid to test how aggregating outcomes across different retirement funding sources, which has previously been unexplored, affects retirement decisions. In particular, we present two studies that experimentally test whether people select systematically different investment risk allocations, wealth drawdown strategies, annuity decisions, and SSA claiming intentions when they are shown the aggregated outcome of the decisions or each piece individually. We find that decisions can be affected by aggregating outcomes, that individuals report higher satisfaction with their decisions when made in an aggregated environment, but that they also indicate that the outcomes they have chosen are less desirable in hindsight than other possible retirement income paths.

Retirees often make retirement income decisions in narrow brackets, myopically considering Old-Age and Survivor's Insurance (OASI) claiming age, pension or 401(k) payouts, annuity purchases, long-term care insurance, and use of home equity as independent and unrelated decisions. Thoughtfully combining these different income sources into a comprehensive decumulation strategy requires mentally combining the risks and benefits associated with different programs and assets, which may be quite challenging. For example, when thinking about decisions for OASI claiming, wealth decumulation, and guaranteed lifetime income products (e.g., annuities), the tradeoffs between longevity risk, stock market risks, and higher future income can make each decision a highly complex task. Thinking of each domain as a separate decision, rather than looking at how they operate in aggregate, may make it even more difficult to evaluate global tradeoffs and also make it difficult to appreciate how potential outcomes can be complementary in generating a smoother path of retirement income.

Research on narrow versus broad framing in financial decisions regularly finds that this type of narrow decision framing can cause individuals to accept lower risk and lower value outcomes, whereas a more broadly bracketed set of options can lead to more optimal aggregated choices (Read, Loewenstein, and Rabin 1999; Webb and Shu 2017). Broadly bracketing outcomes has also been shown to increase risk tolerance (Benartzi and Thaler 1999), especially for individuals seeing investment outcomes aggregated over larger periods of time (Benartzi and Thaler 1995; Gneezy and Potters 1997; Langer and Weber 2001; Thaler et al. 1997). In this project, we test how aggregating outcomes across different sources of retirement income, a topic which has previously been unexplored, affects retirement decisions. We expect that, similar to broad bracketing of other financial outcomes, an aggregate view of sources of retirement income

(OASI benefits, savings wealth, and annuities) may lead to different decisions and thus result in different retirement income outcomes relative to when each decision is made independently.

To explore this possibility, we use a custom-built retirement decision aid to experimentally test whether people select systematically different risk allocations (stocks vs bonds), choose different savings withdrawal strategies, make different annuity decisions, and adjust their Social Security claiming intentions differently when they are shown the aggregated outcome of those decisions or each piece individually. We predict that by combining these risks into a single integrated retirement income metric (broad bracketing), people can more clearly evaluate the risks and understand the impacts that each decision has on their overall circumstances. For example, calculating the exact implications of withdrawing retirement savings more heavily in early retirement in order to delay OASI claiming is a decision that involves complex risk tradeoffs that may be hard to reason about. A decision aid that shows the aggregated impact of these decisions may make it easier for the individual to reason through the costs and benefits of using one income source to make different decisions regarding other income sources.

In what follows, we first briefly review the research on how narrow versus broad framing (bracketing) affects financial decisions. We then describe our experimental setup, and report on the results of an initial study with nearly 400 individuals in the 40 to 63 age range. We find that individuals who make retirement income decisions and receive income and wealth feedback in an aggregated frame, rather than in separate frames, are better able to smooth their retirement income. They also report intentions for claiming OASI benefits at earlier ages, less use of savings withdrawal to bridge the gap before OASI benefits, and less interest in annuities. In a second follow-up study, run with 600 individuals also in the 40 to 63 age range, we repeat this test with some minor study modifications, and also test how positively study participants score the aggregated outcomes of their own decisions relative to other possible outcomes. We find that the income smoothing result of Experiment 1 does not hold after modifying the wealth information provided to participants, but that participants in the aggregate frame now select a higher level of investment risk in their savings decision. In a follow-up wave with 390 of the 600 initial participants we find that people typically score their own selected outcomes better than four of the six other possible outcomes, regardless of condition.

Broad versus Narrow Bracketing

Individuals must often make decisions in which financial outcomes accrue over time and via many different repeated events. For example, investors may receive feedback and make decisions about their investments as narrowly as every day, or as broadly as every few years. If the underlying risk does not change over time, the length of the intervals should not affect the resulting investment strategy since the information is the same. However, it has been well documented that viewing information in an aggregated versus segregated format can have significant impacts on individual decision making. In particular, more frequent evaluations (narrow frames) can generate a perception of more volatile outcomes, while less frequent evaluations (broad frames) can provide a more holistic view of the actual probability distribution of the outcomes. As a result, investors who view aggregated outcomes in broad frames often make decisions that yield higher value long-run outcomes (Benartzi and Thaler 1995).

This idea that choices differ according to how they are aggregated together is known as choice bracketing (Read, Loewenstein, and Rabin 1999). Choice bracketing exploits the asymmetry between gains and losses (Kahneman and Tversky 1979) as well as a narrow or myopic approach to decision making (Kahneman and Lovallo 1993) and may result in both adding-up effects (in which cumulative effects are neglected when outcomes are evaluated separately) and trade-off effects (in which poor results of some choices can be integrated with the positive results of other choices (Read, Loewenstein, and Rabin 1999). Individuals may choose to bracket narrowly because of cognitive constraints (an inability to consider multiple outcomes at once), motivation to achieve specific narrow goals, or a basic tendency to consider decisions as they are sequentially presented (Read, Loewenstein, and Rabin 1999). The effect of narrow bracketing is that decision-makers tend to isolate current choices from a larger portfolio of choices, which may in turn lead them to overweight immediate losses and underappreciate long-run cumulative gains. For example, gambles that are presented in isolated stages are often rejected, while the same set of gambles is accepted when integrated into a single outcome. The decision maker "misses the big picture," failing to see how the individual pieces of information fit into a larger pattern or outcome distribution.

One example of choice bracketing is myopic loss aversion (Benartzi and Thaler 1995), which takes advantage of both loss aversion and the lack of aggregation characteristic of narrow framing. Myopic loss aversion has been tested experimentally and these studies find that narrow bracketing typically leads to more risk aversion while broad bracketing leads to more normative and financially optimal choices in a world of positive EV risks (Thaler, Tversky, Kahneman, and Schwartz 1997; Gneezy and Potters 1997; Langer and Weber 2001; Haigh and List 2005). These myopic loss aversion studies theorize that the narrow bracketing results in a change in preferences rather than a change in beliefs. More recent work has documented that broader bracketing leads to more optimal risk preferences for all risk types, not just positive EV risks, and that the bracketing effect is driven by changes in perceived risk, loss aversion, and cognitive capacity constraints (Webb and Shu 2017). Bracketing effects have also been experimentally explored for repeated plays of identical gambles and other sequences of financial outcomes (Wedell and Böckenholt 1994, Keren 1991, Thaler and Johnson 1990, Klos et al. 2005, Redelmeier and Tversky 1992). Importantly, a common theme in these tests of broad versus narrow bracketing is that a more aggregated frame can often lead to better financial outcomes for individuals. This work extents the upon previous research by having people evaluate outcomes that are income streams over time and includes a richer choice set which allows for trade-off effects in addition to adding up effects.

Overview of Current Research

In the current research, and as described in more detail below, we test how providing individuals with aggregated versus separate outcome information (income and wealth) affects outcomes and choices made about three sources of retirement income: OASI benefits, savings decumulation, and guaranteed income from life annuities. To do so, we developed a retirement income "decision aid" that took participants through each of those decisions and allowed them to iteratively adjust choices to achieve their desired outcomes. We describe this tool along with our research methods, and our preliminary results, in more detail in the sections below. We use the same decision tool, with slight modifications, in two experiments, and we describe the tool in a combined method section before turning to the results from each study. Since data collection is still in progress and analysis is ongoing, all results presented in this paper are preliminary and subject to change.

Method

The Decision Tool. People were asked to make a financial plan for decumulation using an online tool. In order to simplify the decision, we gave people a specific age, income, and savings scenario rather than letting them input information about themselves and then creating a plan using those inputs. This reduces the complexity of the task compared to what a typical household might confront, but we wanted to understand if our intervention would help even in this simplified context. Participants were asked to make a plan using three different financial products: Social Security retirement benefits, retirement savings, and annuities. They received immediate feedback in the form of graphs showing estimated income (and, for some conditions, wealth) over time, along with a probability that they would run out of retirement savings by age 85. This first study is exploratory in nature and thus is not pre-registered because our intervention could impact many outcomes measured in the study. Follow-up studies will have pre-registered designs, data collection, and analyses.

Procedure. After completing the comprehension check, participants walked through an in-depth explanation of the task they were about to complete. The directions started with a general overview of how to navigate and understand the decumulation tool and then stepped through each decision they would be asked to make. In the decumulation tool people saw three different financial products they could make decisions about: Social Security old age (OASI) benefits, savings, and guaranteed income (a single, deferred life annuity). Instructions specific to each element of the tool, and highlighting how the outcome feedback would change according to their decisions, was provided through a series of screenshots and detailed instructions.

Our main dependent variables are based on the retirement income decisions made within each financial product. For the Social Security product, people were able to select what age they would claim from ages 62 to 70. For retirement savings, participants were asked to select a general withdrawal progression from retirement savings: increasing, flat or decreasing withdrawals; were asked whether or not they would like to take extra withdrawals from retirement savings prior to claiming Social Security; and were asked to select one of three different investment allocation paths that varied the ratio of stocks to bonds. In the annuity product space, people were asked to select the percentage of their retirement savings to annuitize and the starting age for the annuity payments. Within each product decision space, individuals saw a graph of estimated income from ages 60 to 100. Within the retirement savings product decision screen, or for anyone in the aggregated outcome condition, a graph of estimated wealth over time from ages 60 to 100 was also provided. Additionally, people were provided with a calculated probability that they will still have positive (non-zero) retirement savings at age 85 (see Figures 1a-d for examples).

Our tool had default selections, but we explicitly encouraged people to try out different options to get an understanding of how varying their selections would impact their resulting estimated wealth and income over time. The default was claiming at age 63, having an increasing withdrawal sequence, not taking extra withdrawals before claiming Social Security, and not purchasing an annuity. Participants were allowed to switch between financial products as much as they liked and change their selections to see the impacts on income and wealth. The median time spent on the decumulation tool was 1.77 minutes and the median time to complete the study was 17 minutes. We now go into details for the construction and measurement for each of the three financial products and the options available to participants.

Social Security. Our estimated Social Security payments are calculated based on someone retiring at age 62 when making $55,000 per year. The benefits payments were adjusted on a yearly basis using the percentage adjustments of 6.66% for ages 62 to 64, 5% for ages 64 to

67 and 8% for ages 67 to 70. Once the participant made a decision of which age they would claim benefits, the yearly adjustment for inflation was 2.5% for calculating the future income path. Benefits continue for the entire age range shown in the graphs (i.e., to age 100). A screenshot of the Social Security decision page (separate condition) is provided in Figure 1a.

Retirement Savings. Recall that participants were asked to imagine that they currently held $250,000 in retirement savings and own a home outright. Questions in this section were related to the liquid savings and did not include housing wealth. To draw down their savings, participants had three different withdrawal patterns that they could choose from. The decreasing withdrawal progression started at 7% of the initial retirement savings balance, $250,000, and slowly decreased over time. The flat withdrawal progression took 4% of the initial retirement savings balance every year. The increasing withdrawal progression took started at 3% of the initial retirement savings balance and slowly increased over time. Each sequence was also adjusted for inflation at a rate of 2.7% per year starting at age 60. For each of these progressions, it was possible for people to run out of money before the age 100, at which point their retirement savings withdrawals would drop to zero for the remaining years.

