RESEARCH

Swiss Finance Institute Research Paper Series N°22-72

Experimental Research on Retirement Decision-Making: Evidence from Reproductions

Kremena Bachmann
University of Zurich and Zurich University of Applied Sciences

Andre Lot
Norwegian School of Economics and Zurich University of Applied Sciences

Xiaogeng Xu
Hanken School of Economics

Thorsten Hens
University of Zurich, Norwegian School of Economics and Business Administration, and Swiss Finance Institute

Experimental Research on Retirement Decision-Making: Evidence from Reproductions *

Kremena Bachmann† Andre Lot‡ Xiaogeng Xu§ Thorsten Hens¶

September 12, 2022

Abstract

We adapt the design of five experimental studies on retirement decision-making and conduct reproductions with a larger sample from the broader population. We reproduce most of the main effects of the original studies. In particular, we find that consumption decisions are less efficient when subjects need to borrow from the future than save from the present. When subjects collect retirement benefits as lump-sum instead of annuities, they choose to retire later. The duration of retirement affects the saving behavior of the subjects. Savings are higher when they are incentivized with matching contributions than with tax rebates. When faced with stochastic survival risk, subjects make partial adjustments to spending paths. We also propose a further experimental research agenda in related topics and discuss practical issues on subject recruitment, attrition, and redesign of complex tasks.

JEL Classification: C91, D15, G51, J26

Keywords: household finance; retirement decision; savings; annuities; life-cycle optimization; income smoothing; experiments; reproductions

Online Repository [click] with supplementary materials.

*This project has received financial support from the Swiss National Science Foundation, [grant number 100018_189107]. Declarations of conflict of interest: none for all authors.

†University of Zurich and Zurich University of Applied Sciences. Corresponding author. Email: kremena.bachmann@bf.uzh.ch

‡Norwegian School of Economics and Zurich University of Applied Sciences

§Hanken School of Economics

¶Swiss Finance Institute Professor at University of Zurich. Adjunct Professor at University of Lucerne and Norwegian School of Economics

1 Introduction

Retirement financial decisions over the life cycle show puzzling patterns in the field, such as subjects not converting savings into life annuities, saving too little before retirement, while also spending their savings too slowly after retirement (Lugilde, Bande, & Riveiro, 2019; Peijnenburg, Nijman, & Werker, 2016; Feigenbaum, Gahramanov, & Tang, 2013; Heimer, Myrseth, & Schoenle, 2019). Some of these patterns may be related to the nature of the decision problem. Financial decisions over the life cycle are complex and require high cognitive skills and financial knowledge. Also, learning from one's own experience is limited by long spans between decisions and observable outcomes and a low decision frequency. Normative institutional settings and strong social norms around these decisions impose further challenges for researchers seeking to identify the underlying drivers of observed behavior.

Experimental studies on retirement decisions address the empirical challenges brought by those characteristics of the decision environment in the field. However, many of these studies rely on age and income homogeneous student samples whose subjects are distant with respect to their chronological age and discretionary income levels, from the contextual experiences of individual retirement planning in the field. Characteristics of typical student samples could be particularly problematic for elicitation procedures and task designs in the existing literature, as noted, among others, by Carbone (2006). Higher cognitive abilities within student samples could also conceal the limitations faced by the representative agent in the population making retirement decisions, such as the use of shorter planning horizons (Ballinger et al., 2011). The lack of experience of students with long-term debt management could also plausibly explain certain sub-optimal life-cycle optimization results, such as those of Meissner (2016).

In this paper, we report the results of five pre-registered reproduction studies. We aim to test the main findings of the original experiments on retirement decision-making using larger online samples that are more representative of the general population than those of the original studies, and introducing financial incentives where they were absent. We selected studies to investigate different aspects of the retirement decision-making problem and to check the feasibility of different experimental designs with samples from the general population. To allow subjects to complete experimental tasks independently without a supervisor in a short period of time, we significantly simplified the original experimental designs.

We successfully reproduced most of the main effects of the selected experimental studies. In particular, we found that subjects optimize consumption less efficiently when they need to borrow from the future than save from the present (as in Meissner, 2016). They are also more likely to delay the timing of retirement when collecting benefits as lump-sum than annuities (like those in Fatas, Lacomba, & Lagos, 2007). When faced with longer or shorter retirement periods, subjects make some adjustments to their savings patterns (following partially the findings of Koehler, Langstaff, & Liu, 2015). When offered incentives to save as matching contributions rather than tax rebates, they have higher effective savings rates (in line with Blaufus & Milde, 2021). Finally, when facing stochastic survival risk, subjects adjust to some extent their spending (within the setup of Anderhub et al., 2000).

We also document evidence of substantial and consistent sub-optimal decision-making across certain subject types. In experiments with enforced lifetime budgets, across different rounds, a group of subjects consistently under-consume their lifetime income, while others consistently do not save enough, going bankrupt when needing to fund mandatory expenses. A considerable fraction of subjects also violate ordinal optimality in their consumption paths in the presence of stochastic survival risks.

Finally, we present and discuss some important methodological challenges and practical issues concerning the modification of original tasks, the implementation of such experiments with online representative panels, and the efficiency of subject decision-making within the tasks. Then, we propose a further experimental research agenda on relevant topics and themes to address lingering questions arising from the current state of the empirical field and experimental literature.

By reproducing the main effects of several experimental studies on retirement decision-making using larger and more representative samples, our paper mainly contributes to the internal and external validity of the findings of this literature. Our results also support the point of view that the complexity of financial retirement decisions per se could be an obstacle to efficient decision-making. Finally, our practical experimental implementation choices and outcomes can be helpful for future experimental research tackling complex tasks with dynamic programming components, in this or in other domains.

In the following Section 2, we present an overview of the relevant experimental literature. Then, in Section 3 we introduce the original studies and present the results of our reproductions. We discuss the implications of our results and propose a future research agenda on this topic in Section 4, and conclude in Section 5.

2 Experimental literature on retirement decision-making

Experiments on individual retirement decision-making investigate the importance of its various driving factors by employing different task designs. In the first subsection, we present an overview of the literature along with the factors that previous studies considered as potential drivers for the observed decision-making behavior. We then discuss, in more detail in the second subsection, the most common experimental task features that distinguish experiments in this domain. Table 1 summarizes the studies in terms of their main findings and distinguishing features with respect to the experimental design.

« Placeholder for Table 1 »

2.1 Drivers of retirement decision-making behavior

One strand of experimental studies investigates how specific features of the decision problem affect people's decision behavior. Carbone and Hey (2004) investigate how people adjust their consumption behavior to the possibility of unemployment and find that people overreact to the risk of unemployment. Koehler, Langstaff, and Liu (2015) vary the length of the retirement phase and find that most participants respond sensibly by saving more of their current income when faced with a long compared to a short retirement phase. Meissner (2016) studies optimal consumption on an increasing and decreasing income path and finds that when people are required to borrow to smooth consumption (i.e., when their income path is increasing), deviations from optimal behavior are more likely. Anderhub et al. (2000) relax the assumption in most experiments that the survival probabilities are constant and find that the average subject reacts in a qualitatively correct way to 'good' and 'bad' news on survival risk. While most studies consider decisions under income distribution risk, Carbone and Infante (2014) study decision-making under risk and ambiguity and find that behavior under ambiguity is characterized by a significant pattern of under-consumption compared to behavior under risk. In terms of the quality of the general decision behavior of the subjects, Hey and Dardanoni (1988) find that the subjects respond optimally to changes in discount factors and the return on savings.

The retirement decision problem has features that can also be determined by the institutional environment. Bohr, Holt, and Schubert (2019) study the introduction of automatic savings schemes and find that individuals save less with such schemes, but the reduction is only partial, in the sense that the total lifetime consumption measures are higher. Duffy and Li

(2019) consider different pension replacement rates and find that subjects achieve the highest experimental payoff when offered a constant life-cycle endowment profile (100% pension replacement rate). Hurwitz, Sade, and Winter (2020) investigate the benefits of implementing a minimum annuity rule and find that it does not guarantee an increase in the demand for annuities and may even reduce it. Beshears et al. (2020) evaluate the benefits of introducing higher withdrawal penalties in retirement savings schemes and find that higher early withdrawal penalties attract more commitment account deposits. Fatas, Lacomba, and Lagos (2007) examine whether the form of pension benefits (lump-sum payments or annuities) affects retirement decisions and find that concentrating payments (shifting from annuity into lump sum) can motivate subjects to postpone retirement.

However, the complexity of this decision problem also raises the question of whether people learn to deal with the problem from experience or from the choices of others. Brown et al. (2009) find that subjects save too little at first, but learn to save close to optimal amounts after three or four rounds (of one simulated life-cycle each). Meissner and Rostam-Afschar (2017) find that people learn to operate under a Ricardian tax scheme (a tax cut in early periods of the experiment, followed by a tax increase of the same magnitude in later periods), but the aggregate effect of taxation on consumption persists, even after eight rounds. Because the subjects in the field made decisions for only one life, important insights can arise from social learning. Carbone and Duffy (2014) find that the provision of social information on past average levels of consumption results in a greater deviation of consumption from optimal paths. Similarly, Feltovich and Ejebu (2014) allow interpersonal comparison and find that it leads to worse outcomes in the form of more under-saving and lower money earnings. In contrast, Ballinger, Palumbo, and Wilcox (2003) analyze learning effects using an intergenerational structure and find that subsequent generations perform significantly better in terms of savings than previous generations.

Few studies analyze the effect of specific behavioral biases on retirement financial decisions. Levy and Tasoff (2020) find that the subject's decision behavior is consistent with the exponential growth bias. Agnew et al. (2008) find that an excessive extrapolation of the past performance of the financial market influences the demand for annuities. Blaufus and Milde (2021) find that different frames of tax-related pension incentives can influence retirement savings, while Brown et al. (2008) also find that the use of different frames can affect the demand for annuities. Several experiments report evidence that subject behave myopically (Carbone & Hey, 2004; Ballinger, Palumbo, & Wilcox, 2003; Carbone, 2005, 2006) and have dynamically inconsistent preferences (Brown et al., 2009). In terms of general decision-making behavior, Carbone (2005) finds that subjects apply common rules-of-thumb to solve the optimization problem. Subjects also build wealth even if it is not optimal to do so (Gechert & Siebert, 2020).

Finally, some studies aim to explain the heterogeneity in behavior based on dynamic decision-making tasks. Ballinger et al. (2011) find that cognitive abilities (but not personality measures) are good predictors of heterogeneity in saving behavior observed as a result of using shorter than optimal planning horizons. Carbone (2005) concludes that demographic characteristics have minor effects on the planning horizon of the subjects and on the strategies applied to solve the optimization problem. Carbone (2006) finds that hyperbolic discounting affects the behavior of students more strongly than that of the general population, which cannot be explained solely by age differences, as younger people are generally considered to be more hyperbolic discounters.

2.2 Design features of the experiments

Most experimental studies on retirement decision-making require sequential decisions over several periods of simulated life (a round). The number of periods can be either fixed or determined by some random process. There is an implicit longevity risk when the number of periods is not fixed, which brings interesting complications into the optimization problem faced by subjects (Agnew, Anderson, & Szykman, 2015; Anderhub et al., 2000; Fatas, Lacomba, & Lagos, 2007; Hey & Dardanoni, 1988).

Another source of uncertainty in the optimization problem could be stochastic income. This type of uncertainty can be used in different ways. It can be linked to the probability of becoming unemployed or later reemployed (Carbone & Hey, 2004). It can also be represented by a simple *i.i.d.* process (Ballinger, Palumbo, & Wilcox, 2003) or by a fluctuating stream of either high or low income (Feltovich & Ejebu, 2014; Carbone, 2005; Carbone & Infante, 2014; Meissner & Rostam-Afschar, 2017). Alternatively, it can be implemented by adding or subtracting a constant error term from an otherwise linear income process (Meissner, 2016). Introducing an uncertain income as an experimental feature is certainly realistic. However, when analyzing deviations from optimal consumption paths, it can be difficult to distinguish between deviations caused by a misperception of probabilities and deviations caused by the general cognitive difficulty of finding the optimal solution. For this reason, some studies use deterministic income paths (e.g. Duffy & Li, 2019).

To incentivize savings, some experiments require subjects to cover some mandatory expenses during the simulated life-cycle (Hurwitz, Sade, & Winter, 2020; Koehler, Langstaff, & Liu, 2015; Agnew, Anderson, & Szykman, 2015). This feature can also determine their survival in experiments.

In approximately half of the reviewed studies, savings are incentivized through an interest-bearing savings account. While offering interest increases the attractiveness of saving versus immediate consumption, this can increase the computational burden to participants and cause sub-optimal decisions.