People could also choose whether or not to take extra withdrawals from retirement savings until they claimed Social Security if they chose to delay claiming beyond age 62. These extra withdrawals were equal to the income someone would have received for a particular age if they had claimed at that age. If people took extra withdrawals they could withdraw a maximum of $29,450.71 at age 62, if they did not take extra withdrawals the maximum was $16,992.54 at age 62.

Finally, people were allowed to select an investment allocation plan that varied their level of investment between stocks and bonds. They were reminded that retirement savings would fluctuate more from year to year if they selected options with higher levels of stocks. For each of the plans, participants had an initial allocation in stocks of either 30%, 60%, or 90%, which steadily decreased until age 77 to the levels of 0%, 30%, or 60% respectively. The allocation to stocks and bonds then remained constant from age 77 until age 100.

An important element of the feedback on this decision screen was the path of retirement wealth and the associated probability of still having non-zero wealth at age 85. In order to estimate the uncertainty associated with investing we used bootstrap sampling. For each combination of choices people could make for their retirement savings, we calculated 400 bootstrap samples of returns, the associated withdrawals, and wealth states. In each sample, we randomly selected both the S&P 500 and 10-year Treasury bond returns from the years 1926 to 2016, inclusive. We decided to fix future inflation to 2.7% and thus wanted to calibrate the stock and bond returns to this level of inflation. We subtracted 3.5 and 2 percentage points from all stock and bond returns, respectively, in order to better match expectations over the next several decades. This gave us an average return of 8.4% for stocks and 2.9% for bonds. Returns were independently and randomly selected for each age from 60 to 100. Using these returns and the options selected we had 400 possible paths for each combination of options. Since we were providing purely graphical feedback and our risk communication was less precise than a standard deviation of wealth states we did not have thousands of bootstrap samples. A screenshot of the retirement savings decision page (separate condition) is provided in Figure 1b.

Annuity. Individuals were able to make two decisions related to the purchase of a guaranteed income product, a deferred single life annuity. Participants were allowed to annuitize 0, 25%, 50%, 75%, or 100% of their retirement savings at age 60 to purchase the annuity. They could also choose to start payments when they were either 65, 75, or 85. To determine the values for the annuity income paths, we priced deferred annuities from immediateannuities.com. Specifically, we assumed the individual was currently 60, male, and living in California. The annuity was for life and had a 3% cost of living adjustment with no additional riders. We selected the New York Life offerings since they had relatively high payouts compared to the other offerings and an A++ rating. A screenshot of the guaranteed income decision page (separate condition) is provided in Figure 1c.

Experimental Conditions. This study is a between-subjects design with two conditions to test the impact of broad versus narrow framing on retirement income decisions. Participants either saw income and wealth graphs that added values across the three financial products (the aggregated condition) or saw only outcome information specific to the financial product they were making decisions about (the separate condition). In the aggregate condition, participants saw estimated income graphs that would add together their Social Security payments, their estimated retirement savings withdrawals, and their yearly annuity payments. They also saw a wealth graph at all times, and the associated probability of running out of savings. Those in the separate condition saw three different income graphs as they switched between the financial products. For example, when viewing the Social Security tab, people would only see the income from Social Security, and not the income from retirement savings or guaranteed income. Since people have little control over the wealth associated with their Social Security payments and annuity income, we did not show participants estimated wealth graphs while viewing the Social Security or guaranteed income decisions in the separate condition in Experiment 1. When people in the separate condition viewed the retirement savings product they saw a wealth graph over time, while those in the aggregate condition saw a wealth graph when making decisions for all three financial products. This element was changed in Experiment 2, and participants in the separate condition saw a wealth graph when making decisions for all three financial domains. A screenshot of the Social Security decision within the aggregated condition is provided in Figure 1d.

Feedback. As noted above, on the right side of the screen for each decision space, people were shown important outcome feedback for their decisions depending on the assigned condition. This outcome feedback had three important components: two graphs and one probability. Both graphs went from ages 60 to 100. The top graph showed estimated yearly income and the bottom showed estimated wealth at the start of each year (note that the wealth graph was only displayed for participants in the separate condition of Experiment 1 when making savings decisions). Participants saw the median withdrawals and retirement savings balance for each age from the bootstrap samples we calculated for each combination of choices. Below the graphs was displayed the probability that they would have a non-zero retirement savings balance at age 85. All values shown to participants were adjusted for anticipated inflation, in this case a constant 2.7%.

Psychological Outcomes. Right after completing the decumulation planning task people were asked to answer a series of questions about how they felt. These questions were meant to capture people's subjective evaluation of their chosen outcome, stress around making a decumulation plan, confidence in making a reasonable decumulation plan for themselves, likelihood of making a decumulation plan for themselves, motivation to save for retirement, and subjective knowledge of retirement planning. A final question asked people if they would want to use a similar tool when making a decumulation plan for themselves.

Demographics and Psychographics. Heterogeneity plays an important role in financial decision making. For example, basic demographic factors such as age, sex, education, income, marital status, dependents, and ownership of 401K or pension are either related to actual claiming age or claiming age intentions (Knoll 2011). Risk and time factors, such as loss aversion and temporal discounting, have also been found to affect decumulation decisions such as Social Security claiming and annuity choice (Shu, Zeithammer, and Payne 2016; Schreiber and Weber 2016). We thus collected several demographic and psychographic measures as controls for our data. Participants were asked to complete a discount elicitation task using a nine-row matrix of binary choices with one option being for money tonight and the other for money in one year (Harrison, Lau, and Williams 2002). They then completed a 10-question task using gambles to elicit a loss aversion parameter. Finally, they gave their birth year, their annual household income[1], how much they have in retirement savings[2], education level, and number of people in their household. Cloud Research provided gender information for each respondent.

Experiment 1

Sample. We recruited 605 participants from Amazon Mechanical Turk (AMT), using their panel feature to screen for people ages 40 to 63. Previous research has found that AMT has demographics that are reasonably similar to the general population in the US (Berinsky, Huber, and Lenz 2012). Importantly, a variety of experiments and correlational studies have found similar results using participants recruited from AMT as compared to participants recruited from

[1] The options were $0-$15,000, $15,001-$25,000, $25,001-$35,000, $35,001-$50,000, $50,001-$75,000, $75,001-$100,000, $100,001-$150,000, and $150,001+

[2] The options were $0-$10,000, $10,001-$20,000, $20,001-$35,000, $35,001-$50,000, $50,001-$80,000, $80,001-$120,000, $120,001-$200,000, $200,001-$300,000, $300,001-$450,000, $450,001-$600,000, $600,001-$1,000,000 and $1,000,001+

representative or probability samples (Coppock 2018; Mullinix et al. 2015; Snowberg and Yariv 2018).

To ensure that our stimuli were relevant for our sample (e.g., not college students, and not already receiving Social Security retirement benefits), participants were prescreened to be between the ages of 40 and 63. This screening process was imperfect, so we have 13 people who report birth years that put them outside of this range. Based on the nature of the task we targeted people who were more likely to have thought about creating a decumulation plan but also unlikely to have started decumulation. Participants were given a brief overview of the study and then told about the features of the three main financial products in the tool. Given the negative associations with the word annuity, we opted to use the term guaranteed income throughout the study. We asked participants to image that they were single, currently 60, retiring at 62, making $55,000 a year before taxes, have $250,000 in retirement savings, and own a home free and clear. Participants were also screened to make sure they understood the instructions, using a comprehension check of three simple questions about how the financial products described within the study worked. People were given two attempts to get the questions correct before they were screened out. Based on this comprehension test 178 people were excluded. Additionally, 28 people started but did not finish the study. After exclusions and attrition, we have 399 participants (median age = 48, 44.9% female).

Study Results

Average retirement income per condition. We begin by evaluating how overall retirement income changes by condition. We expected that individuals in the aggregate condition, who are able to have a more complete picture of their possible outcomes while making choices, would be more likely to take tradeoffs between the products into account. To test for differences in average retirement income per condition, we consider four different measures: average monthly income across ages 62 to 100, NPV of income from 62 to 100, standard deviation of income from 62 to 100, and average year-by-year differences in income for those years. Note that income for years 60 and 61 is excluded since no decisions within the tool affect that income, even though it was displayed on the income graphs seen by participants.

Graphs of the average estimated income per year for each condition are shown in Figure 2. Observation suggests that the average income path for participants in the aggregated condition is smoother than for those in the separate condition. This is borne out in regression analyses of the income measures. Table 1 shows the results of four regression models. The first two output columns represent a model using average retirement income as the dependent variable, run either with or without controls. For this measure, the effect of condition is marginally significant when controls are included; average income is slightly lower for participants in the aggregate condition. If the sequence of average income is instead captured as an NPV with future years discounted, the effect of condition is now positive (i.e., higher NPV for the aggregate condition), but the effect is again only marginally significant, economically trivial ($2 to $3), and only appears when controls are not included.

Of more interest is what happens to the variability of income in the two conditions. Table 2 provides the output of four regression models with dependent variables capturing the smoothness of income over time. The dependent variable of the first two columns is the standard deviation of income in years 62 through 100. The model shows a significant decrease in the average variability of income sequences selected by participants in the aggregate condition ($b = -715.05$, $t(393) = -2.84$, $p = 0.005$), which also holds when adding controls ($b = -967.48$, $t(362) = -3.72$, $p < 0.001$). In an effort to include the temporal nature of the data we also constructed a measure of the absolute difference in expected income from each year to the previous year and then averaged those differences for each person. For this outcome we see that on average, participants in the aggregate condition had lower average lagged differences ($b = -108.9$, $t(391) = -3.88$, $p < 0.001$). This difference holds when we control for demographic and psychographic variables ($b = -118.61$, $t(362) = -4.07$, $p < 0.001$).

Inspection of the most commonly selected scenarios within each condition aids understanding of these differences. Table 3a shows the five most common scenarios for the aggregated condition, while Table 3b shows the five most common scenarios for the separate condition. Visual inspection suggests that the commonly chosen scenarios in Figure 2a often have a smoother consumption path than those in Figure 2b, suggesting that study participants were most likely to make a set of decisions that led to lower income variation in the aggregate condition where they were able to see all outcomes at once while making their choices. It's also worth noting that OASI claiming appears to more often start at age 62 in the aggregate condition (the two most common scenarios) but often starts later in the separate condition, where only one scenario starts claiming at age 62. In the sections below, we explore these patterns in more depth to see whether there are significant differences in choices between the conditions.

Claiming decision. We next turn to evaluating the outcomes for each of the three financial product domains, starting with the OASI benefits claiming decision. Recall that participants only made one decision in this space of what age to begin claiming Social Security benefits, with choices from age 62 to age 70. Regression results for this dependent variable are provided in Table 4, and a graphical display of the choices per condition (including mean and 95% confidence intervals) is provided in Figure 3. For this outcome, we find that on average people in the aggregate graph condition claimed nine months earlier than those in the separate condition (b = -0.75, t (397) = -2.67, p = 0.008). This impact held when controlling for demographics and psychographics (b = -1.06, t (365) = -3.69, p <0.001).

Savings decisions. When comparing the selection of retirement savings withdrawal strategies, we do not find evidence to suggest a difference between the selections in the aggregate and separate graph conditions ($\chi^2(2) = 4.2$, p = 0.123). In both conditions, the majority of participants selected an increasing withdrawal strategy, with 51% of participants choosing it in the separate condition and 49% choosing it in the aggregate condition. In a regression model that also controls for demographic and psychographic variables, we do find a significant difference in the proportion of people in the aggregate condition choosing the flat withdrawal sequence as opposed to the decreasing sequence (b = 2.39, z = 2.32, p = 0.02).

For the choice of whether or not to take extra withdrawals from savings in the years before claiming Social Security, we see a marginally significant difference in the proportion of people taking this option between the aggregate and separate graph conditions (b = -0.07, t(397) = -1.75, p = 0.081). When we control for demographic and psychographic variables, we see that on average people in the aggregate condition choose this option significantly less frequently (b = -0.092, t(365) = -2.13, p = 0.034). This may be partly the result, however, of participants in the aggregate condition claiming their benefits earlier, thus reducing the need to withdraw from savings to bridge those years of delay.

The final choice within the retirement savings space was about the level of risk participants were willing to take with their retirement savings, as captured through the ratio of stocks versus bonds they wished to take for their investment portfolio. We do not find evidence to suggest there is a relationship between condition and the risk allocation choice ($\chi^2(2)$= 2.43, p = 0.297); a breakdown of the choices per condition is provided in Table 5. In both conditions, the majority of participants selected the middle risk allocation option (75% in the separate condition, and 77% in the aggregate condition). When controlling for demographics and psychographics we do not see evidence to suggest that people on average choose different final risk allocations by condition (medium vs low stock: $b = 0.584$, $z = -1.2$, $p = 0.229$); medium vs high stock: $b = 0.921$, $z = -0.3$, $p = 0.765$).

Guaranteed income decisions. The final product domain was the set of choices for guaranteed income. In this domain, individuals were asked what proportion of their retirement savings they would be willing to use to purchase an annuity, and if they did purchase one, they were asked what age they would want income to begin (deferred to ages 65, 75, or 85). People in the aggregate condition on average had significantly lower annuitization rates of their retirement savings ($b = -8.76$, $t(397) = -3.21$, $p = 0.001$), see Figure 4 for condition means and 95% confidence intervals. When controlling for demographic and psychographic variables we see that this general relationship holds ($b = -9.84$, $t(365) = -3.53$, $p < 0.001$); results for these regression models are provided in Table 6. For analysis of the annuity start years we only include people who have annuitized some part of their retirement savings. Conditional on buying an annuity, we see no significant differences in the average start year of the payments between the aggregate and separate conditions ($b = -0.34$, $t(186) = -0.42$, $p = 0.68$). In both conditions, the majority of participants desire to start their annuity income at age 65.

Confidence and Engagement. The final set of measures concern how the decision tool affect individuals' confidence in their retirement planning and engagement with the task. These are secondary outcomes, and have less economic impact on the lives of retirees, but are an important indicator of how much a retirement income tool such as this one may be adopted by potential retirees to assist with their decision making process.

As described in the methods section, we collected several psychological outcome variables from participants at the end of the retirement income planning task. We asked about anxiety ("How stressed do you feel about creating a plan for yourself of income in retirement?" with options from 1: Not at all stressed to 4: Very stressed), confidence in a plan ("How confident are you in your ability to make a reasonable plan for income in your retirement?" with options from 1: Very confident to 4: Not at all confident), likelihood of making a plan ("How likely are you to make an explicit plan for income over time in retirement?" with options from 1: Extremely likely to 5: Extremely unlikely), confidence in their selections ("How confident are you that the plan for income and wealth over time you created with this tool is good?" with options from 1: Very confident to 4: Not at all confident), motivation to save for retirement ("How motivated do you feel to save for retirement?" with options from 1: Very motivated to 5: Very demotivated), subjective knowledge about retirement income planning ("In general, how knowledgeable do you feel about retirement income planning?" with options from 1: Very knowledgeable to 4: Not at all knowledgeable), and their likelihood of using a decision tool such as this one to assist in their planning ("Would you want to use a similar tool for planning your own retirement income?" with options from 1: Definitely to 4: Not at all). We checked for differences in each of these measures by condition; results are shown in Table 7. Of these measures, the only one that shows a significant difference per condition is that participants in the aggregate condition expressed higher confidence that they made a good plan with the tool (1.78 vs. 1.54; $b = 0.24$, $t(397) = 2.83$, $p = 0.005$). This difference held when controlling for demographic and psychographic variables ($b = 0.26$, $t(365) = 2.98$, $p = 0.003$).

To capture engagement, we measured the amount of time each participant spent using the tool itself. A more engaging tool should lead to more time spent due to more exploration of different options within each decision space. However, we do not see any statistically significant differences in the amount of time spent on the decumulation tool or the number of changes in decumulation options (which includes asset class changes for both conditions) by condition.

Experiment 2

To further test and extend the findings of Experiment 1, we ran a second version of the experiment using the same tool with a new population of participants. As described below, several key changes were made to the tool based on feedback from the first experiment. We also sought to better understand how well the decisions and outcomes that participants selected in the tool reflected their individual preferences. In Experiment 1, we had no benchmark of what was "optimal" for each participant, and evaluated outcomes such as average monthly income, income variability, and claiming age without knowing whether those outcomes matched to the utility functions of the participants. Thus, in Experiment 2, we run a second stage of the experiment with each participant in which they evaluate their own outcome path (not identified as such) against several other possible outcomes. The results of this evaluation provide us with a set of revealed preferences with which to judge the individual optimality of the choices each participant made in the first stage. Based on the results from our first experiment we also pre-registered (http://aspredicted.org/blind.php?x=yp275c) a specific set of hypotheses to test in our second experiment.

Decision Tool. In our second study we made two changes to the decision tool. First, people in the separate condition saw estimated retirement savings wealth across all three asset classes. Previously people in the separate condition would not see a wealth graph when making decisions about OASI or guaranteed income. Effectively this makes it so the only difference between the conditions is whether or not the income graphs show aggregated income or income corresponding to each financial product. Second, in every graph we started the estimated dollar values at age 62. None of the decisions made by participants impacted wealth or income at ages 60 and 61 so they are excluded from the feedback graphs.

Method. After completing the retirement decision-making tool, participants were asked to do a follow-up study at least two days after they completed the decision tool. In this second wave people were first asked to rate six selected aggregate income graphs as well as the aggregate income graph for their selected scenario in wave one. They were not told that one of the graphs represented their decisions from the first wave, and we expected that most participants would not recognize it as such when set among the other graphs. The six selected scenarios were chosen to have a clear spread in the tradeoff between standard deviation of income over time and NPV of total income. Two of the six scenarios have very similar NPVs and standard deviations, but one annuitizes 25% of retirement savings wealth and the other does not annuitize any retirement savings. With these two options we hope to elicit a revealed preference for annuities in a realistic tradeoff. Participants were asked, “How much would you like to get the income shown in the plot above at each age?” on a scale of 1 = Not at all to 10 = Very much. For each graph participants were told all of the options selected for the financial products, such as when the scenario would claim OASI benefits, as well as the probability of running out of retirement savings by age 85. The order of the scenarios was counter balanced.

Next, participants explicitly ranked each of the seven scenarios. Each scenario was represented by the aggregate income graph along with the six financial product selections and the probability of running out of retirement savings by age 85. Presentation order of the scenarios was randomized. Finally, participants rated four statements about the importance of certain considerations when selecting a decumulation plan. The four statements were about running out of money, having enough money for late life medical expenses, having more income early on to spend on travel, and passing money to heirs.

Sample. In the first wave of experiment 2, we have 600 participants from AMT, using the panel field in Cloud Research to select for participants who were 40 to 62. A total of 639 people were recruited and 39 were excluded because they tried to complete the study on a mobile device. Of the 600 people, 69 are outside our specified range with the vast majority being one year away from the target range. An additional 75 participants were excluded from the main analyses because they did not get all three comprehension questions correct on their first or second attempt. In each model below we excluded participants with absolute standardized residuals greater than four. Data collection for the second week started one week after the first wave launched. We kept at least two days between when someone completed the first wave and the second wave of the study. A total of 390 people completed the second wave for a response rate of 65%. People in the second wave had to get all of the comprehension questions correct on their first or second attempt to be included in the primary analysis.

Results

Estimated Retirement income per condition. As before we test whether condition had an impact on the estimated income using the same measures as before. In Figure 5 the average income per year is plotted for the aggregate and the separate conditions. Individual sequences selected by participants are plotted as transparent grey lines. In contrast to the first study, we do not see a significant difference between conditions for both metrics of variability in the sequences selected (see Table 8 for details). The average standard deviation between conditions does not differ, with ($b = 232.1$, $t(506) = 1.08$, $p = 0.3$) or without controls ($b = 182.28$, $t(520) =$

0.85, p = 0.4). When looking at the lagged absolute differences there also is not a statistically significant difference between conditions (b = 7.4, t(522) = 0.3, p = 0.77), which does not change substantially when we add controls (b = 14.01, t(508) = 0.55, p = 0.58).

Similar to experiment 1, the two variables capturing the value of the income sequence over time do not show differences across conditions. In the final metric, the probability of having non-zero retirement savings by age 85, there are not significant differences between the two conditions.

Claiming decision. Now we will report the results for each product choice across the three financial products. In Experiment 2 we do not find that average claiming age was significantly different between the aggregate and separate income graph conditions (b = 0, t (523) = -0.01, p = 0.99). This lack of a difference held when controlling for demographics and psychographics (b = 0.06, t (509) = 0.26, p = 0.8). Regression results for these comparisons are in Table 10.

Savings decisions. The pattern of withdrawal strategies across the two treatments does not significantly differ by condition ($\chi^2(2) = 2.77$, p = 0.25). A vast majority of people selected either the flat or increasing sequence of withdrawals. In the separate condition, the plurality, 46.9%, of people chose the flat withdrawal sequence. As for the aggregate condition, the plurality, 48.3%, chose the increasing withdrawal sequence. When adding controls, we see similar results.

People also had to choose whether or not to take extra withdrawals from their retirement savings in the years before they claimed OASI. We do not find a significant difference in the likelihood that someone takes extra withdrawals between conditions (b = 0.03, t(523) = 0.66, p = 0.51), which does not change when adding in controls (b = 0.027, t(509) = 0.71, p = 0.48). Given the lack of differences in claiming age, which we saw in experiment 1, people may not be using extra withdrawals to compensate for later OASI claiming in the separate condition.

Finally, people had to choose an investment allocation for their retirement savings. The three options, all of which mimicked a target date fund, varied in their level of stock. We see that people in the aggregate condition have a significantly different pattern of selections compared to the aggregate condition ($\chi^2(2) = 9.23$, $p = 0.01$). The difference comes from more people in the aggregate condition choosing the highest stock allocation compared to the middle stock allocation. When adding controls and using a multinomial model we see that people selected the highest stock option more frequently than the middle stock option in the aggregate compared to the separate condition (medium vs high stock: $b = 2.084$, $z = 3.17$, $p = 0.002$), while there was not a significant difference in the selection of the lowest risk option relative to the middle risk option between conditions (medium vs low stock: $b = 1.366$, $z = 0.95$, $p = 0.344$). Full regression results can be found in Table 11 and a plot of selections by condition can be found in Figure 6.