Some studies introduce a retirement phase as part of the inter-temporal optimization problem (Blaufus & Milde, 2021; Bohr, Holt, & Schubert, 2019; Duffy & Li, 2019; Feltovich & Ejebu, 2014; Koehler, Langstaff, & Liu, 2015). In the retirement phase, there is no uncertainty about exogenous income, which is set to zero, i.e. subjects will only be able to consume and/or pay expenses in the retirement phase from their savings that they accumulate during the working phase. The solution to inter-temporal optimization problems with and without such a retirement phase may differ depending on whether subjects misinterpret the probabilities concerned, for instance, by overreacting to events occurring with certainty (periods with zero income) as compared to events occurring with very high/low probability (periods with unemployment or income shock risk).

Only a few studies enforce a life-time budget, whereby any wealth left at the last period is automatically spent (Blaufus & Milde, 2021; Bohr, Holt, & Schubert, 2019; Brown et al., 2009; Koehler, Langstaff, & Liu, 2015; Meissner, 2016; Meissner & Rostam-Afschar, 2017). This feature simplifies the analysis of experimental decisions and facilitates calibration of several theoretical models underpinning the experimental designs, but it may potentially obfuscate instances of sub-optimal behavior or misunderstanding of the experimental tasks.

Finally, to motivate subjects to optimize their consumption paths, most studies link subjects' consumption choices to their payoffs. Some studies specify the link by using a particular (induced) utility function. When there are no interest earned on savings, and payoffs are based on lifetime outcomes, inducing an utility function is essential. Otherwise, subjects might just assign most of their lifetime consumption at one or some of the periods, then consume little (or save just enough for expenses, if applicable), as many possible combinations of period consumption would yield the same lifetime outcome. Experiments without an induced utility can motivate consumption smoothing by linking compensation to choices in one random period. This latter task design is much simpler for subjects to understand, although it carries the small drawback of allowing risk-seeking subjects to gamble by concentrating most consumption in just one period in hope that this period is eventually selected for payoff.

3 Reproductions of adapted experimental designs

Taking into account the existing body of previous experimental studies on retirement decision-making (see Table 1), we selected five experiments that spanned a heterogeneous set of research topics and experimental design features. In terms of research topics, we selected three studies investigating different characteristics of decision, such as ambiguous survival probabilities (Anderhub et al., 2000), different income paths (Meissner, 2016), and different relative lengths of working and retirement phases (Koehler, Langstaff, & Liu, 2015). Another selected study addresses the relevance of behavioral effects, more specifically framing effects (Blaufus & Milde, 2021). In addition, we select one study that investigates the effects of the characteristics of the institutional environment (Fatas, Lacomba, & Lagos, 2007). In terms of experimental design characteristics, we selected studies, within each topic, that vary with respect to the task features summarized in Table 1. Taking into account the uneven distribution of the characteristics of the experiments (within this research agenda on retirement decision-making), and the technical feasibility (or lack thereof) of reproduction of certain experimental designs using our online sample (for instance Brown et al., 2009), we selected these five aforementioned experiments for reproductions within our time and budget constraints.

Reproductions and, in some cases, additional analyses of individual experiments were pre-registered on AsPredicted.[1] Each of these studies addresses a different research question; henceforth, we do not propose any joint analysis of individual reproduction results with respect to their original hypotheses.

In the reproductions, we focus on one or two main effects of each study. We intended to reproduce the studies using subjects from the general population (four of the selected studies originally used students in the lab, and one was originally conducted online with an Amazon M-Turk sample), who needed to perform the tasks online without any assistance. For this purpose, we modified, as needed, the original experimental design and adjusted the tasks. We used the same large pool of subjects, deployment method, quality control mechanism, and common design and interface features in all reproductions, to the extent applicable.

[1] See Pre-registration (1), (2), (3), (4) and (5).

In the next subsections, we first discuss our approach and procedures to modify and adjust experimental designs and their tasks, and general engagement and performance metrics of subject participation. Then we discuss the specific reproduction results for each study. For parsimony, we will skip most or all of the discussions of the models and hypotheses used and developed by the authors of the original studies and refer interested readers to the respective original published research papers instead.

3.1 Redesign and adaptation of experimental tasks

The reproductions were subject to a main completion time constraint: a subject's session must have been kept under 30 minutes on average and ideally below 25 minutes. Four of the original experiments had lab sessions that lasted much longer than that, up to 90-120 minutes. Their original sessions included extensive subject education and training. In addition, some of the experiments had a very complex set of instructions, including direct mathematical formulae presented to subjects to explain the induced utility and complex payoff mechanisms. One experiment (Koehler, Langstaff, & Liu, 2015) was shorter, having originally been implemented on Amazon M-Turk; however, it was not incentive compatible, as we intended all our reproductions to be.

These characteristics of the original studies would make any attempt to closely replicate all of the original experiments unfeasible. To address this challenge, while aiming to preserve the main mechanisms we wanted to reproduce, we modified and redesigned the experimental tasks, to varying degrees.[2]

In three experiments, we reduced the number of rounds and/or periods per round, preserving the structure of lifetime budget constraints and the relative scale of income paths, expenses, and other environmental variables, where applicable. There is a long-standing concern in the literature about the elicitation of decision-making sets for subjects that need to engage in dynamic programming and the minimum necessary number of periods over which optimization is to be done. However, we believe that a partial reduction in the length of each round, or the number of rounds, is not much of an issue in our reproductions as it would have been in other experiments that rely on stochastic environmental variables that persist over many periods (such as in the first task of Brown et al., 2009).[3]

[2]We would like to thank Thomas Meissner, Derek Koehler and Kay Blaufus for providing us with additional materials from their experiments; we did not ask for materials to authors of other older papers.

[3]The reduction in the number of rounds would have affected the analyses of within-subject learning over rounds. We did not study learning across rounds except in the pre-registered additional analysis in Fatas, Lacomba, and Lagos (2007).

Three experiments originally used a large number of sequential computer screens for feedback on results, reassurance of procedures, and indirect attention checks. Compounded over dozens of periods and several rounds per subject, this approach greatly lengthens the total session time. In our reproductions, we streamlined the interface so that the information and decision screens and action buttons for each round (i.e., one experimental life) could fit on one screen. We used dynamic tables, one per round, that were progressively filled with each period's decision and populated from the beginning with information on constant or predetermined environmental variables (such as a predetermined income path).[4] Where not obvious, we implemented hovering text balloons that quickly expanded the concept of variables at the top of the dynamic tables.

For input on consumption and savings decisions in all relevant experiments, we used sliders (automatically adjusted to the boundaries of budget constraints, if any) instead of typing boxes. Changing the decision slider(s) would also reveal to subjects the simple accounting mechanics on savings and cash balances, where relevant, and also give feedback on expected payoffs in future periods (as in Blaufus & Milde, 2021). Together with the one-dynamic-screen-per-round approach, this greatly reduced the need to navigate through different screens, reducing substantially the time required to complete the otherwise repetitive multi-period decision tasks.

Other experimental design features that substantially contribute to the session's completion time in the original studies are instructions and training on the actual task. Although we showed and asked subjects to read the instructions at the beginning of the session, we let them know that the instructions would always be available during the main task. This was implemented using clickable tabs at the bottom of the dynamic screens. Each tab had a small, self-contained piece of information that addressed only one aspect of the experimental task. To further improve the accessibility of instructions, we replace explicit complex mathematical formulae (such as the induced utility in Meissner, 2016) with graphs that showed, more intuitively, the relevant functional relationship between variables.

The session flow in all reproductions is illustrated in Figure 1. Once the subjects completed reading the instructions, they started a trial round.[5] This allowed them to learn by doing the main experimental task, with ready access to the instructions in tabs at the bottom of the

[4] All screenshots for all treatments of the reproductions are available in the Online Appendix.

[5] In experiments adopting a within-subject treatment, the trial round is always identical to the treatment the subjects will undergo in the first live round.

screen.[6] The subjects then answered a quiz with four or five questions on the basic mechanics or features of the task, before moving on to the rounds of the main task. Random elements of the payoff determination, such as the selection of one period of one round for compensation, were only revealed at the very end of the session. After the main task, subjects were asked basic demographic information (age, sex, education, income range, and financial training experience). We elicited their risk preference with an assignment task (of their main task earnings) from Gneezy and Potters (1997), and their time preferences (patience) as their willingness to delay their variable payoff by 1, 2 or 3 months for a 5% monthly interest.[7] The final payment was determined by the earnings with the main experimental task, the outcome of the risk taking task and the choice of the time preference task. Subjects were only informed at the end of the experiment about their final payoff and its components.

« Placeholder for Figure 1 »

Subjects were recruited from the Germany recruitment pool of the market research company *Bilendi*. The experiments could be taken in English, however, less than 2% of the subjects opted out of the German default. The experiments were programmed in oTree (Chen, Schonger, & Wickens, 2016). Power analyzes were computed with GPower (Erdfelder et al., 2009).

3.2 Subject engagement, quality control and decision efficiency

We implemented strict quality control on responses. Subjects were dropped if they skipped too fast through the instruction screens at the beginning of the sessions (thresholds of 10 to 60 seconds). On the quiz, the subjects were dropped if they answered more than two or three wrong questions on a first attempt or any wrong answer in a second attempt.[8] They were also automatically removed from the experiment if more than 60 minutes had passed after the quiz was completed.[9]

The experimental sessions were conducted in individual batches for each experiment between

[6] The trial round is not relevant for the payoff.

[7] All payments were credited to the subject accounts directly by the market research company, upon receipt of a master payment file from us. Since subjects in their pool often participate in a few surveys or activities per month and are used to be paid regularly, it is unlikely that the options for delayed payment would have been avoided due to concerns about administrative and time costs to recover delayed payments.

[8] A second quiz attempt highlighted the questions they got wrong and displayed a reminder with the relevant snippet from the instructions that had the relevant information needed to correct the wrong answer(s). We shuffled the order of the options of the quiz questions in the second attempt.

[9] Very few subjects appear to have been removed from the experiment for taking too long while continuously engaged in the tasks. In all cases, this removal procedure ensured that subjects who abandoned their screens and browser tabs would not be able to resume the experiment many hours or days later.

September 2021 and March 2022. 6213 subjects clicked on e-mail invitations sent[10] from the market research panel.[11]

Panel A of Table 2 details the attrition at each step for all the reproductions. Overall completion rates ranged from 21.5% to 50.6% of invitation clicks.[12]

« Placeholder for Table 2 »

Panel B of Table 2 summarizes the incentive component of the payoff of the subject[13] and the completion time for the experiments. All reproductions could produce zero incentive payoff for subjects, and the largest subject incentive payoff was €101.16. The panel also summarizes the completion time of the subjects who answered all questions.[14]

The randomization of subjects to treatment cells in all experiments seems satisfactory with respect to the demographics of the subjects, as seen in Table 3. The *t-tests* show that the means of *age* are significantly different ($p = 0.06$) between treatments in the reproduction of Anderhub et al. (2000). The means of *female* are significantly different ($p = 0.06$) between treatments in the reproduction of Meissner (2016), and ($p = 0.08$) of Anderhub et al. (2000). ANOVA tests show that the means of the variable *patience* are different ($p = 0.02$) in the reproduction treatments of Fatas, Lacomba, and Lagos (2007). For all other characteristics and treatments of each experiment, there are no significant differences within each experiment at the 10% level. It should be noted that variables *risk-taking* and *patience* were generated after the main tasks, so the subjects' expectations about their earnings from the main task could affect their decisions on the risk-taking task and the time preference task.[15] Although subjects were initially assigned to balanced treatment cells in all experiments, different attrition rates produced slightly unbalanced cells in all reproductions.[16] The imbalance ratios (between the largest and smallest cell size in each experiment) ranged from 1.07 to 1.16.

[10] Any subject that gave consent and started the trial of one reproduction experiment was automatically excluded from participating in any other.

[11] The invitation emails are brief, informing subjects mostly of the expected length of the task and expected payoff.

[12] The abnormally high drop-out rates of Blaufus and Milde (2021) was due to a database load surge that slowed down the interface for some hours of the second day of data collection, which motivated some subjects to abandon the task.

[13] In addition to variable payoffs, subjects who finished the experiment earned €4.76 as a show-up fee.

[14] A few subjects who answered all the questions but forgot to click 'Finish' skew the maximum completion time shown in the table.

[15] Even though the resolution of uncertainty would only be resolved at the end of the experiment, subjects who performed poorly in the main task on all rounds could consider their low expected payoff when deciding on the risk-taking task.

[16] We had many subjects simultaneously completing each experiment at once, which required predetermined treatment assignment upon click-through. Therefore, it was not possible to rearrange the empty rows on the online input database based on live completion information as subjects ended the experiment.