Guaranteed income decisions. People were asked whether they would like to purchase guaranteed income and when payments should start if they chose to purchase one. We do not see a difference between conditions in the average annuitization percentage ($b = -1.76$, $t(523) = -0.79$, $p = 0.43$). When adding controls this result does not change ($b = -1.67$, $t(509) = -0.75$, $p = 0.46$). Conditional on choosing to annuitize some wealth we look at the average starting age between conditions. Here we do not see a significant difference in the starting year ($b = -0.42$, $t(237) = -0.56$, $p = 0.58$). With the addition of controls we still do not see a significant difference in average starting year ($b = -0.12$, $t(226) = -0.16$, $p = 0.87$). The vast majority, 83.4% in the separate condition and 81.5% in the aggregate condition, of people in both conditions chose to start receiving payments at 65.

Confidence and engagement. The same set of confidence and engagement measures as in Experiment 1 were analyzed for this experiment. We do not find significant differences between conditions in the ratings for all of the confidence measures (see Table 13 for means and standard deviations by condition). As for the engagement metrics, both the number of minutes spent on the decumulation tool and the number of times people made changes to their selection (which includes asset class changes for both conditions), we did not see any significant differences between conditions.

Scenario rating and ranking. In the follow-up survey people were asked to rate, "How much would you like to get the income shown in the plot above at each age?" The responses were on a 1 = Not at all to 10 = Very much scale. People were shown the aggregate income graph of the scenario that they selected in the first wave, along with six other selected scenarios. On average people in the aggregate and separate conditions do not rate their chosen graph differently ($b = -0.38$, $t(318) = -1.51$, $p = 0.13$). When adding in controls there is a marginally significant difference with people in the aggregate condition rating their scenarios lower on average than those in the separate condition ($b = -0.42$, $t(306) = -1.65$, $p = 0.099$). Full regression results can be found in Table 14. In addition to rating each scenario, we asked people to explicitly rank each of the seven options. The average rank given by people in the two conditions is not significantly different ($b = -0.13$, $t(318) = -0.6$, $p = 0.55$).

When looking across conditions we see the average rating for their scenario is 6.4. Two other scenarios have higher scores (6.7 and 6.5), while the other four have average ratings less than 6. About 31% of people gave their highest rating to their scenario. The scenario with the highest average rating has a claiming age of 63, increasing withdrawals, extra withdrawals before beginning OASI benefits, the highest stock allocation, and no annuitization. The second highest scenario has a claiming age of 66, increasing withdrawals, extra withdrawals, the highest stock allocation, and no annuitization.

Discussion and Conclusion

As consumers approach retirement, they are faced with many difficult decisions regarding decumulation. Typically, these decisions – when to claim Social Security OASI benefits, whether to purchase an annuity, and when (and how much) of their retirement savings they should withdraw – are done in a siloed fashion. Although consumers may intuitively understand that all of these decisions are part of one overall decumulation strategy, it can be cognitively taxing to balance the effect of each independent decision on one's overall financial picture in retirement. As a result, we created and tested the effectiveness of a tool that allows consumers to make such decisions in an aggregate, rather than separate, manner.

In Experiment 1, we found that making decisions in aggregate had several effects. Perhaps the most robust and notable one, however, was that the participants who used the aggregate version of the tool had significantly smoother consumption patterns (marked by lower variability over time) than participants who used the separate version of the tool. Given decreasing marginal utility of consumption, all else equal, life cycle theory (Modigliani, 1966) suggests smoother consumption patterns are preferred to more variable patterns when sacrifices in NPV are not required. More generally, it is possible (if not likely) that the aggregate presentation permits consumers not just to maximize smoothness, but rather to choose the most-preferred consumption stream independent of the variability of its components, a hypothesis that led to direct testing of participant preferences in Experiment 2. Given that workers sometimes exhibit a preference for increasing consumption patterns over flat consumption patterns at the expense of NPV (e.g., Loewenstein and Sicherman, 1991), decision tradeoffs within the tool may not lead to maximizing smoothing, but rather seeking to maximize a combination of smoothness and growth. Our initial findings that participants were more confident in their decisions in the aggregate condition rather than the separate condition lends some credence to the notion that they were better able to choose the aggregate consumption pattern that more closely matched their preferences.

While participants in the aggregate condition were better able to smooth their retirement income, we also see differences in the particular financial decisions they made. In particular, individuals in the aggregate condition indicated interest in claiming SSA retirement benefits approximately a year earlier than those in the separate condition. Individuals in the aggregate condition were also 10 percentage points less likely to choose to annuitize part of their wealth than those in the separate condition. However, there was no significant difference in the savings decisions between the two conditions, with most participants in both conditions choosing increasing withdrawals and moderate risk. Since individuals have a variety of options for arranging these decisions to achieve the income path they prefer, and we had no explicit predictions for how broad bracketing would affect these individual decisions, the particular reasons for these differences are unclear.

In Experiment 2, we attempted to extend these results by repeating Experiment 1 with a new set of participants with similar demographics. We made some changes to the tool, the largest being adding a display of wealth over time to all decision screens in the separate frame condition. Our a priori prediction was that the change in the wealth graphs would not be significant enough to change participants' behavior in most decisions but this turned out to not be the case. The differences between conditions in income smoothing, SSA benefit claiming, and annuitization found in Experiment 1 did not hold in this version of the experiment, and both frame conditions had similar results for these decisions. Whether the lack of significant differences are a function of the changes to the tool or simply experimental noise is an open question, and one we will examine further in additional research. We did find a difference between the conditions in a new area: the amount of risk selected for retirement savings was higher for individuals in the aggregate frame condition. Experiment 2 also went further in trying to understand whether the outcomes selected by participants reflect stable underlying preferences or are a dynamic function of the experience of using the tool to explore a variety of settings and aggregated options. Participants in this experiment were asked to both rate and rank their selected outcomes against several other possible outcome scenarios. While there was no difference in the ranking of the outcomes per condition, it is notable that participants ranked their own scenarios as the third best out of the seven possible outcomes. Analysis is ongoing to use these ratings to decompose the primary attributes that drive these preferences, to generate a better normative model of how individuals value the components of an ideal retirement income scenario.

This work is preliminary and it contains limitations. First, these decisions are necessarily hypothetical, and it is possible that consumers who are faced with consequential decisions about their own money and their own retirements may use the tool in different ways. To the extent that this early work shows some promising effects on decision-making outcomes, it is our hope that future work employs this tool with consumers who are making consequential decisions. Second, the research participants on the Mechanical Turk platform may not have had as much experience making these sorts of decisions with their own savings and retirement plans. Future work should assess the effectiveness of this tool with consumers who have already accumulated savings and/or are actively considering retirement. Yet it is important to note that, by its very nature, consumers do not have substantial experience making decumulation decisions at the time they must make them: learning from past mistakes may be impossible if those mistakes foreclose future decision paths. Furthermore, the high stakes involved in such decisions implies that early mistakes have substantial impact. This combination of minimal chances for learning and high stakes contributes to an environment where decision making quality is likely to suffer.[3] Planners, simulations, and other decision aids have a critical role to play in such highly consequential one-shot decisions since they introduce two of the main ingredients needed for learning: frequent practice and immediate feedback. While not a perfect solution, we hope that tools such as the one tested here provide opportunities for retirees to build retirement income solutions that best match their preferences.

[3] "If learning is crucial, then as stakes go up, decision making quality is likely to go down." Thaler, 2015, pg. 51.

References

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics* 110(1) 73–92.

Benartzi, Shlomo, and Richard H. Thaler. 1999. "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments." *Management Science* 45 (3): 364–81.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3), 351-68.

Coppock, Alexander. 2018. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 1-16.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk-Taking and Evaluation Periods." *The Quarterly Journal of Economics* 112(2) 631–645.

Haigh, Michael S., and John A. List. 2005. "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis." *The Journal of Finance* 60(1) 523–534.

Hlavac, Marek. 2018. stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decisions under Risk." *Econometrica* 47(2) 263–291.

Keren, Gideon. 1991. "Additional tests of utility theory under unique and repeated conditions." *Journal of Behavioral Decision Making* 4(1) 297–304.

Klos, Alexander, Elke U. Weber, and Martin Weber. 2005. "Investment Decisions and Time Horizon: Risk Perception and Risk Behavior in Repeated Gambles." *Management Science* 51(2) 1777–1790.

Langer, Thomas, and Martin Weber. 2001. "Prospect Theory, Mental Accounting, and Differences in Aggregated and Segregated Evaluation of Lottery Portfolios." *Management Science* 47(5) 716–733.

Litman, Leib, Jonathan Robinson, and Tzvi Abberbock. 2017. "TurkPrime.com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences." *Behavior Research Methods* 49 (2), 433-42.

Loewenstein, George, and Nachum Sicherman. 1991. "Do workers prefer increasing wage profiles?" *Journal of Labor Economics 9*(1), 67-84.

Modigliani, Franco. 1966. “The life cycle hypothesis of saving, the demand for wealth and the supply of capital.” *Social Research* 160-217.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. “The Generalizability of Survey Experiments.” *Journal of Experimental Political Science* 2 (2), 109-38.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. “Choice Bracketing.” *Journal of Risk and Uncertainty* 19 (1): 171–97.

Redelmeier, Donald A., and Amos Tversky. 1992. “On the Framing of Multiple Prospects.” *Psychological Science* 3(3) 191–193.

Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz. 1997. “The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test.” *The Quarterly Journal of Economics* 11(2) 647–661.

Thaler, Richard H., and Eric J. Johnson. 1990. “Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice.” *Management Science* 36(6) 643–660.

Webb, Elizabeth C., and Suzanne B. Shu. 2017. “Is broad bracketing always better? How broad decision framing leads to more optimal preferences over repeated gambles.” *Judgment and Decision Making* 12 (4).

Wedell, Douglas H., and Ulf Böckenholt. 1994. “Contemplating Single versus Multiple Encounters of a Risky Prospect.” *The American Journal of Psychology* 107(4) 499–518.

Figure 1a. Screenshot of Social Security benefits claiming decision (separate condition).

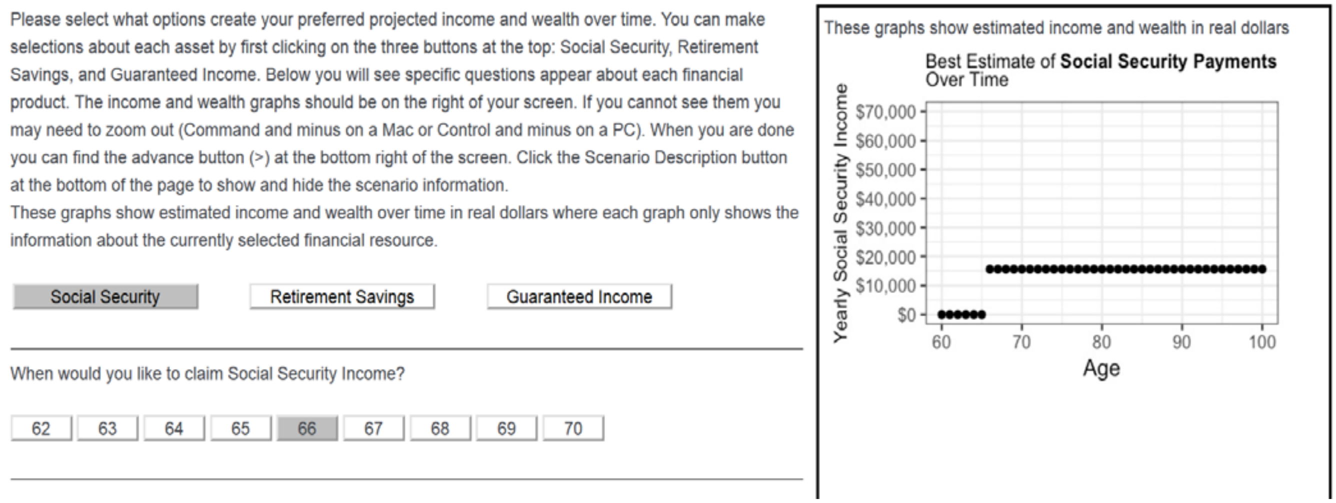


Figure 1b. Screenshot of retirement savings decisions (separate condition).