« Placeholder for Table 3 »

Finally, we evaluate the effects of individual subject characteristics on their economic efficiency of decisions across the experiments.[17] Across four experiments,[18] *female* subjects perform worse than males. In three experiments, higher *risk-taking* subjects perform significantly worse.[19] The effects of *financial ability* are ambiguous between the experiments. These differences do not seem to arise from different effect levels between the treatment assignments, as shown in specifications incorporating round × treatment fixed-effects for each experiment.

« Placeholder for Table 4 »

3.3 Reproduction results of Anderhub, Strobel, Müller and Gühl (2000)

The original study addresses uncertainty (and its resolution) on survival risks in the context of spending an endowment over multiple periods. It tackles consumption smoothing in a life-cycle task with longevity risk. In the experiment, subjects faced three possible survival risk functions in the form of colored dice,[20] which are discarded (reducing uncertainty) in the first two periods, until only one remains, which is then applied to determine survival in subsequent periods. The treatment effect, which we explore, is the form of the induced lifetime utility: the lifetime *Summation* of the square root of the period consumption or *Product* of the period consumption. Subjects completed six rounds, comprising all permutations of the sequence of resolution of the uncertainty of survival risk. We used the same initial endowment for both treatments (11.92 ECU) as in the original study.

In our reproduction, 339 subjects completed the experiment (176 in treatment *Product* and 163 in treatment *Summation*). In the *Product* treatment, there are 1066 subject × round and the average payoff is €3.33: in the *Summation* treatment, there are 985 rounds and the average payoff is €4.16.

We show the univariate statistics for the average decision per period according to the uncer-

[17]This analysis was not preregistered.

[18]Fatas, Lacomba, and Lagos (2007) does not have a within-subject dynamic endogenous (to the main task) benchmark for decision efficiency, given its task design.

[19]We cannot exclude an instance of gambling, as the risk preference elicitation follows the main task: subjects who know to have performed badly in the main tasks might well decide to take more risk in the following risk taking task to recover perceived 'losses' in the main task.

[20]Replaced by card decks in the reproduction.

tain resolution path in Figure 2 and Figure 3. The red (blue) outline of the nodes indicates that on average the subjects on that treatment, period, and uncertainty resolution path spent above (below) the optimal consumption levels for that node. The spending patterns for each period are very similar to those of the original study. The average efficiency is defined as U/U^*, where U is the average payoff in all six rounds, and U^* is the expected optimal payoff. The average efficiency is 0.50 for *Product*, and 0.88 for *Summation*. Both efficiency rates are lower than in the original study. However, our results are consistent with the original study in that the efficiency rate is higher in condition *Summation* than in condition *Product*.

« Placeholder for Figure 2 and Figure 3 »

In contrast to the original study, we find that there is substantially less differential adjustment to the resolution of the risk of ambiguity. For example, in the second period (X2) for *Summation* condition, the average spending ranges from 2.47 to 2.55 points, while in the original study the averages ranged between 2.56 and 3.23. Furthermore, the variation in average spending per treatment, period, and ambiguity resolution path was also considerably higher in our reproduction than in the original study. Furthermore, the ordinal rank of average spending violates the rank of optimal consumption paths.

We checked whether the reactions to the first removed die are qualitatively correct. Subjects supposedly consume more when the die with a low termination probability is removed compared to the die with a high termination probability, as implied by the condition $(\frac{x_2}{S_2}\Big|\neg\text{green}) > (\frac{x_2}{S_2}\Big|\neg\text{yellow}) > (\frac{x_2}{S_2}\Big|\neg\text{red})$., where $\neg$ is the removal of a card deck (set of survival probabilities). There are 23 of 176 subjects in the treatment product and 30 of 163 in the treatment *Summation* who met this condition. We also checked whether the reactions to the final die are qualitatively correct. Subjects supposedly consume more when the termination probability of the final card deck is high than low, as implied by the condition $(\frac{x_3}{S_3}\Big|\text{red}) > (\frac{x_3}{S_3}\Big|\text{yellow}) > (\frac{x_3}{S_3}\Big|\text{green})$. There are 29 out of 176 subjects in treatment *Product* and 25 out of 163 subjects in *Summation* who fulfilled the condition.

Regarding cumulative consumption in the first three periods (that is, until the ambiguity of survival risk is resolved), we find that, contrary to the original study, the fraction of endowment consumed in these initial periods does not vary substantially according to optimal levels under either treatment condition, as seen in Table 5. In particular, the fraction of consumption under the *Summation* condition ranged only from 0.69 to 0.72 (compared to optimal levels of 0.70 to 0.89 and decisions in the original study of 0.70 to 0.83). In general, in our reproduction, subjects seem to react much less to changes in their underlying survival risk than subjects in the original study.

« Placeholder for Table 5 »

As the survival probabilities in periods 3 to 6 are always below 1, it follows that consumption should always be larger in the earlier periods when the survival probabilities remain ambiguous. In Table 7, we tabulate the percentage of cases where this condition is met, according to the termination periods of each subject in each round. As in the original paper, a large fraction of cases do not adhere to relatively relaxed conditions arising from bounded rationality. In the right column of panel A for the *Product* condition, for example, only 35.6% of the cases who reached period 6 had a monotonically decreasing consumption between periods 3 and 6 (35.5% in the original study). In Panel B for *Summation*, 35.3% of the cases met the same conditions (48.7% in the original study). Most of the discrepancies between our results and the original study in this analysis are due to our relatively smaller differences, in each termination period and treatment, between the fraction of subjects who satisfy the condition strictly (as before) or weakly ($x_3 \geqslant x_4 \geqslant x_5 \geqslant x_6$). In the original study, instead, a large number of cases violated the strict condition, but kept consumption numerically constant between two rounds. This specific difference between violations of strictly and weak conditions is arguably attributable in part to our use of a slider, with precision at the 0.01 point, rather than requiring a numerical input.[21]

« Placeholder for Table 7 »

3.4 Reproduction results of Fatas, Lacomba and Lagos (2007)

This study investigates the impact of the form of retirement payoff (*Annuity*, *Lump-sum* or *Combined* from the previous two, which are the between-subject treatment conditions) on the timing of retirement, i.e., the choice of subjects of when to (start to) collect payoff(s). Subjects face longevity risk in every period and can only earn a reward while they are still 'alive'. Among the selected studies, this is the only one in which subjects make one decision per round, which is the timing of retirement.

In our reproduction, 530 subjects completed the experiment (177 in the *Annuity*, 170 in the

[21] Subjects ex-ante intended on keeping consumption constant could just still not bother to fit the very exact same number. This behavior would plausibly increase violations in both directions, that is, consumption that increases or decreases slightly from one period to the next, reducing the proportion of cases where numerically equal consumption is observed in two sequential periods.

Combined and 183 in the *Lump-sum* condition). Following the analysis of the original study, we find that subjects earning *Lump-sum* payments chose to retire later than those earning *Annuity* or *Combined* (with all payoffs actuarially equivalent), as shown in Figure 4. On average, subjects in the *Annuity* condition, chose to retire after 5.49 periods, those in the *Lump-sum* condition retired the latest after 6.32 periods[22] and those in *Combined* after 6.13 periods.

« Placeholder for Figure 4 »

As in the original study, we found significant treatment effects between the treatment conditions of *Lump-sum* and *Annuity*. Following the original study, using *Lump-sum* as a baseline, we regressed the retirement periods chosen on our own measures of *risk-taking* and *patience*, and on the treatment dummies. The results are shown in Table 8. In the full specification (4), subjects in the *Annuity* payoffs choose to retire 0.916 periods earlier than those in *Lump-sum*. The difference is smaller (0.863 periods) but still significant before controlling for patience in (5). The coefficients for the treatment effect are significant in both our reproduction and in the original study. The differences in the choice of timing of retirement between *Lump-sum* and *Combined* are not significant in either specification, and the difference in the coefficients of *Annuity* and *Combined* (0.635 and 0.645 periods in (4) and (5)), shown in the bottom panel, is significant as in the original study.

« Placeholder for Table 8 »

Similarly to the original study, we find that higher risk-taking is significantly associated with a later choice of retirement timing: each additional percentage point allocated to a risky asset in a Gneezy and Potters (1997) task is associated with a delayed retirement timing of 0.026 to 0.028 periods. Patience is also positively associated with a delay in retirement. Each month that subjects choose to wait for their payoff in exchange of interest delays the choice of retirement period between 0.184 and 0.255 periods.

In an additional pre-registered analysis that was not part of the original study, we analyze how termination at round one and/or two affects the choice of timing of retirement in later rounds. In our reproduction, the subjects made repeated decisions in three rounds. Although subjects did not know which round would be relevant for their compensation until they completed the experiment, they experienced resolution of longevity risk through a random draw of a

[22] In the original experiment, subjects under the *Lump-sum* condition retired, on average, much later after 9 periods.

card without replacement (beginning with 1 red card and 14 green cards) in each round, until a red card was drawn and subjects were terminated in that round.[23] The results of this additional analysis are presented in Table 9.

« Placeholder for Table 9 »

Generally, we find that a later termination in earlier round(s) is associated with a significantly delayed choice of retirement in subsequent round(s). In specification (3), controlling for the treatment, a first round that last one period longer delays the retirement timing chosen in the second round by 0.06 periods. A much more salient event, notwithstanding, is subject survival at least until the period they had chosen to earn (or start earning) their payoffs. In specification (4), we regress the choice of timing of retirement in round two on whether the subject survived until their chosen timing of retirement during round one. In round one, surviving at least until the chosen period delays the subsequent choice of timing of retirement chosen in round two by 2.782 periods. The direct effect of one later period for termination is then a further delay of 0.277 periods for round two. In specifications (5-8), we test retirement timing in round three given outcomes of the first two rounds: the effect of termination period and survival until the chosen period remains significant in all the specifications.

3.5 Reproduction results of Koehler, Langstaff and Liu (2015)

The original study aimed to develop a simplified and standalone saving task that could be implemented online, without assistance, with non-student subjects. This is the only study, among the ones we reproduce, that was originally implemented with an online sample (Amazon's M-Turk subjects). It was also the only original study that did not have a variable incentive structure, which we introduced in our reproduction.

In the reproduction, we focus on the main treatment of relative length of the retirement phase (*Short* or *Long* retirement) in the life-cycle. As in the original study, subjects play two rounds under one condition, then change to the other condition for the final two rounds, with a random assignment of the starting condition. Subjects are 'retired' (earning no income, but having to pay mandatory expenses) for four periods in the *Short* retirement condition, and for eight periods in *Long*, out of 16 periods per round.[24] In our reproduction, subjects earn variable incentives based on their consumption decision in one period of one round. During

[23]The original study used colored balls.

[24]In the original study, each round last 24 periods, with *Short* retirement consisting of 6 periods and *Long* of 12 periods.

working periods, income is always larger than mandatory expenses, so that even a subject who always consumed everything available would still have enough income to meet his or her expenses. In retirement periods, subjects who did not save enough in the working phase will be unable to meet expenses, i.e., become 'bankrupt'.[25] To prevent strategic but unwanted behavior on consumption decisions,[26] in our design, we introduce a penalty for 'bankrupt' subjects: If a round in which a subject went 'bankrupt' is selected at the end of the session for compensation, the incentive payoff is automatically set to zero, regardless of any other spending decisions in the other periods of that round. Compared to the original design, this bankruptcy penalty strengthens the incentive for subjects to – at the very least – save enough during the working periods to meet the mandatory expenses known to await them during retirement.

We collected valid responses from 344 subjects, of whom 166 started the first two rounds under the retirement condition *Long* and 178 started under the condition *Short*, then switched to the other condition for another two rounds.

Following the original study, we analyze how treatment affects the total retirement savings balance in the last period of the working phase and the variability in spending. As in the original study, we find that the duration of retirement is a highly significant determinant of retirement savings. In ANOVA analyzes, the retirement length treatment has a significant effect on retirement savings, with $F\,(1375) = 1445,\ \textit{adjusted}\ R^2 = 0.752,\ p < 0.001$ (effect size 0.52, statistical power > 0.999). In contrast to the original study, we find that neither the *Long* nor the *Short* retirement condition has an effect on the variability of spending, with $F\,(1375) = 0.52,\ \textit{adj.}R^2 = 0.648,\ p < 0.471$ (effect size 0.01, statistical power 0.071). These results do not change qualitatively after controlling for *risk-taking* and *patience*.

As in the original study, we do not force subjects to automatically spend all points they have in the last period and allow them to end a round with the remaining points in the savings account.[27] As part of our additional analysis, we investigated the implication of this feature on the subjects' decision behavior.

[25] Mechanically, this is represented by negative involuntary savings forced upon subjects when their savings balance is smaller than the current period's mandatory expenses.

[26] For instance, suppose that a subject who, as periods advance, sees that the random realization of expenses will back-load the high expense periods during the retirement phase. This subject could decide to spend more during the lower expenses period, while his budget slack to spend is higher, even while knowing that he or she would eventually become bankrupt, in order to maximize lifetime spending before bankruptcy.