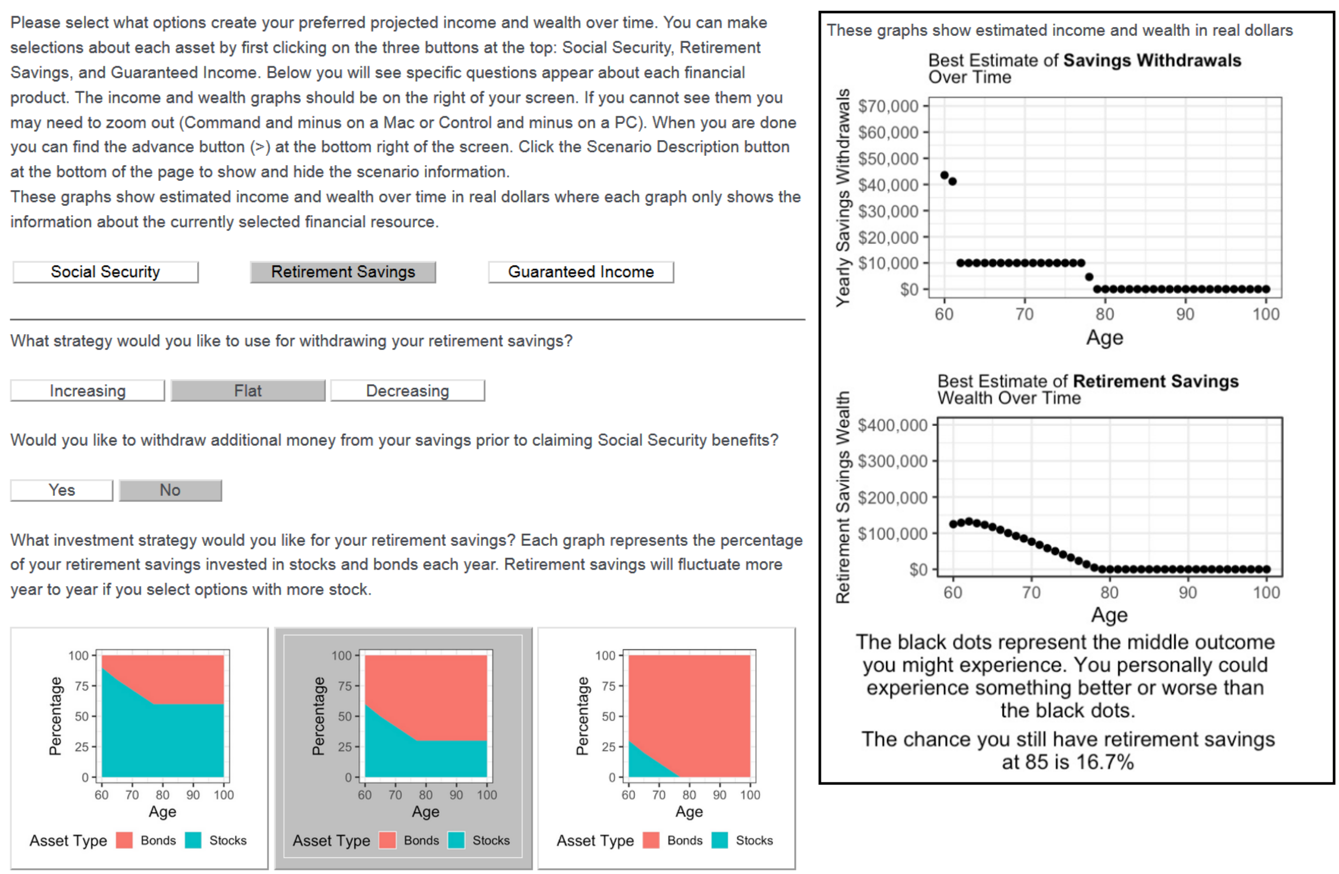


Figure 1c. *Screenshot of guaranteed income decisions (separate condition).*

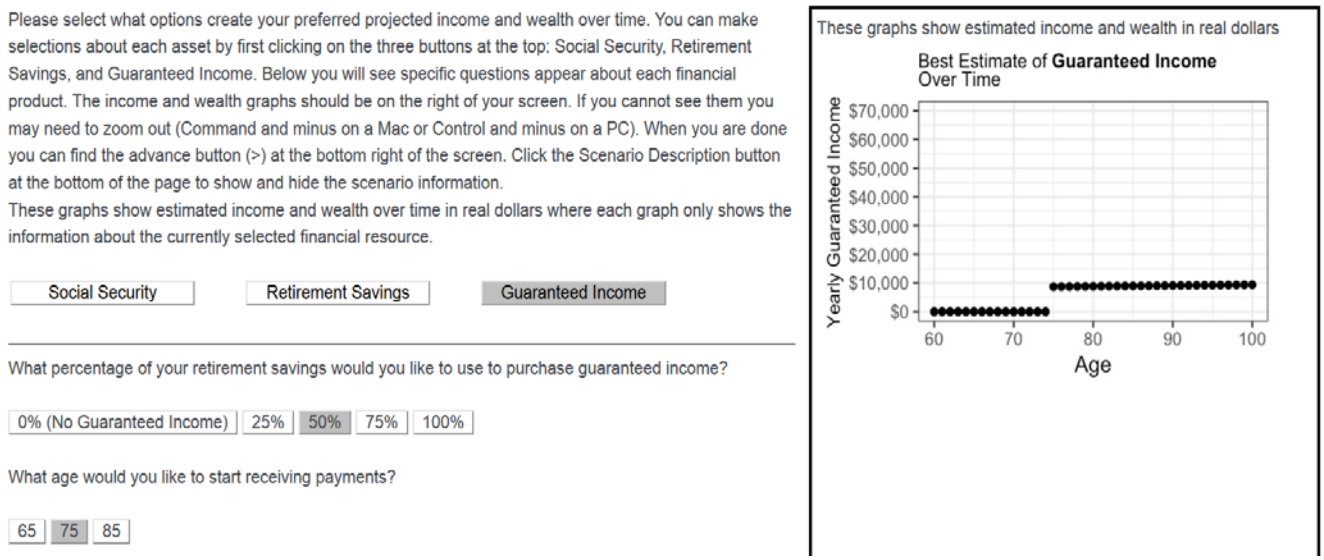


Figure 1d. Screenshot of Social Security benefits claiming decision (aggregate condition).

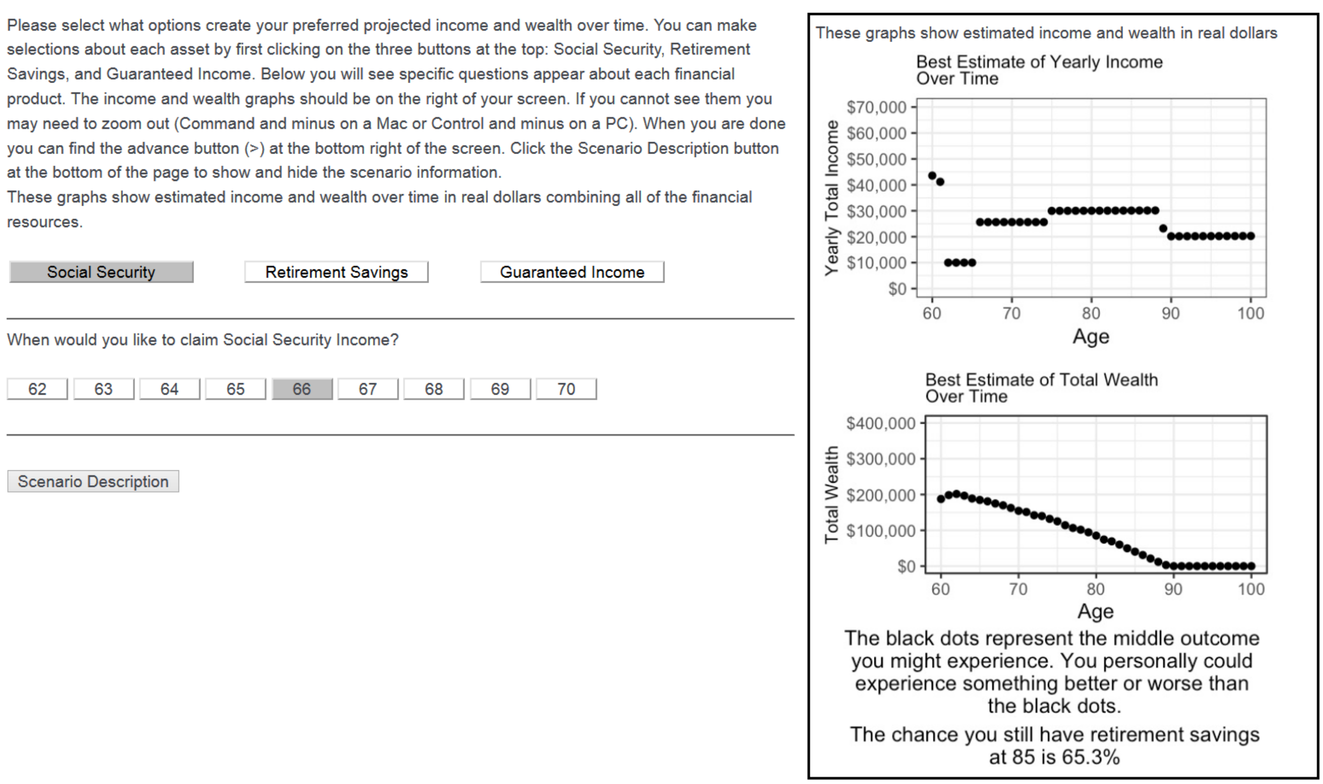
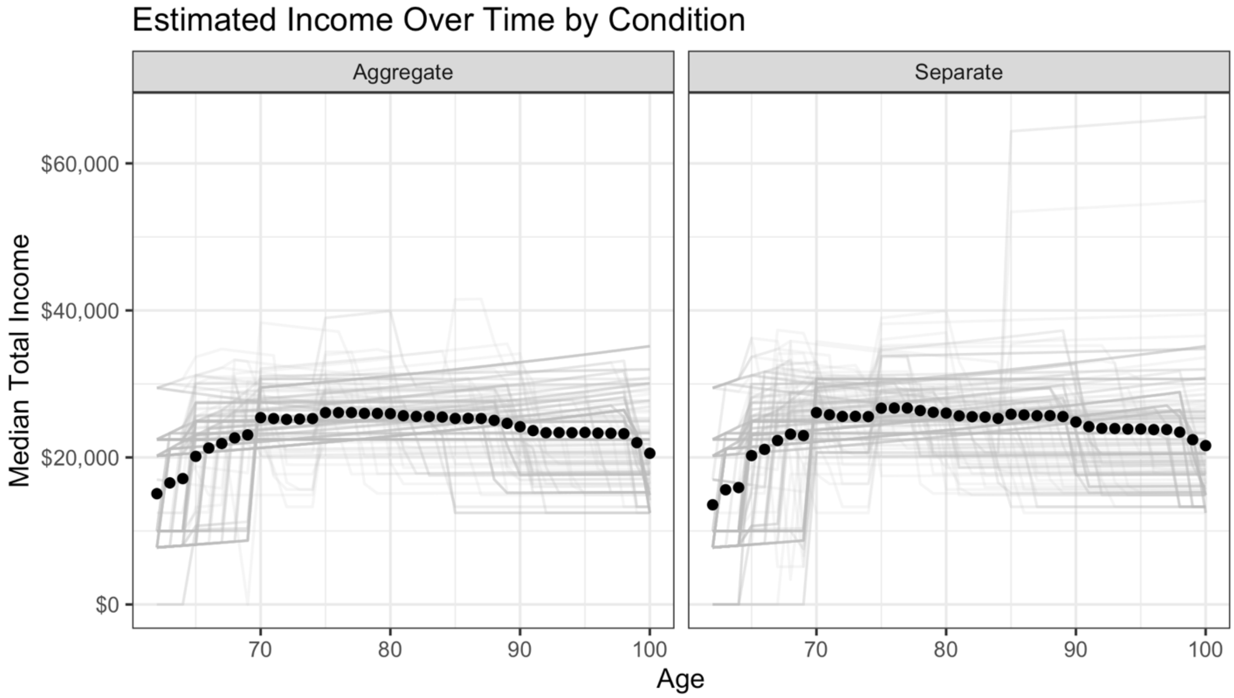
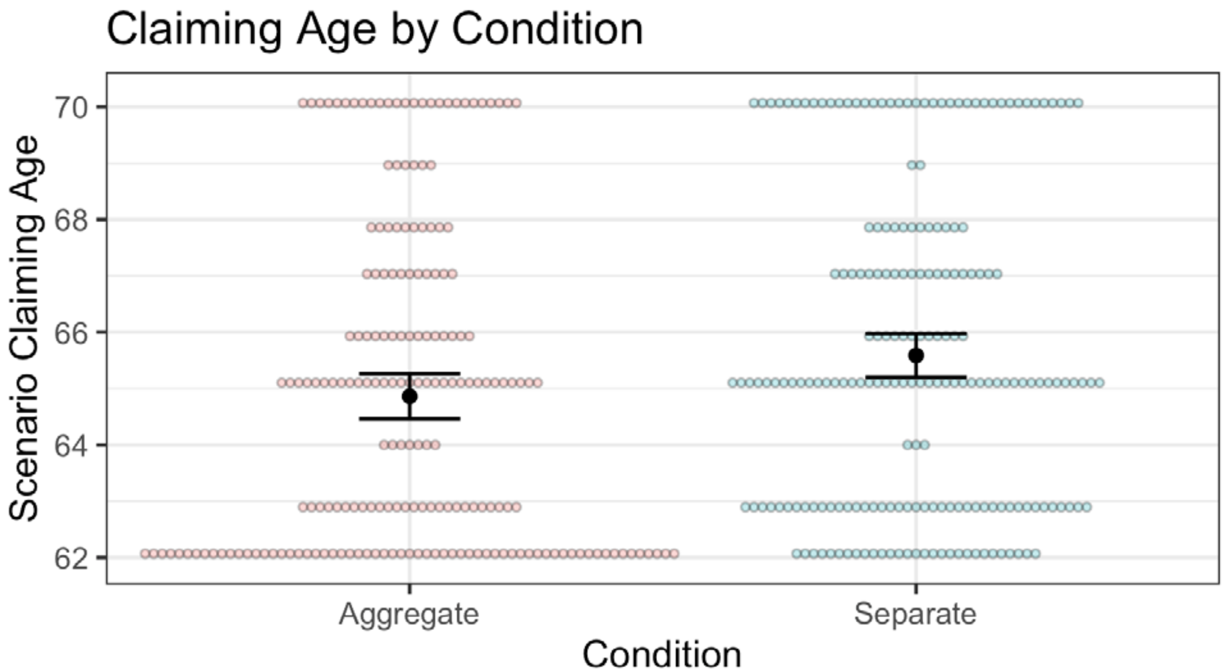