[27] However, we informed the subjects about this feature in the instructions. In addition, in the quiz that subjects had to pass before the main task, we tested whether they understood that the payoff would be determined by a randomly chosen period.

Table 10 presents the means of several experimental environmental and decision variables. Lifetime income is fixed at 1620 points and lifetime expenses at 720, leaving a budget of 900 points for lifetime consumption. However, the average observed lifetime spending ranges between 591 and 693 points only. This means that on average subjects are leaving substantial amounts of savings unspent at the end of their experimental life. We therefore classify subjects into three types according to their lifetime savings and spending pattern: 'bankrupt', 'endlife non-spenders', and 'effective planners'. 'Bankrupt' subjects did not save enough to cover the remaining mandatory expenses during retirement. 'Endlife non-spenders' did not spend all their points in the last period of a round, wasting them. All other subjects are 'effective planners'.

« Placeholder for Table 10 »

Subjects have saved, on average, 52.7% of their income and 72% of their available budget in the first round in the *Long* retirement condition and 36.9% and 57.1%, respectively, in the *Short* condition. In these same first rounds, 13.9% of the subjects went bankrupt and, of those who did, their average deficit (that is, $savings - expenses$) was 96 points in condition *Long*. So did 9% of the subjects in the first round *Short*, for an average deficit of 54 points. Furthermore, 68.7% of the subjects in condition *Long* ended the first round with an unspent savings balance (average savings lost of 383 points among those who did), as did 70.2% of the subjects in the condition *Short*. The fraction of subjects who did lose savings by not spending them appears high, but also does not change noticeably between rounds. We do not have original study results to compare its prevalence for each type of subject there.

The average savings and consumption paths for each type and for the entire sample are shown in Figure 5.

« Placeholder for Figure 5 »

Since income is increasing along the periods during the working phase, while expenses are not, savings and spending are naturally less constrained over time. In both treatments, 'bankrupt' subjects increase spending at a faster rate and save much less than other subjects. They also take too long, on average, to reduce consumption after retirement given their low savings. Subjects who leave unspent savings seem to spend too little (and save too much) throughout the periods, without other obvious decision patterns that might explain why they leave so much unspent savings behind.

3.6 Reproduction results of Meissner (2016)

We focus our reproduction on the main treatment effect of two symmetrically opposed income paths that need to be smoothed. In condition *Savings*, subjects face a downward income path and need to set aside part of their high income in earlier periods to offset low income in later periods. In condition *Borrowing*, subjects face an increasing income path and need to borrow (incur in negative savings, that is, debt) in earlier periods to be repaid (without interest) with future high income in later periods. Treatments differ in the sequence of conditions that the subjects face. In the *Savings First* treatment, the subjects play two rounds in the *Savings* path, then switch to *Borrowing* for another two rounds and vice versa for the *Borrowing First* treatment.[28]

To simplify the instructions, we modified the variables in the experimental environment of the original study. In our reproduction, subjects earn income in points and have variable incentives in the form of induced CARA utility, which is then converted into Euro ('Eurocent Rewards').[29] We simplified the variable incentive to be the lifetime sum of 'Eurocent Rewards' in one randomly chosen round. Unlike the original study, voluntary negative spending is not allowed in our design in periods before the last. In the last period, as in the original study, the subjects do not make decisions: spending (negative if necessary) is automatically calculated to impose the budget constraint of lifetime income equal to lifetime consumption.[30]

In total, 278 subjects completed the experiment, of whom 147 in the *Borrowing First* treatment and 131 in *Savings First*.[31]

« Placeholder for Figure 6 »

In Figure 6, we see – as in the original study – that subjects in the *Borrowing* condition have a greater variance in their consumption path than in *Saving*, and do not borrow against

[28]In the original study, subjects played three rounds of each condition.

[29]In the original experiment, subjects earned 'Talers' which they converted into utility-induced 'points', summed across each round and then converted into monetary units. We essentially turned the exchange into a currency bypassing an intermediate utility computational variable to reduce the complexity of the experimental task.

[30]Negative spending (not savings) is a very hard feature to conceptualize for subjects, which would require significant increase of the instruction set's size. In the original study, which allowed negative spending as an induced CARA utility function that could be defined in the negative domain, only 24 of 9120 subjects $\times$ period $\times$ round spending decisions were negative.

[31]In the original pre-registered plan, we had proposed to exclude subjects who, in a first attempt, got more than one mistake in the instruction quiz. This resulted in an unexpectedly high rejection rate that was not acceptable for our market research panel vendor. After one live day when only ten subjects completed the experiment, we suspended data collection, discarded these observations altogether, and restarted data collection the following day, with a relaxed restriction to allow two initial mistakes in a first attempt at the quiz while maintaining the requirement of no mistakes in a second attempt; see Subsection 3.1.

future income to smooth consumption in earlier periods. Income smoothing works better under the *Saving* condition, since its subjects have to save a part of the income they have already earned at present. However, it appears that, compared to the original study, the median consumption among *Saving* subjects in our study was not as smooth. The order effects of conditions *Borrowing First* or *Savings First* do not appear to affect the results in each treatment of our experiment.

« Placeholder for Figure 7 »

The deviation from optimal consumption (the dependent variable) also resembles the general stylized pattern in the original study. Its most interesting specification is *Measure 2* – the absolute deviation from conditionally optimal consumption,[32] shown in Figure 7. As in the original study, subjects deviate the most from optimal consumption paths when in the *Borrowing* condition (i.e., rounds 1-2 for *Borrowing First* and 3-4 for *Savings First*).

In bivariate analyzes with Mann-Whitney U tests, reported in Table 11, we find that the deviation and absolute deviation of conditionally optimal consumption (*Measure 1* and *2* as defined in the original study) are statistically significantly different between treatments, in all rounds (effect size – in the first round – 0.470 and 0.412; statistical power (0.05 level) 0.973 and 0.916 for *Measure 1* and *2* respectively). The utility loss from the deviation from the unconditionally optimal consumption path (*Measure 3*) is significant only for the first two rounds.

In an additional preregistered analysis, we control for the impact of *risk-taking* and *patience*, and find that the results are unchanged: treatment has significant effects on *Measure 1* and *Measure 2*, but not on *Measure 3*.

« Placeholder for Table 11 »

3.7 Reproduction results of Blaufus and Milde (2021)

For this reproduction, we are interested in the main treatment effects of different but economically equivalent taxation regimes on retirement savings decisions. The experiment consists of a 'working' phase and a 'rest' phase. During the working phase, subjects decide between saving and spending. Each round has 10 working periods (with fixed wages) and 5 resting periods. The treatment conditions vary the taxation regime for savings. In *Immediate* taxation, subjects pay income taxes immediately, but their savings are tax-free upon withdrawal

[32]Recalculated at every period for each subject.

during retirement. In *Deferred* taxation, subjects do not pay income taxes on their savings (they get a tax rebate from income taxes) but are taxed later when they withdraw savings during retirement. Finally, in a *Matching* condition, subjects receive matching contributions to their savings and pay taxes later, upon withdrawal, during retirement. The balance in all savings accounts earns an interest of 5% per period. Earned interest is taxable or tax-exempt according to the tax rule applied to the principal amount of savings. Withdrawals after retirement are automatically calculated and made equal for all periods of the rest phase.[33]

To simplify the experimental design, we removed an attention check and reassurance screen of tax return filings and integrated the projections of retirement income directly into the main interface screen. Further, we replaced the real effort task generating income in the working phase (a time-consuming transcribing task requiring printed handouts) with a simplified version of Gill and Prowse (2012) sliders' task. In terms of control variables, we retain age and gender, but use our own risk taking measure for identification of *High risk-taking* subjects taking its 75th percentile cut-off as in the original study. Furthermore, we use our measure of *financial ability* as a replacement for the original study measure of financial knowledge. Due to session-time constraints, we do not collect information on tax aversion or procrastination.

As in the original study, our main dependent variables are *saving rate* (naive rates compared to wages) and *effective savings rate* (which accounts for the different taxation regimes on withdrawal). With the tax rate ζ, the (naïve) savings rate for all treatment conditions is defined as $\left(\frac{savings}{wage(1-\zeta)}\right)$. The effective savings rate that makes the (after-tax) withdrawals economically equivalent to those in the *Immediate* condition is defined as $\left(\frac{savings}{wage(1-\zeta)}\right)\times(1-\zeta)$ for the *Deferred* condition. With the matching contribution rate ϕ, for the *Matching* condition the effective savings rate is defined as $\left[\frac{savings(1+\phi)}{wage(1-\zeta)}\right] \times (1-\zeta)$.

For our reproduction, we collected 486 valid responses, of which 172 in the *Immediate* treatment condition, 148 in *Deferred* and 166 in *Matching*. We first calculate the unconditional means of the compatible savings rates across treatments, with 95% confidence intervals (see Figure 8). As in the original study, we observe that the savings rates do not change significantly between the first and second round, and *Immediate* savings rates are higher than *Deferred* effective savings rate.

« Placeholder for Figure 8 »

[33]Interest is still paid on the savings balance during retirement, and accrued interest is considered when calculating the fixed withdraw amount for all rest periods.

Both savings measures are reasonably stable over periods, as their aggregate levels per period and round show in Figure 9.

« Placeholder for Figure 9 »

Following the analysis of the original study, we regressed savings rates and effective savings rates, observed at the subject × period × round level[34] on the binary indicators of treatment and the aforementioned covariates. The results of the estimation are presented in Table 12. All models include subjects of the *Immediate* treatment. For treatment contrasts, models (1-5) include *Deferred* subjects only, while models (6-10) add *Matching* subjects only.

« Placeholder for Table 12 »

Both treatment coefficients are statistically significant in all estimation models, and for all treatments, the magnitudes of our coefficients are similar to those of the original study. In our reproduction, both the *Deferred* and the *Matching* tax-protected savings schemes increase the base savings rate from the *Immediate* condition (models 1-3 and 6-8 in Table 12). In our estimation (2), *Deferred* subjects save 8.6% more of their income than *Immediate* subjects in the first round (Cohen's $f^2 > 0.99$). In estimation (8), subjects in the *Matching* condition have base save rates 7.8% higher than in the *Immediate* condition. Tax rebates and matching contributions appear to attract savings in nominal terms, as in the original study.

However, that comparison of base savings rates ignores the fact that, in both *Deferred* and *Matching* conditions, withdrawals will be taxed, whereas *Immediate* withdraws are tax-exempt. Evaluating effective savings rates – as in the original study – shows that the economically equivalent savings of *Deferred* subjects are on average 9% lower than the savings of subjects in the *Immediate* treatment group. *Matching* subjects, however, have an effective saving rate, which is on average 7.6% higher than *Immediate* after accounting for future withdrawal taxation. In other words, the *Matching* contribution tax regime generates higher average post-tax net pension savings than the baseline *Immediate* taxation scheme.

However, in contrast to the original study, we find that *male* was a significant negative predictor of savings rates between the *Immediate* and *Deferred* treatment group. Furthermore, in our reproduction, *High Risk Taking*'s coefficient was significant and positive in all specifications, while in the original study, this variable was not statistically significant. Furthermore, we find that *Period* has a positive and significant coefficient in our sample, while in the orig-

[34]Therefore, we have 10 observations per subject per round, covering its working periods.

inal study it had a significant negative coefficient. However, the effect is small in economic terms. In period 10, subjects in our sample would save 1 to 3% more from their income than in period 1. In the original study, savings and effective savings rates decrease over periods.

4 Discussion and implications for future research

Combining insights from our reproductions and the current state of various strands of experimental research, we discuss possible implications for future experimental design for studies on individual retirement decision-making. We also highlight some of the opportunities we envision for future controlled experiments that could address some limitations to the generalization of extant findings in the empirical field-based (non-experimental) literature. We also offer a modest suggestion of an agenda for future experimental research on retirement decision-making, considering the current state of the literature and its limitations.

4.1 Reproduction of modified tasks, task design features and implementation challenges

We reproduced most of the main effects of the five studies we reviewed. We introduced variable incentives where they were absent, compressed, or reduced the scope of the original studies to fit a short time limit, and we used simplified instructions for online general population samples. These modifications, as we found, did not systematically affect the main outcomes. This suggests the potential for adopting general features of simplified life-cycle experimental tasks, like those we used. Furthermore, this result suggests that the manipulations and treatment conditions could be layered on the suggested flexible basic design, echoing Koehler, Langstaff, and Liu (2015). Some important considerations and precautions, discussed in the following, should still apply to future experiments.

We observed that the subject's consumption smoothing still is fairly sub-optimal. This failure concerns specifically two groups: a smaller group of subjects who do not save enough and go into experimental bankruptcy, and a larger group of subjects that leaves unspent savings (foregoing part of their potential compensation) at the end of a life-cycle task.

In particular, underconsumption (or oversaving) in later life periods has been identified in other intertemporal allocation tasks outside the context of retirement-like decision-making (e.g. Yamamori, Iwata, & Ogawa, 2018). In our experiments, this behavior was extensively observed. Future experiments that impose lifetime budget constraints and then study lifetime outcomes (such as induced utility from spending or consumption in all periods) should be concerned about the impacts of these constraints that self-resolve in the last period. Simultaneous aggregation of lifetime utility from subjects who leave money unspent at the end of a round and subjects who end a round in debt or go bankrupt does not allow distinguishing between these different decision-making anomalies. If bankrupt subjects and subjects with unspent savings are both present in the sample, while some concave utility is induced, and the task imposes automatic decisions in the last period to meet a lifetime budget constraint, then aggregated results might not show such inefficient decisions in opposing directions. Additionally, the coefficients of the treatment effects could be downward-biased.

Our strict subject retention criteria eliminated more than half of all subjects initially recruited through our market panel vendor (see Subsection 3.2). We suggest that future researchers adopt similar quality control measures using similar online panels elsewhere to reduce noise in the data generated by subjects with limited participation or interest. Departing from the usual practices, we allowed subjects to proceed immediately from instructions to a practice round and a quiz afterward. We did not pay any compensation (not even a show-up fee) for subjects who did not pass the post-trial quiz. With such procedures, we attempted to impose a minimum engagement that resembles the requirements of subjects in an in-person lab session to successfully answer all questions of a quiz correctly, before being allowed to proceed further. Meanwhile, we allowed the subjects to revisit the instructions throughout the quiz and all subsequent tasks. Our implementation allowed for the use of a much more representative sample of the adult population than typical university student pools. We hope that our experience encourages fellow researchers to use representative online samples in future experimental studies on life-cycle optimization that might seem *prima facia* impracticable to implement with such a sample and medium.

We reduce the number of discrete periods and/or rounds. Such changes did not materially affect the results of the original longer designs. A further drastic reduction to fewer periods should be approached with caution to avoid degenerating the natural computational and sequencing complexity present in life cycle optimization decisions in the field or in the laboratory.[35]

Further experiments might help inform the particular impact of other features on life-cycle experimental tasks. These often sidestep any implementation of time-discounting factors

[35]Discrete life length of less than 15 periods is uncommon both in the experimental and numerical optimization literature on optimization over the life-cycle.

across periods, other than interest on savings. Relatively complex utility forms can be imposed through incentive-reward functions. However, our knowledge is still limited on how subjects would react if decisions were measured non-parametrically (as in Abdellaoui, Attema, & Bleichrodt, 2010), when – for instance – longevity uncertainty and changes in the institutional environment are simultaneously introduced in the same task.

4.2 Heuristics

The overall complexity of life-time optimization and the cognitive demands it places on the average person who actually makes retirement decisions should attract more systematic studies on the specific heuristics and rules of thumb adopted by subjects with respect to the different characteristics of those decisions. Such characteristics refer to the longevity risk profile, income uncertainty, survival expenses, probability, and size of stochastic shocks, or risky investment choices in retirement savings. The use of heuristics in individual decisions and the possible biases embedded in these decisions could extend beyond the issue of whether voluntary retirement savings levels adhere to some normative model optimality (Benartzi & Thaler, 2007).

Winter, Schlafmann, and Rodepeter (2012) show that utility losses relative to the combined adoption of simple heuristics do not accrue substantially in relation to optimal solutions from a normative perspective of standard intertemporal preferences. If utility losses are in general low, a social planner might promote and disseminate the use of rules-of-thumb to increase savings, or achieve any other desired aggregate outcome. There is some survey evidence (Binswanger & Carman, 2012) implying that engagement with retirement financial preparation through rules-of-thumb can substitute for strategic planning, producing better outcomes in retirement savings wealth than those who do not adopt any structured approach. The potential of stylized simple rules to improve retirement planning should be investigated in more depth when interacted with different characteristics of retirement decisions.

More experimental work, based on the Carbone (2005) finding of the limited relevance of demographic characteristics, is also necessary to identify the impact of contextual rules of thumb that individuals might acquire from their social environment, peers, financial advisors, media, and institutions. Another potential open question is whether subjects are more likely to stick to a savings plan when it rests on easy-to-understand endogenous rules than on outcome-based recommendations re-optimized every period.

Taking a step back, for some features of retirement decisions, the theoretical or simulation-based literature does not even sufficiently agree on what the necessary assumptions are for certain normative outcomes, as in the unsettled and unsolved annuity puzzle. Simulations conducted by Peijnenburg, Nijman, and Werker (2016) question some previous assumptions about the (lack of) attractiveness of pension wealth annuitization for many subjects. They imply that normative prescriptions for rational annuitization decisions are less likely to breakdown than, for example, in the earlier work of Davidoff, Brown, and Diamond (2005) or Brown et al. (2008). Experimental work should contribute to assessing how individuals break down the complex input of decisions such as the annuitization choice and their interactions with other factors that determine decision behavior in controlled settings. This will possibly allow descriptive models to emerge and better explain whether, why, and to what extent subjects should (or not) annuitize their pension wealth. To achieve this goal, larger experiments are required to simultaneously implement several key features of the annuitization decision. This would allow evaluating the interaction and combined outcome effects of different heuristics that subjects adopt to address different features of the annuitization decision. Such experiments could also identify consistent subject types with regard to annuitization preferences.

4.3 Institutional setting and mandatory-participation schemes

Many countries have some form of mandatory participation pension scheme, in various forms and scopes. Such compulsory schemes complicate the empirical analysis of field data on voluntary retirement decisions, as they are bounded by their institutional setup. These mandatory schemes (such as presented in Bohr, Holt, & Schubert, 2019; Hurwitz, Sade, & Winter, 2020) and the legislation that enables them are rarely, if ever, harmonized between countries, making it difficult to compare them even on a country-by-country aggregate basis. There are several opportunities to investigate how other characteristics of individual voluntary retirement decisions are impacted by obligatory schemes that force people to save; redistribute mortality risk (other than annuitization), income, and investment risks across individuals; or provide exogenous income support at old ages.

Pension reforms, often focusing on increasing individual control over certain financial choices in retirement, have produced a large empirical literature evaluating the impact of these reforms (see Gough & Niza, 2011, for an overview). In recent decades, these reforms have commonly relaxed compulsory elements, creating opt-outs, flexible arrangements, or similar changes. Pension reforms are often paired with simultaneous tax or social welfare reforms, complicating the proper identification of the impact of pension reforms on individual retirement decisions.

In this context, further experimental research should address the impact of institutional rule changes within controlled environments, and collective risk-sharing features and solidarity mechanisms between individuals not restricted to a single institution or market (Tausch, Potters, & Riedl, 2013). More specifically, the complex features of retirement decisions should be added to more stylized and simple risk-sharing experiments.

4.4 Nudges and choice architecture

Certain characteristics of retirement decisions, discussed in Section 3, make them ex-ante good candidates for many categories of behaviorally informed interventions and modifications in the field. This is the case for the conditional commitment device designed and studied by Thaler and Benartzi (2004).

The power of changing default options within constrained choice settings (Madrian & Shea, 2001) has been found to be one of the most effective interventions when one of the options in the choice menu is assumed to be optimal or dominant by policymakers or a social planner. So far, the decision-making mechanisms that act at the individual level to generate the outcomes found in field interventions have not been fully understood.

Field interventions using nudges or changes in decision architecture are more often than not highly specific to a single institution (pension fund or company) and their particular arrangements. As there is substantial heterogeneity across long-term savings in households (Bernheim, Skinner, & Weinberg, 2001), there are opportunities to expand the understanding of the factors driving heterogeneous response to such interventions in the first place, and whether such heterogeneity is orthogonal to other ex-ante motivators of different levels of savings, different investment risk-taking and so forth.

4.5 Financial literacy, education and information

Relatively simple decision-making biases could result in substantial differences in financial outcomes of retirement decisions. Undersaving at younger ages could lead to severe budget constraints upon retirement, for example, a potentially serious concern given the lower interest of younger and lower-income individuals in planning for retirement (Clark, Knox-Hayes, & Strauss, 2009). Low numeracy can lead to substantial misunderstandings of the expected outcomes of retirement portfolio allocation. Longevity pessimism (Wu, Stevens, & Thorp, 2015) could lead to an underestimation of the retirement savings needed at any level of consumption. The economic benefits could not be fully understood by people whose marginal utility would be the highest.

All these stylized outcomes could and often have been subject to financial literacy or financial education interventions in the field. These interventions consist of making outcomes more explicit, providing people with simulation and planning tools, or training them to be more aware of their own biases. On the one hand, high financial literacy scores (or reliance on professional advice) are associated with improved general investment outcomes (Von Gaudecker, 2015). On the other hand, considering the timeline of common financial retirement decisions, the long-term effectiveness of financial education might be questionable (Fernandes, Lynch, & Netemeyer, 2014). There may also be ethical considerations related to the effort and personal commitment that effective training programs could require from participants (Willis, 2009).

Ambitious longitudinal lab-in-the-field experiments should tackle the long-term impact of financial education programs that aim to improve individual decision-making on repeated decisions, with limited learning opportunities before their main outcomes are revealed after retirement. Future research should also explore heterogeneity on the willingness to participate in retirement decision-oriented financial education programs and identify possible interactions of unwillingness to participate and ex-post dissatisfaction with experimental outcomes.

More research is also warranted on all aspects of institutional communication on pension decisions. For instance, it is common for financial institutions (such as pension funds) offering open-market annuities or the choice between lump-sum and annuities to prominently display a comparison between one large certain lump-sum and multiple conditional small periodical payoffs. This could be challenging for certain groups of individuals deciding on whether to buy an annuity by relinquishing a large sum of money. Likewise, portfolio allocation online interfaces often focus on annual returns and volatility, instead of compounding them over multi-decade horizons. We still do not know enough about how the cognitive demands for the more fundamental life-cycle smoothing task interact with these other informational characteristics of the decision environment that individuals face in the field.

5 Conclusion

Individual retirement financial decisions are complex, which makes them prone to magnification of biases and cognitive mistakes in their outcomes. The sub-optimal outcomes are likely persistent, since retirement saving decisions also offer limited learning opportunity due to long lags between time of decision-making and of its outcomes. Experimental research on retirement decisions and on how heterogeneous individuals engage in these decisions is thus particularly helpful to advance our understanding of many empirical field outcomes that cannot be easily reconciled with theoretical normative models addressing these decisions.

To that end, we redesigned five experimental studies, each addressing different topics and incorporating different features of the retirement decision problem, and attempted to reproduce their main findings. We used reduced-scope tasks and/or a simplified decision environment to make the tasks suitable for implementation with online samples of a general adult population in incentivized settings. We reproduced most of the main effects of the original studies we selected for this exercise, which might raise the external validity of their findings.

Finally, the persisting limitations of the extant simulation-based and field empirical literature, on several topics concerning retirement decision making, offer a promising future research agenda for experimental research, which we proposed and briefly commented on.

References

Abdellaoui, M., Attema, A. E., & Bleichrodt, H. (2010). Intertemporal Tradeoffs for Gains and Losses: An Experimental Measurement of Discounted Utility. *Economic Journal*, *120*(545), 845–866.

Agnew, J. R., Anderson, L. R., Gerlach, J. R., & Szykman, L. R. (2008). Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults. *American Economic Review*, *98*(2), 418–422.

Agnew, J. R., Anderson, L. R., & Szykman, L. R. (2015). An Experimental Study of the Effect of Market Performance on Annuitization and Equity Allocations. *Journal of Behavioral Finance*, *16*(2), 120–129.

Anderhub, V., Güth, W., Müller, W., & Strobel, M. (2000). An experimental analysis of intertemporal allocation behavior. *Experimental Economics*, *3*(2), 137–152.

Ballinger, T. P., Hudson, E., Karkoviata, L., & Wilcox, N. T. (2011). Saving behavior and cognitive abilities. *Experimental Economics*, *14*(3), 349–374.

Ballinger, T. P., Palumbo, M. G., & Wilcox, N. T. (2003). Precautionary Saving and Social Learning Across Generations: an Experiment. *The Economic Journal*, *113*(490), 920–947.

Benartzi, S., & Thaler, R. H. (2007). Heuristics and Biases in Retirement Savings Behavior. *Journal of Economic Perspectives*, *21*(3), 81–104.

Bernheim, B. D., Skinner, J., & Weinberg, S. (2001). What Accounts for the Variation in Retirement Wealth Among U.S. Households? *American Economic Review*, *91*(4), 832–857.

Beshears, J., Choi, J. J., Harris, C., Laibson, D., Madrian, B. C., & Sakong, J. (2020). Which early withdrawal penalty attracts the most deposits to a commitment savings account? *Journal of Public Economics*, *183*, 104144.