Figure 2. Experiment 1: Average estimated income from ages 62 to 70, per condition



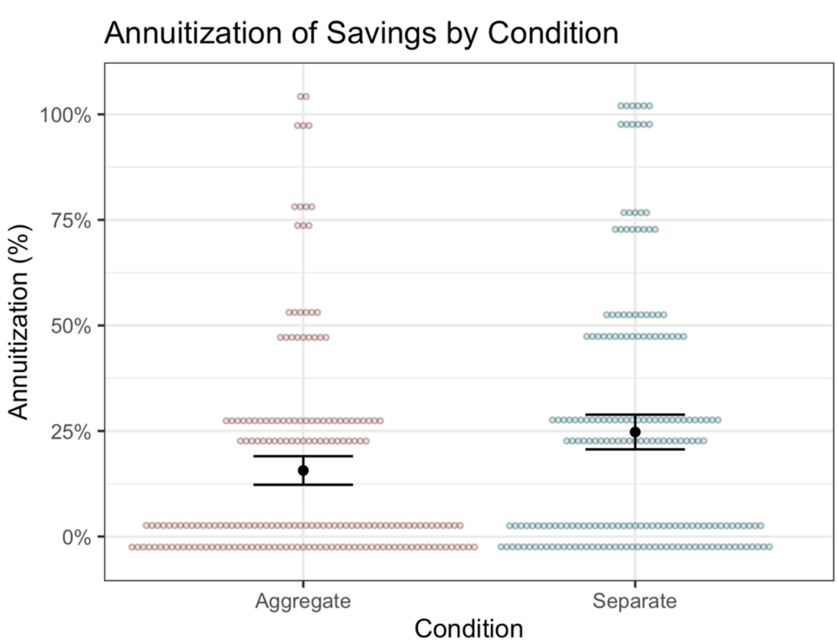
Note: Figure 2a is on the left showing the income sequences selected in gray and the average total income for people in the aggregate condition. Figure 2b is on the left showing the income sequences selected in gray and the average total income for people in the separate condition

Figure 3. Experiment 1: Claiming age by condition.



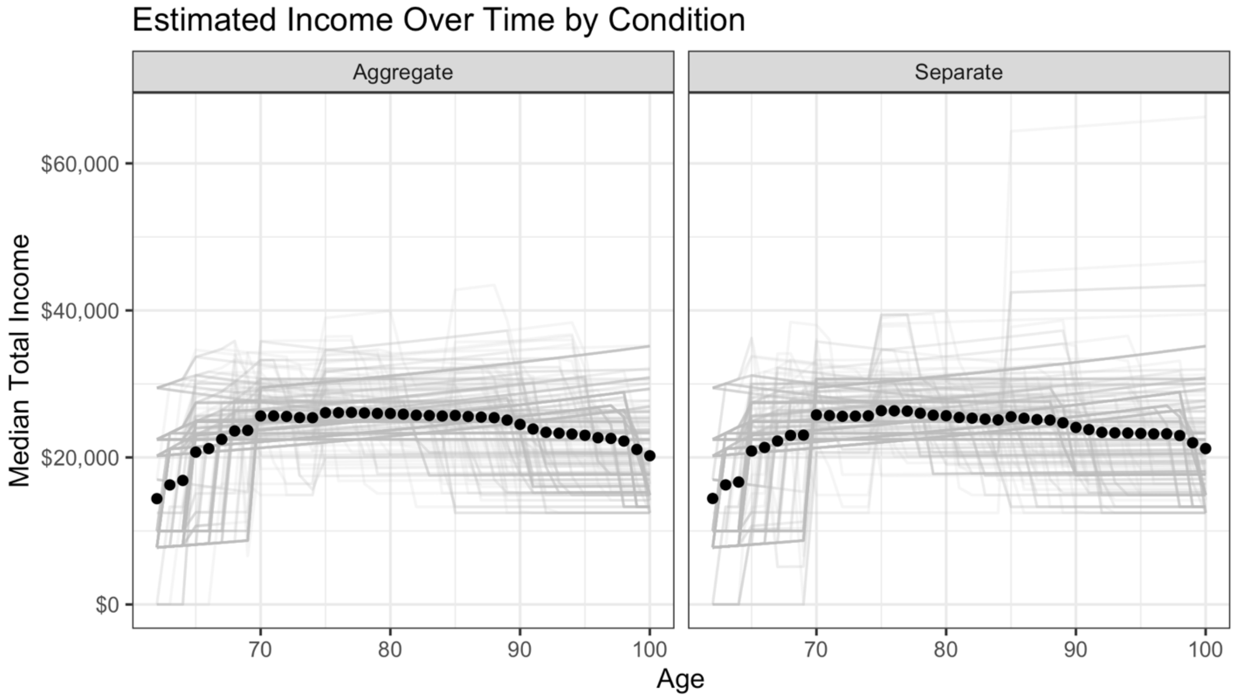
Note: Error bars represent 95% confidence intervals

Figure 4. Experiment 1: Annuitization by condition



Note: Error Bars represent 95% confidence intervals

Figure 5. Experiment 2: Separate Asset Condition



Note: Figure 5a is on the left showing the income sequences selected in gray and the average total income for people in the aggregate condition. Figure 5b is on the left showing the income sequences selected in gray and the average total income for people in the separate condition

Figure 6. Experiment 2: Stock allocation by condition

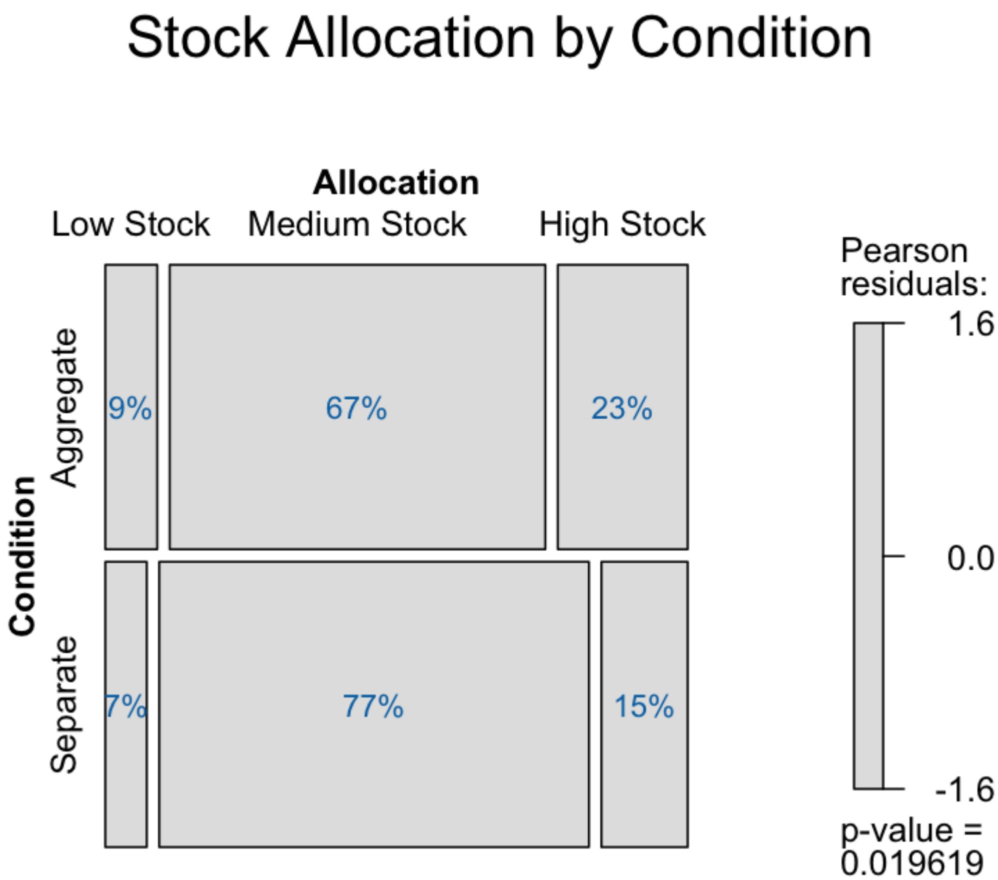


Table 1. Experiment 1: Effects of condition, with and without controls, on average income and NPV of income.