Binswanger, J., & Carman, K. G. (2012). How real people make long-term decisions: The case of retirement preparation. *Journal of Economic Behavior and Organization*, *81*(1), 39–60.

Blaufus, K., & Milde, M. (2021). Tax Misperceptions and the Effect of Informational Tax Nudges on Retirement Savings. *Management Science*, *67*(8), 5011–5031.

Bohr, C. E., Holt, C. A., & Schubert, A. V. (2019). Assisted savings for retirement: An experimental analysis. *European Economic Review*, *119*, 42–54.

Brown, A. L., Chua, Z. E., Camerer, C. F., & Brown, A. L. (2009). Learning and Visceral Temptation in Dynamic Saving Experiments. *Quarterly Journal of Economics*, *124*(1), 197–231.

Brown, J. R., Kling, J. R., Mullainathan, S., & Wrobel, M. V. (2008). Why Don't People Insure Late-Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle. *American Economic Review*, *98*(2), 304–309.

Carbone, E. (2005). Demographics and Behaviour. *Experimental Economics*, *8*(3), 217–232.

Carbone, E. (2006). Understanding intertemporal choices. *Applied Economics*, *38*(8), 889–898.

Carbone, E., & Duffy, J. (2014). Lifecycle consumption plans, social learning and external habits: Experimental evidence. *Journal of Economic Behavior and Organization*, *106*, 413–427.

Carbone, E., & Hey, J. D. (2004). The Effect of Unemployment on Consumption: An Experimental Analysis. *Economic Journal*, *114*(497), 660–683.

Carbone, E., & Infante, G. (2014). Comparing behavior under risk and under ambiguity in a lifecycle experiment. *Theory and Decision*, *77*(3), 313–322.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Clark, G. L., Knox-Hayes, J., & Strauss, K. (2009). Financial Sophistication, Salience, and the Scale of Deliberation in UK Retirement Planning. *Environment and Planning A: Economy and Space*, *41*(10), 2496–2515.

Davidoff, T., Brown, J. R., & Diamond, P. A. (2005). Annuities and Individual Welfare. *American Economic Review*, *95*(5), 1573–1590.

Duffy, J., & Li, Y. (2019). Lifecycle consumption under different income profiles: Evidence and theory. *Journal of Economic Dynamics and Control*, *104*, 74–94.

Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fatas, E., Lacomba, J. A., & Lagos, F. (2007). An Experimental Test on Retirement Decisions. *Economic Inquiry*, *45*(3), 602–614.

Feigenbaum, J., Gahramanov, E., & Tang, X. (2013). Is it really good to annuitize? *Journal of Economic Behavior and Organization*, *93*, 116–140.

Feltovich, N., & Ejebu, O.-Z. Z. (2014). Do positional goods inhibit saving? Evidence from a life-cycle experiment. *Journal of Economic Behavior & Organization*, *107*(PB), 440–454.

Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, *60*(8), 1861–1883.

Gechert, S., & Siebert, J. (2020). Preferences over wealth: Experimental evidence. *Journal of Economic Behavior & Organization.*

Gill, D., & Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, *102*(1), 469–503.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645.

Gough, O., & Niza, C. (2011). Retirement Saving Choices: Review of the Literature and Policy Implications. *Journal of Population Ageing*, *4*(1-2), 97–117.

Heimer, R. Z., Myrseth, K. O. R., & Schoenle, R. S. (2019). YOLO: Mortality Beliefs and Household Finance Puzzles. *Journal of Finance*, *74*(6), 2957–2996.

Hey, J. D., & Dardanoni, V. (1988). Optimal Consumption Under Uncertainty: An Experimental Investigation. *The Economic Journal*, *98*(390), 105.

Hurwitz, A., Sade, O., & Winter, E. (2020). Unintended consequences of minimum annuity laws: An experimental study. *Journal of Economic Behavior and Organization*, *169*, 208–222.

Koehler, D. J., Langstaff, J., & Liu, W. Q. (2015). A simulated financial savings task for studying consumption and retirement decision making. *Journal of Economic Psychology*, *46*, 89–97.

Levy, M. R., & Tasoff, J. (2020). Exponential-growth Bias in Experimental Consumption Decisions. *Economica*, *87*(345), 52–80.

Lugilde, A., Bande, R., & Riveiro, D. (2019). Precautionary Saving: a Review of the Empirical Literature. *Journal of Economic Surveys*, *33*(2), 481–515.

Madrian, B. C., & Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *Quarterly Journal of Economics*, *116*(4), 1149–1187.

Meissner, T. (2016). Intertemporal consumption and debt aversion: an experimental study. *Experimental Economics*, *19*(2), 281–298.

Meissner, T., & Rostam-Afschar, D. (2017). Learning Ricardian Equivalence. *Journal of Economic Dynamics and Control*, *82*, 273–288.

Peijnenburg, K., Nijman, T., & Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, *70*, 18–35.

Tausch, F., Potters, J., & Riedl, A. (2013). Preferences for redistribution and pensions. What can we learn from experiments? *Journal of Pension Economics and Finance*, *12*(3), 298–325.

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, *112*(1), S164–S187.

Von Gaudecker, H. M. (2015). How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice? *Journal of Finance*, *70*(2), 489–507.

Willis, L. (2009). Evidence and Ideology in Assessing the Effectiveness of Financial Literacy Education. *San Diego Law Review*, *46*(2008-6).

Winter, J. K., Schlafmann, K., & Rodepeter, R. (2012). Rules of Thumb in Life-Cycle Saving Decisions. *Economic Journal*, *122*(560), 479–501.

Wu, S., Stevens, R., & Thorp, S. (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase. *Journal of Economic Dynamics and Control*, *55*, 39–56.

Yamamori, T., Iwata, K., & Ogawa, A. (2018). Does money illusion matter in intertemporal decision making? *Journal of Economic Behavior and Organization*, *145*, 465–473.

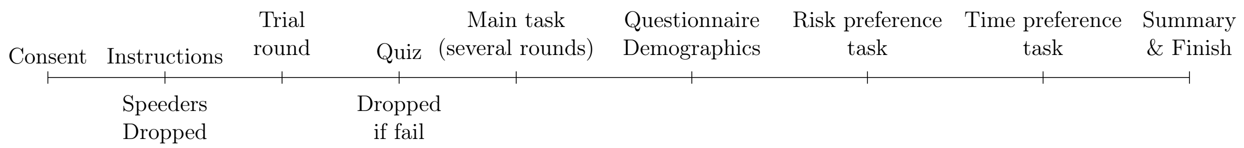

Figure 1: Sequence of the steps for each reproduction

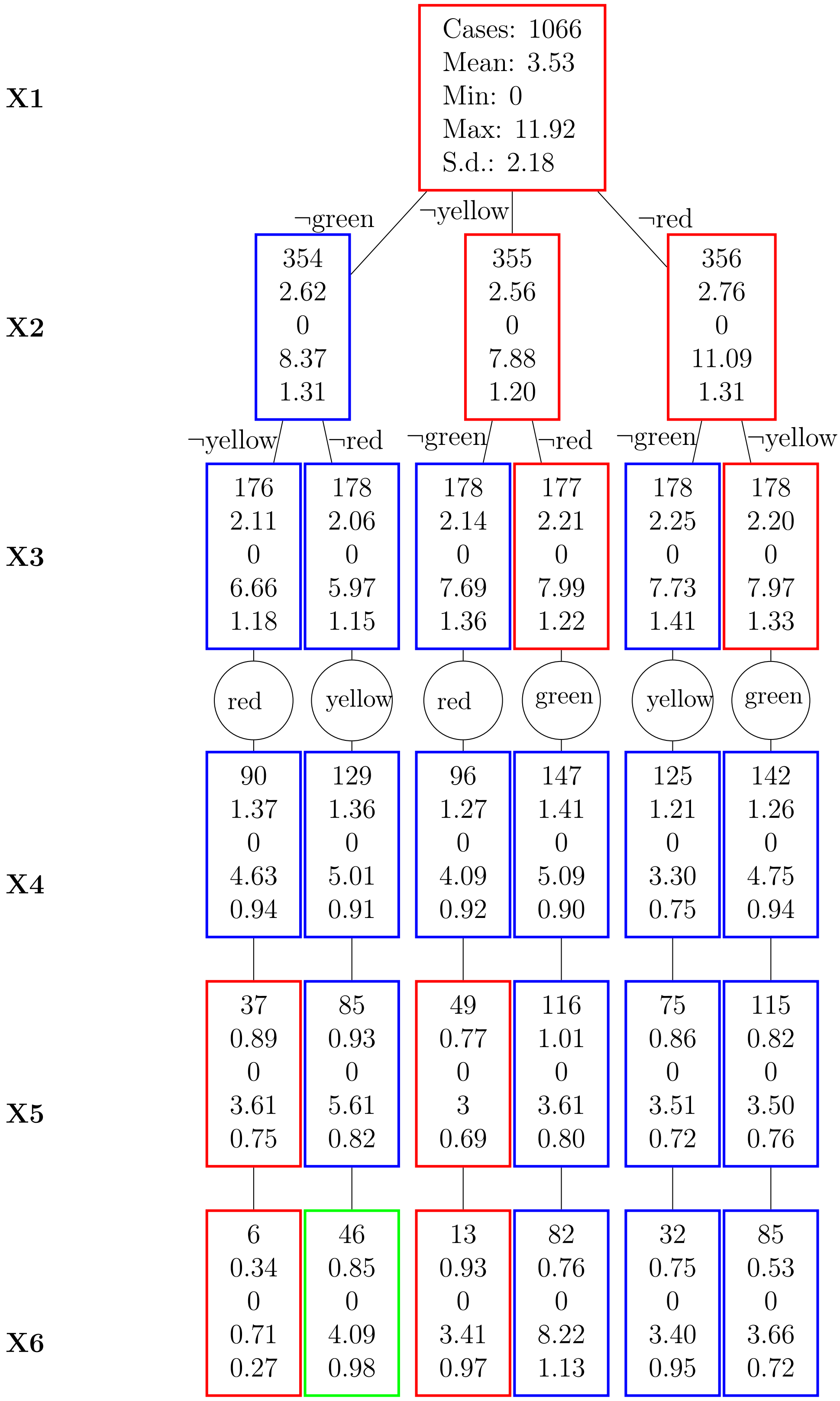

Figure 2: **Anderhub et al. (2000) Reproduction: Average Behavior on Product Treatment**. The labels on the left X1 to X6 indicate the period of the spending decision for *Product* condition. Each node box includes number of observations, mean, maximal, minimal, and standard deviation. The color of the frames indicate the observed mean of spending decision compared to the optimal spending paths: red means over-spending, blue means under-spending, and green means same as the optimal spending.

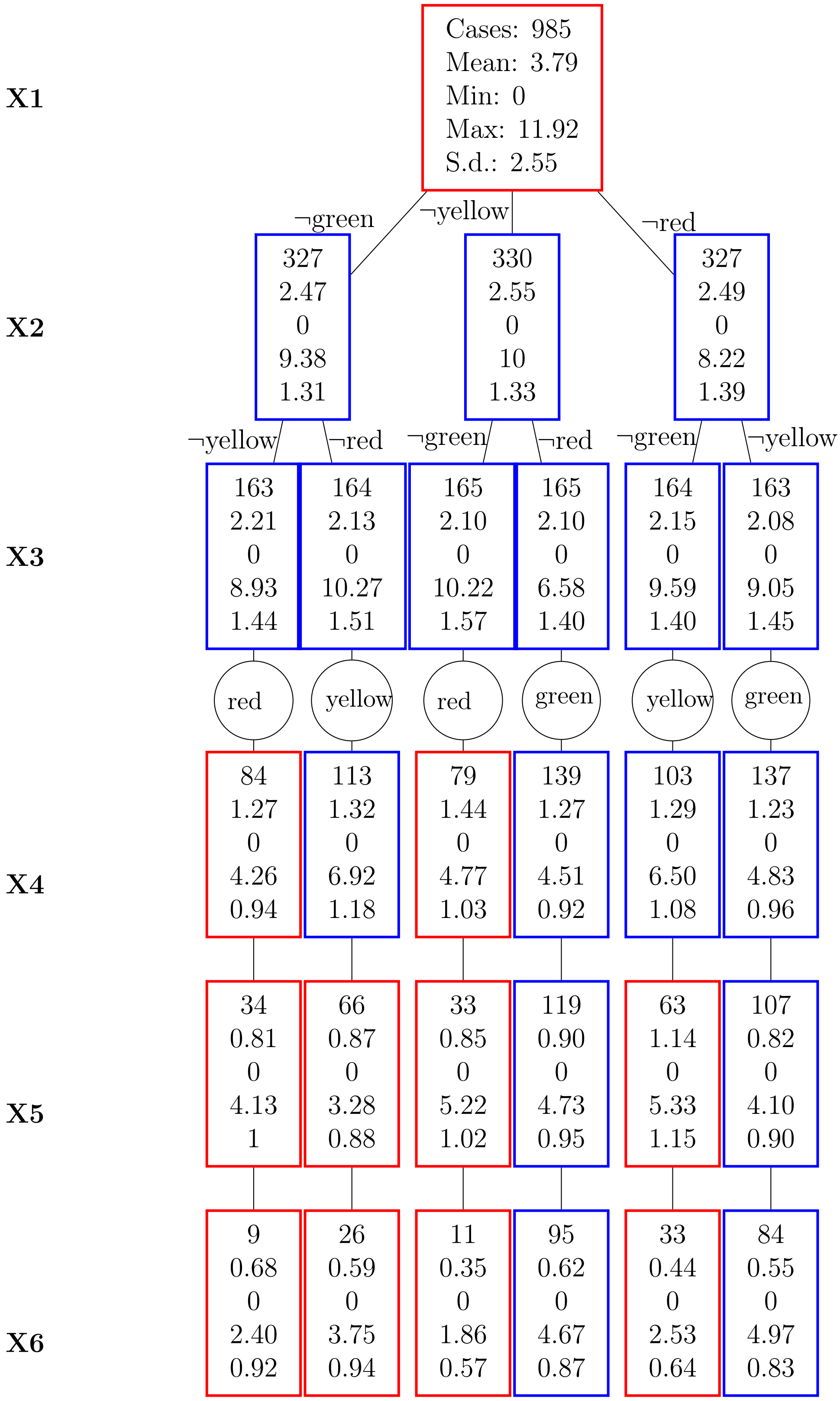

Figure 3: **Anderhub et al. (2000) Reproduction: Average Behavior on Summation Treatment**. The labels on the left X1 to X6 indicate the period of the spending decision for *Summation* condition. Each node box includes number of observations, mean, maximal, minimal, and standard deviation. The color of the frames indicate the observed mean of spending decision compared to the optimal spending paths: red means over-spending, blue means under-spending, and green means same as the optimal spending.

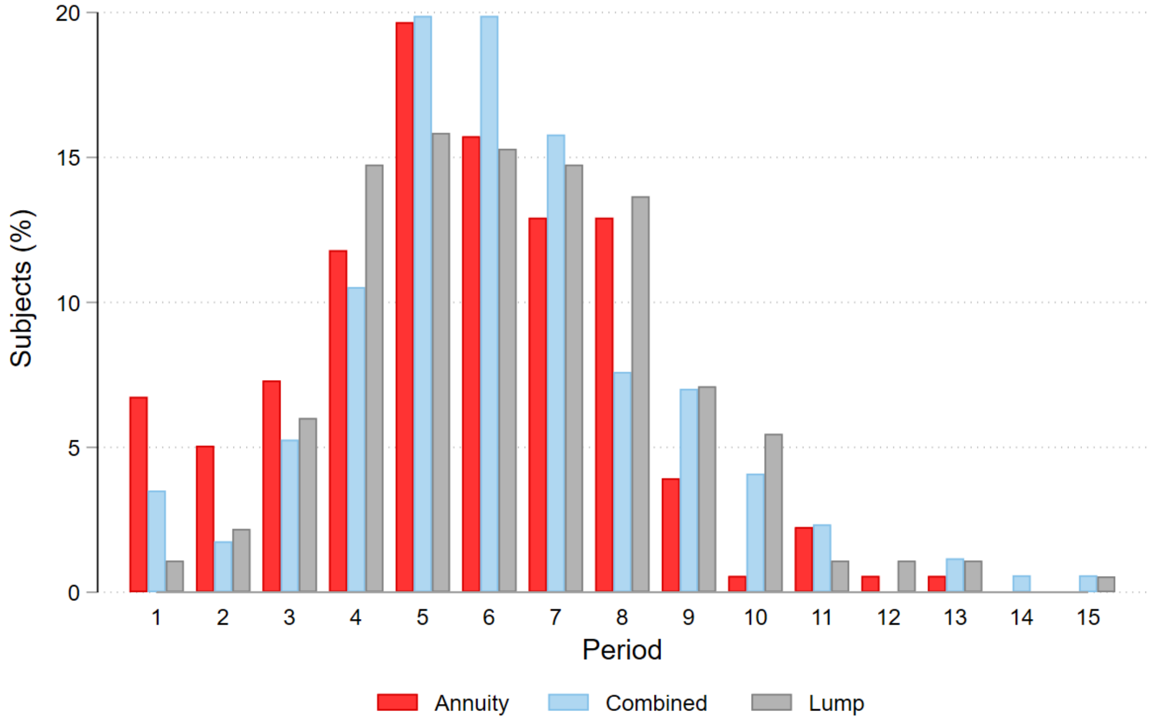

Figure 4: **Fatas, Lacomba, and Lagos (2007) Reproduction: Timing of Retirement**.
Period chosen by subjects to (start) collecting payoffs, conditional on not having been terminated.

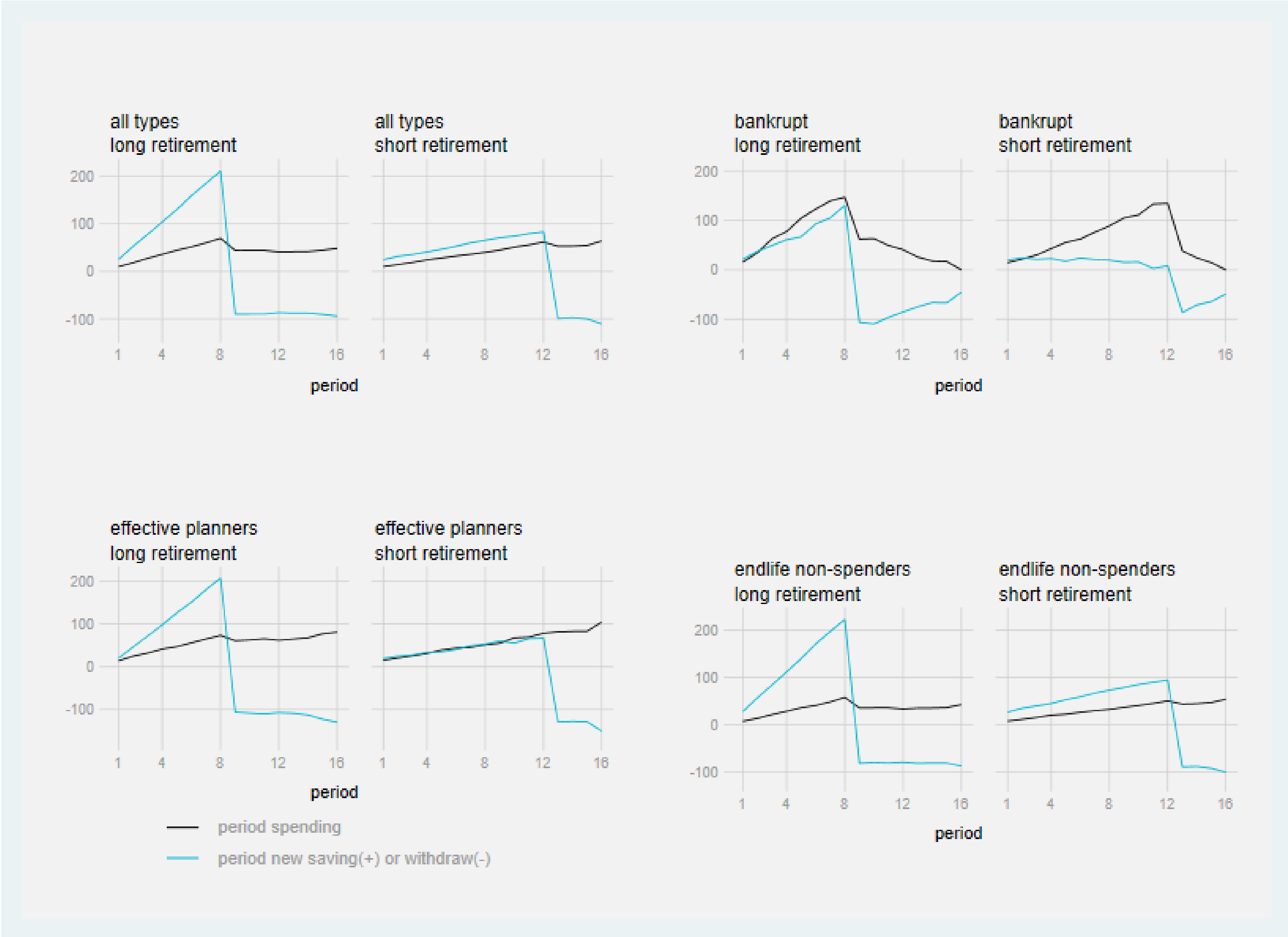

Figure 5: **Koehler, Langstaff, and Liu (2015) Reproduction: Spending and Saving per Period.** Negative savings are withdraws in the rest phase. *Bankrupt* subject types saved below mandatory expenses, *endlife non-spenders* left unspent savings at last period, and *effective planners* did neither.

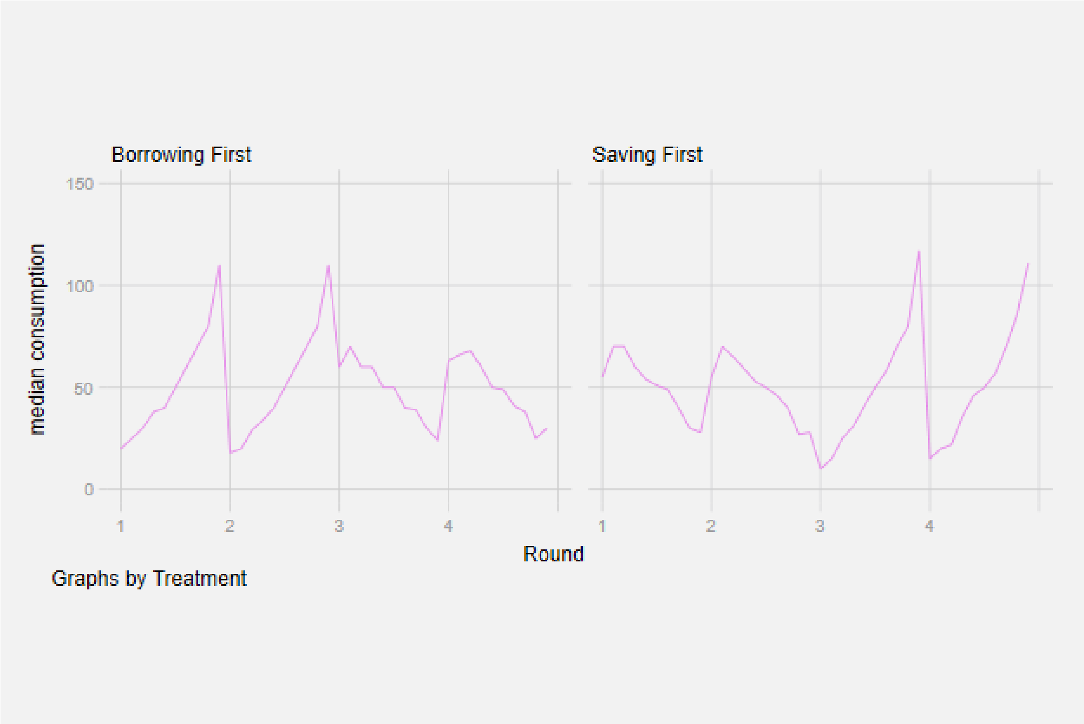

Figure 6: **Meissner (2016) Reproduction: Median Consumption per Period over Sequential Rounds.** Borrowing (saving) first subjects play rounds 1 and 2 in the borrow (saving) condition; and rounds 3 and 4 in the saving (borrow) condition.