	Dependent variable:			
	Ave Estimated Income	Ave Estimated Income	NPV Estimated Income	NPV Estimated Income
Aggregate	-260.009	-396.841*	2.423*	2.062
	(174.023)	(176.339)	(1.182)	(1.139)
Discount Rate		-65.778		-0.299
		(44.448)		(0.287)
Loss Aversion		-26.773		-0.304
		(29.835)		(0.193)
Age		13.760		-0.097
		(13.775)		(0.089)
HH Income		175.087**		-1.294**
		(64.551)		(0.417)
Retirement Savings		-27.461		0.443
		(35.867)		(0.232)
Education		70.477		0.430
		(72.294)		(0.467)
HH Size		-115.030		0.969*
		(66.039)		(0.427)
Constant	24,676.410***	23,806.520***	44,810.710***	44,818.350***
	(87.011)	(914.397)	(0.591)	(5.908)
Observations	399	374	399	374
R^2	0.006	0.054	0.010	0.051
Adjusted R^2	0.003	0.033	0.008	0.030
Residual Std. Error	1,737.913 (df = 397)	1,689.344 (df = 365)	11.804 (df = 397)	10.915 (df = 365)
F Statistic	2.232 (df = 1; 397)	2.593** (df = 8; 365)	4.203* (df = 1; 397)	2.451* (df = 8; 365)

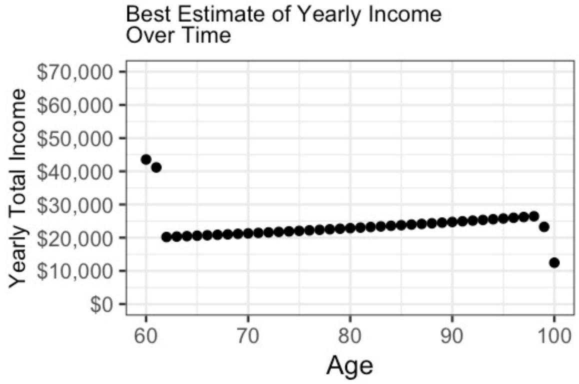
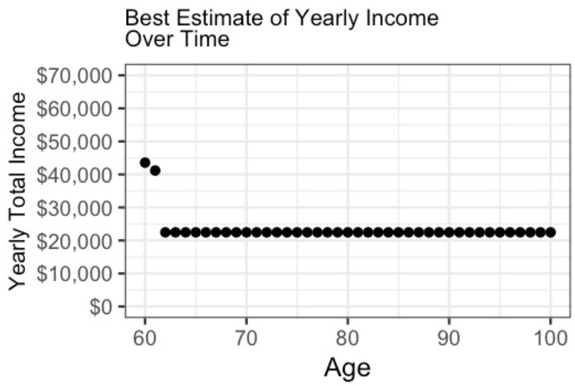
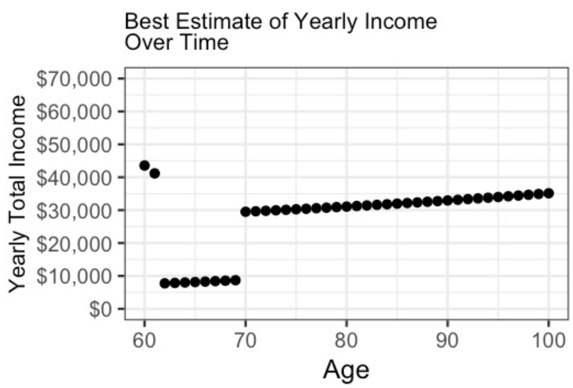
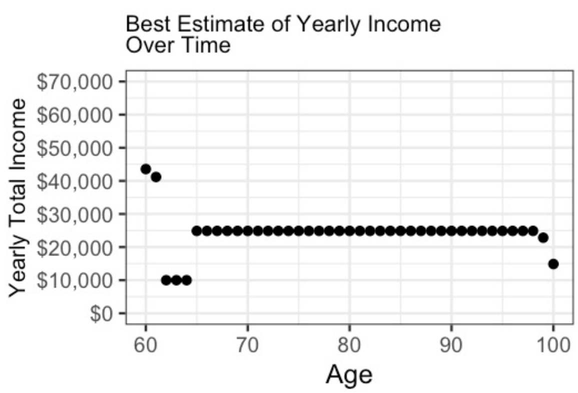
Note: ****p* < .01. ***p*<.01. **p*<.05. Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous measures.

Table 2. Experiment 1: Effects of condition, with and without controls, on variability of income.

	Dependent variable:			
	SD Estimated Income	SD Estimated Income	Abs Lag Diff Est Income	Abs Lag Diff Est Income
Aggregate	-715.045**	-967.476***	-108.898***	-118.606***
	(251.979)	(260.206)	(28.066)	(29.140)
Discount Rate		-45.982		5.382
		(65.742)		(7.357)
Loss Aversion		-3.315		0.891
		(44.086)		(4.930)
Age		11.526		1.627
		(20.297)		(2.281)
HH Income		305.270**		22.897*
		(95.001)		(10.653)
Retirement Savings		-68.643		-10.803
		(52.939)		(5.945)
Education		147.266		0.649
		(107.647)		(11.997)
HH Size		-144.785		5.002
		(97.568)		(10.910)
Constant	5,183.178***	3,554.083**	636.364***	448.602**
	(125.989)	(1,359.195)	(14.033)	(152.298)
Observations	395	371	393	371
R^2	0.020	0.069	0.037	0.062
Adjusted R^2	0.018	0.049	0.035	0.041
Residual Std. Error	2,503.981 (df = 393)	2,481.298 (df = 362)	278.189 (df = 391)	277.829 (df = 362)
F Statistic	8.053** (df = 1; 393)	3.366*** (df = 8; 362)	15.055*** (df = 1; 391)	2.968** (df = 8; 362)

Note: ***p < .001. **p<.01. *p<.05. Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous measures.

Table 3a. Experiment 1: Five Most Common Selections and Total Income in Aggregate Condition

Graph	Frequency	Selection
	13	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 62, Allocation: Target 60 %
	8	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: flat, Extra Withdrawals: no, Claiming Age: 62, Allocation: Target 60 %
	7	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 70, Allocation: Target 60 %
	6	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: flat, Extra Withdrawals: no, Claiming Age: 65, Allocation: Target 60 %

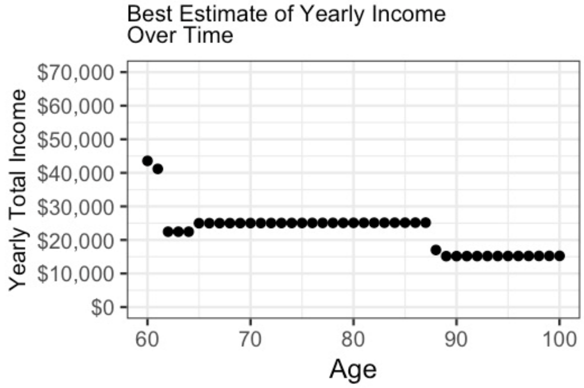
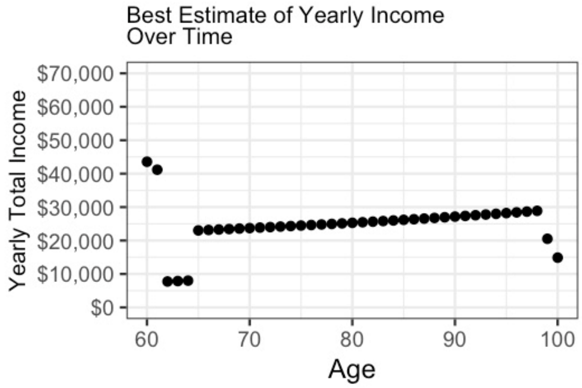
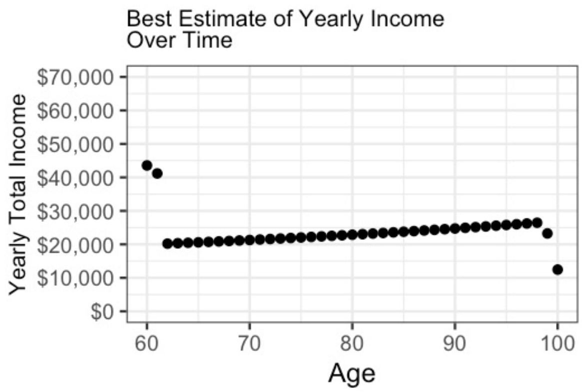
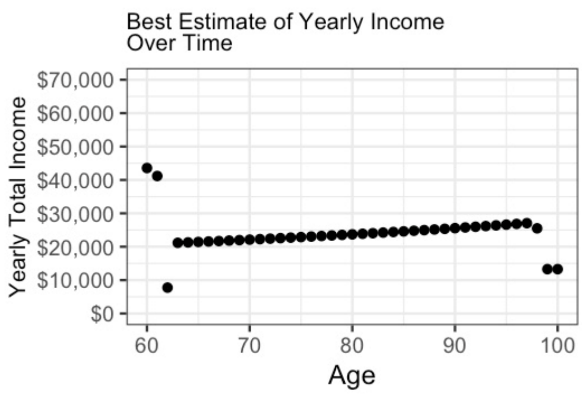
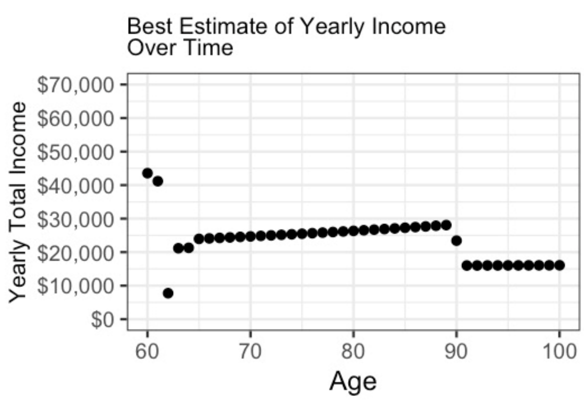
Graph	Frequency	Selection
	5	Annuitization Percentage: 25, Annuity Start Age: 65, Withdrawal Progression: flat, Extra Withdrawals: no, Claiming Age: 62, Allocation: Target 60 %

Table 3b. Experiment 1: Five Most Common Selections and Total Income in Separate Condition

Graph	Frequency	Scenario
	9	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 65, Allocation: Target 60 %
	6	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 62, Allocation: Target 60 %
	6	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 63, Allocation: Target 60 %
	5	Annuitization Percentage: 25, Annuity Start Age: 65, Withdrawal Progression: increasing, Extra Withdrawals: no, Claiming Age: 63, Allocation: Target 60 %

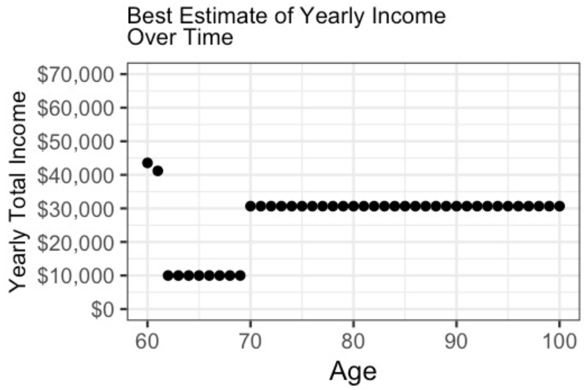
Graph	Frequency	Scenario
	4	Annuitization Percentage: 0, Annuity Start Age: 65, Withdrawal Progression: flat, Extra Withdrawals: no, Claiming Age: 70, Allocation: Target 90 %

Table 4. Experiment 1: Claiming age

	Dependent variable:	
	Claiming Age	Claiming Age
	(1)	(2)
Aggregate	-0.752**	-1.058***
	(0.282)	(0.287)
Discount Rate		-0.167*
		(0.072)
Loss Aversion		0.018
		(0.049)
Age		0.016
		(0.022)
HH Income		0.322**
		(0.105)
Retirement Savings		-0.080
		(0.058)
Education		0.224
		(0.118)
HH Size		-0.189
		(0.107)
Constant	65.213***	63.650***
	(0.141)	(1.487)
Observations	399	374
R^2	0.018	0.090
Adjusted R^2	0.015	0.070
Residual Std. Error	2.815 (df = 397)	2.747 (df = 365)
F Statistic	7.109** (df = 1; 397)	4.496*** (df = 8; 365)

Note: ****p* < .001. ***p*<.01. **p*<.05. Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous measures.

Table 5. Experiment 1: Risk allocation for retirement income

	Separate	**Aggregate**
Low Stock	17	9
Medium Stock	150	152
High Stock	35	36
Total	202	197

Table 6. Experiment 1: Annuitization results

	Dependent variable:	
	Annuitization Percentage	Annuitization Percentage
	(1)	(2)
Aggregate	-8.763**	-9.835***
	(2.728)	(2.784)
Discount Rate		1.243
		(0.702)
Loss Aversion		-0.949*
		(0.471)
Age		-0.078
		(0.217)
HH Income		1.962
		(1.019)
Retirement Savings		-0.970
		(0.566)
Education		0.286
		(1.141)
HH Size		-1.043
		(1.042)
Constant	20.371***	20.181
	(1.364)	(14.434)
Observations	399	374
R^2	0.025	0.064
Adjusted R^2	0.023	0.043
Residual Std. Error	27.242 (df = 397)	26.667 (df = 365)
F Statistic	10.319** (df = 1; 397)	3.120** (df = 8; 365)

Note: ****p* < .001. ***p*<.01. **p*<.05. Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous measures.

Table 7. Means (SDs) of Attitude Questions by Condition

Question	Aggregate	Separate
Plan Anxiety	1.39 (0.89)	1.42 (0.92)
Confidence to Make Plan	1.65 (0.82)	1.5 (0.91)
Likely to Make Plan	4.06 (0.95)	4.05 (0.95)
Confidence in Selections**	1.77 (0.78)	1.54 (0.92)
Saving Motivation	1.46 (0.83)	1.36 (0.84)
Subj Knowledge of Income Planning	1.47 (0.81)	1.41 (0.82)
Use Tool	2.24 (0.66)	2.23 (0.65)

Note: ****p* < .001. ***p*<.01. **p*<.05.

Table 8. Experiment 2: Effects of condition, with and without controls, on average income and NPV of income.

	Dependent variable:			
	Ave Est Income	Ave Est Income	NPV Estimated Income	NPV Estimated Income
Aggregate	0.545	58.888	-0.216	-0.311
	(143.951)	(142.822)	(0.607)	(0.611)
Discount Rate		-110.574**		0.485**
		(35.552)		(0.152)
Loss Aversion		-26.062		-0.028
		(23.405)		(0.100)
Age		-5.583		0.038
		(10.847)		(0.046)
HH Income		-89.089		-0.061
		(49.940)		(0.214)
Retirement Savings		12.672		0.098
		(28.705)		(0.123)
Education		108.560		0.317
		(59.955)		(0.257)
HH Size		132.708*		-0.260
		(53.781)		(0.230)
Constant	24,602.330***	25,156.100***	44,809.870***	44,804.860***
	(71.975)	(700.354)	(0.304)	(2.998)
Observations	525	518	525	518
R^2	0.00000	0.039	0.0002	0.028
Adjusted R^2	-0.002	0.024	-0.002	0.012
Residual Std. Error	1,648.922 (df = 523)	1,617.907 (df = 509)	6.957 (df = 523)	6.925 (df = 509)
F Statistic	0.00001 (df = 1; 523)	2.561** (df = 8; 509)	0.127 (df = 1; 523)	1.803 (df = 8; 509)
Note:				**p<0.05; **p<0.01;* ***p<0.001

Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous in the models above.

Table 9. Experiment 2: Effects of condition, with and without controls, on variability of income.

	Dependent variable:			
	SD Estimated Income	SD Estimated Income	Abs Lag Diff Est Income	Abs Lag Diff Est Income
Aggregate	182.284	232.104	7.424	14.011
	(214.612)	(214.896)	(25.063)	(25.320)
Discount Rate		-99.803		-9.159
		(53.503)		(6.298)
Loss Aversion		8.586		-1.070
		(35.255)		(4.153)
Age		-6.276		-0.297
		(16.320)		(1.923)
HH Income		-46.327		-6.267
		(76.711)		(8.856)
Retirement Savings		-16.735		-2.963
		(43.344)		(5.085)
Education		278.769**		22.574*
		(90.595)		(10.629)
HH Size		-78.176		3.942
		(84.504)		(9.540)
Constant	5,192.493***	5,306.985***	622.836***	626.471***
	(107.306)	(1,051.619)	(12.531)	(124.259)
Observations	522	515	524	517
R^2	0.001	0.030	0.0002	0.014
Adjusted R^2	-0.001	0.015	-0.002	-0.002
Residual Std. Error	2,451.009 (df = 520)	2,428.518 (df = 506)	286.802 (df = 522)	286.600 (df = 508)
F Statistic	0.721 (df = 1; 520)	1.972* (df = 8; 506)	0.088 (df = 1; 522)	0.874 (df = 8; 508)

Note: **p<0.05; **p<0.01;* *******p<0.001

Aggregate is contrast coded [aggregate: 0.5; separate: -0.5]. Discount rate, loss aversion, household income, retirement savings, and education are all ordinal variables that are treated as continuous in the models above. The standard deviation and lag difference outcomes exclude income from years 60 and 61.

Table 10. Experiment 2: Effects of condition on Claiming age

	Dependent variable:	
	Claiming Age	Claiming Age
	(1)	(2)
Aggregate	-0.002	0.061
	(0.236)	(0.237)
Discount Rate		-0.138^{*}
		(0.059)
Loss Aversion		0.009
		(0.039)
Age		-0.007
		(0.018)
HH Income		-0.095
		(0.083)
Retirement Savings		0.024
		(0.048)
Education		0.205^{*}
		(0.100)
HH Size		0.034
		(0.089)
Constant	65.181^{***}	65.621^{***}
	(0.118)	(1.164)
Observations	525	518
R^2	0.00000	0.022
Adjusted R^2	-0.002	0.006
Residual Std. Error	2.706 (df = 523)	2.690 (df = 509)
F Statistic	0.0001 (df = 1; 523)	1.414 (df = 8; 509)
Note:		*$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$*

Table 11. Experiment 2: Effects of Condition on Allocation

	Dependent variable:	
	30% Stock	90% Stock
	(1)	(2)
Aggregate	0.312	0.734^{**}
	(0.330)	(0.232)
Discount Rate	0.030	-0.094
	(0.082)	(0.057)
Loss Aversion	0.083	-0.075^{*}
	(0.058)	(0.036)
Age	-0.008	-0.033
	(0.025)	(0.018)
HH Income	-0.047	-0.148
	(0.116)	(0.078)
Retirement Savings	0.014	0.020
	(0.067)	(0.045)
Education	-0.037	0.124
	(0.136)	(0.098)
HH Size	-0.013	0.085
	(0.128)	(0.083)
Constant	-2.116	1.080
	(1.647)	(1.107)
Akaike Inf. Crit.	806.643	806.643
Note:	**p<0.05; **p<0.01;* *******p<0.001	

All values shown are in log odds.

Table 12. Experiment 2: Effects of Condition on Annuitization

	Dependent variable:	
	Annuitization Percentage	Annuitization Percentage
	(1)	(2)
Aggregate	-1.757	-1.666
	(2.214)	(2.235)
Discount Rate		0.378
		(0.556)
Loss Aversion		-0.077
		(0.366)
Age		-0.243
		(0.170)
HH Income		-0.567
		(0.781)
Retirement Savings		-0.358
		(0.449)
Education		-0.191
		(0.938)
HH Size		1.271
		(0.842)
Constant	17.920***	29.241**
	(1.107)	(10.959)
Observations	525	518
R^2	0.001	0.022
Adjusted R^2	-0.001	0.007
Residual Std. Error	25.357 (df = 523)	25.317 (df = 509)
F Statistic	0.630 (df = 1; 523)	1.429 (df = 8; 509)
Note:		**p<0.05; **p<0.01;* ***p<0.001

Table 13. Experiment 2: Means (SDs) of Attitude Questions by Condition

Question	**Aggregate**	**Separate**
Plan Anxiety	1.47 (0.89)	1.4 (0.89)
Confidence to Make Plan	1.48 (0.9)	1.54 (0.92)
Likely to Make Plan	4.12 (0.89)	4.07 (0.92)
Confidence in Selections	1.59 (0.85)	1.61 (0.84)
Saving Motivation	1.37 (0.91)	1.34 (0.85)
Subj Knowl of Income Planning	1.32 (0.78)	1.31 (0.83)
Use Tool	2.24 (0.67)	2.2 (0.59)

Note: ***$p < .001$. **$p<.01$. *$p<.05$.

Table 14. Experiment 2: Effect of Condition on Rating

	Dependent variable:	
	Scenario Rating	Scenario Rating
	(1)	(2)
Aggregate	-0.383	-0.423
	(0.253)	(0.256)
Discount Rate		0.032
		(0.067)
Loss Aversion		-0.059
		(0.042)
Age		0.028
		(0.020)
HH Income		0.135
		(0.094)
Retirement Savings		0.003
		(0.053)
Education		-0.015
		(0.116)
HH Size		-0.111
		(0.104)
Constant	6.434***	4.941***
	(0.127)	(1.355)
Observations	320	315
R^2	0.007	0.034
Adjusted R^2	0.004	0.009
Residual Std. Error	2.257 (df = 318)	2.252 (df = 306)
F Statistic	2.289 (df = 1; 318)	1.346 (df = 8; 306)
Note:		*p<0.05; **p<0.01; ***p<0.001