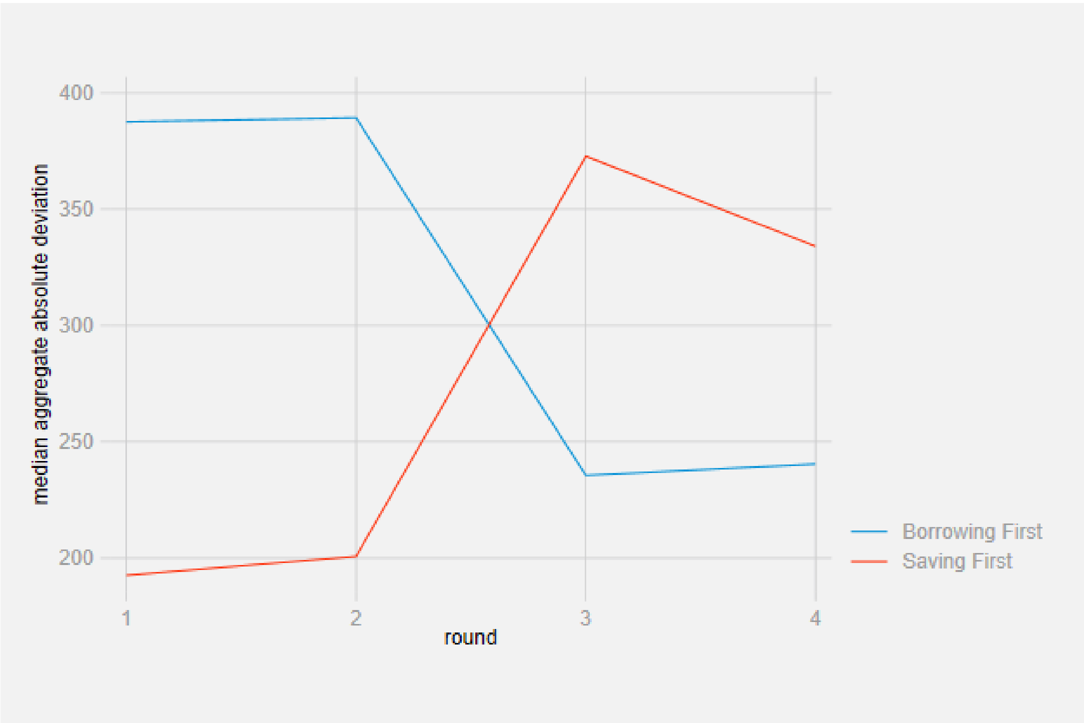

Figure 7: **Meissner (2016) Reproduction: Sub-Optimal Consumption.** Medians of *Measure 2* (mean absolute deviation of consumption from optimal path at each round, per subject X round) by treatment condition. Borrowing (saving) first subjects play rounds 1 and 2 in the borrow (saving) condition; and rounds 3 and 4 in the saving (borrow) condition.

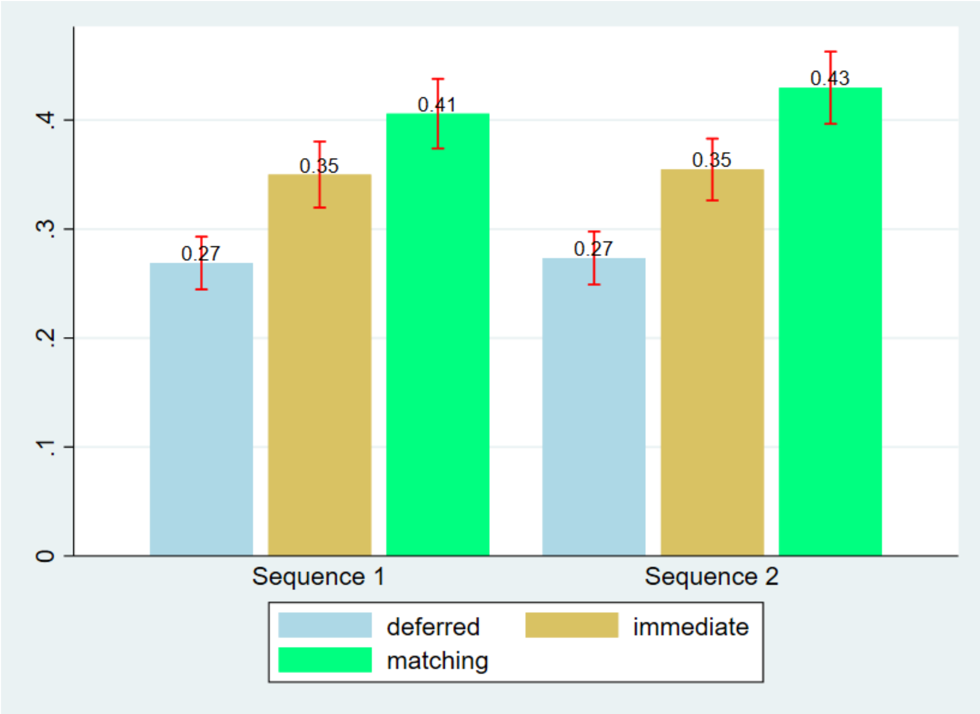

Figure 8: **Blaufus and Milde (2021) Reproduction: Average Savings Rates.** Direct (total) saving rates used for *Immediate* condition, and effective savings rates for *Deferred* and *Matching*, per round.

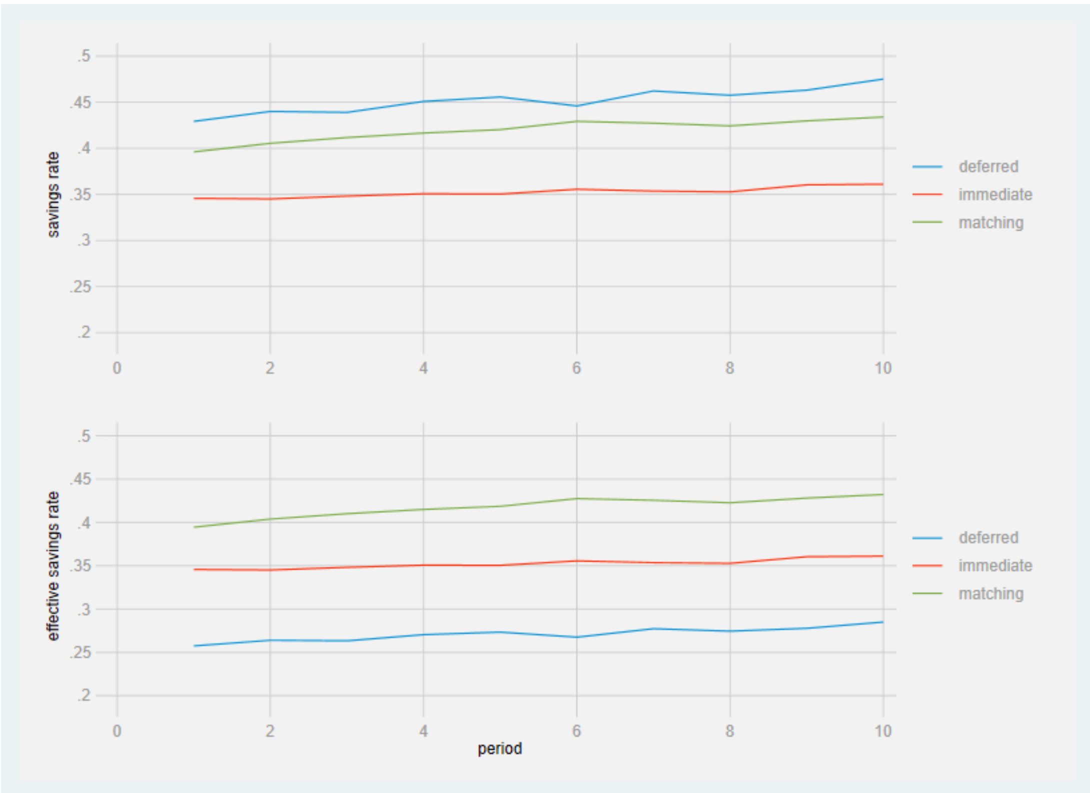

Figure 9: **Blaufus and Milde (2021) Reproduction: Savings Persistence.** Average (effective) saving rates per period across rounds.

Table 1: **An Overview of Experimental Studies on Retirement Decision-Making**. *N* is the number of observations. *MP* is whether there are Multiple Period decisions per Life. *LR* is Longevity Risk. *InU* is Income Uncertainty. *ME* is Mandatory Expenses. *IoS* is Interest on Savings. *ELB* is Enforced Lifetime Budget. *SWRP* is Separate Work and Retirement Phase. *IU* is Induced Utility. *Sample* includes the population where the sample is drawn and the platform (lab or online). The gray-honed rows are the five studies that are included in this reproduction study.

Study	Focus	Main Dependant Variable(s)	Main Finding(s)	N	MP	LR	InU	ME	IoS	ELB	SWRP	IU	Sample
Agnew, Anderson, and Szykman (2015)	biases	demand for annuity	Past market performance influences the demand for annuities.	1093	X	X		X					general population (lab)
Anderhub et al. (2000)	decision problem feature	consumption	Observed consumption paths are qualitatively correct with respect to optimal ones.	50	X	X						X	students (lab)
Ballinger et al. (2011)	heterogeneity	decision performance	Cognitive abilities predict performance.	192	X		X					X	students (lab)
Ballinger, Palumbo, and Wilcox (2003)	learning	consumption	Later generations perform better than earlier generations.	36	X		X					X	students (lab)
Beshears et al. (2020)	institutional features	endowment allocation to commitment account	Higher early-withdrawal penalties attract more commitment account deposits.	1045									Rand American Life Panel
Blaufus and Milde (2021)	biases	savings rate	Matching contributions attract higher savings than deferred or immediate taxation regimes.	306	X				X	X	X		students (lab)
Bohr, Holt, and Schubert (2019)	institutional features	optimal consumption	Mandatory (vs. voluntary) savings improves total lifetime consumption.	45	X				X	X	X	X	students (lab)
Brown et al. (2009)	learning	deviation from optimal consumption	Subjects save too little at first, but learn to save optimally over repeated life-cycles.	72	X		X			X		X	students (lab)
Brown et al. (2008)	biases	choice of life annuity	Individuals prefer an annuity over alternative products when presented in a consumption frame; non-annuitized products are preferred when presented in an investment frame.	1342									Internet survey (participants age > 50)
Carbone (2005)	heterogeneity	consumption	There is only a minor link between the strategies employed by the subjects and their demographic characteristics.	495	X		X		X				CentER family expenditure panel
Carbone (2006)	biases (general)	consumption	Discounting model gives the best explanation, but subjects are myopic.	594	X		X		X			X	CentER panel and students
Carbone and Duffy (2014)	learning	consumption	Provision of social information on past average levels of consumption results in a greater deviation of consumption from optimal paths.	60	X				X			X	students (lab)
Carbone and Hey (2004)	decision problem feature	deviation from optimal consumption	Over-sensitivity of consumption to income changes due to unemployment.	96	X				X			X	students (lab)
Carbone and Infante (2014)	decision problem feature	consumption-to-wealth	Ambiguity (vs. risk) triggers savings	30	X		X		X			X	students (lab)
Duffy and Li (2019)	institutional features	optimal consumption	100% pension replacement rate yields the highest experimental payoff.	119	X				X		X	X	students (lab)
Fatas, Lacomba, and Lagos (2007)	institutional features	choice of retirement period	Subjects retire later with lump-sum payoffs instead of annuities or combination thereof.	82		X							students (lab)
Feltovich and Ejebu (2014)	learning	optimal saving	Inter-personal comparisons (by assigning subjects to groups and displaying rankings based partly on consumption) increases under-saving and leads to lower money earnings.	170	X		X				X	X	students (lab)
Gechert and Siebert (2020)	biases (general)	savings	Participants on average form and maintain a stock of wealth although not optimal.	180	X		X						students (lab)
Hey and Dardanoni (1988)	decision problem feature	consumption	Actual behavior differs significantly from optimal behavior; the comparative static implications of actual behavior appear to be optimal	128	X	X			X			X	students (lab)
Hurwitz, Sade, and Winter (2020)	institutional features	division of savings between annuity and lump sum	Providing a mandatory minimum annuity rule creates an anchoring effect that reduces annuitization.	277				X					students (lab)
Koehler, Langstaff, and Liu (2015)	decision problem feature	accumulated savings at start of retirement	Most subjects save enough, and longer retirement attracts higher saving	149	X			X		X	X		M-Turk
Levy and Tasoff (2020)	biases	overconsumption	Observed behavior consistent with behavior predicted by exponential growth bias.	399	X				X			X	students (lab)
Meissner (2016)	decision problem feature	deviation from consumption smoothing	Consumption smoothing is worse when subjects need to borrow from the future than save from the present.	76	X		X			X		X	students (lab)
Meissner and Rostam-Afschar (2017)	learning	consumption / save (borrow)	Some subjects learn to comply with Ricardian Equivalence.	176	X		X			X			students (lab)

Table 2: **Overview of Attrition, Payoff and Completion Time**

Panel A: **Subject Attrition.** Subject participation according to their furthest stage reached per experiment. *Dropped out at instructions* include subjects rejected for having skipped through instruction screens too fast (10s to 60s threshold depending on experiment). *Dropped out at quiz* include subjects rejected for failing to answer a five or six multiple-choice quiz on experimental instructions, after the trial round.

	Koehler *et al* (2015)		Meissner (2016)		Blaufus & Milde (2021)		Anderhub *et al* (2000)		Fatas *et al* (2007)	
	Obs	%	Obs	%	Obs	%	Obs	%	Obs	%
No consent	64	5.5	77	5.9	128	7.4	89	9.3	80	7.6
Drop out at instructions	160	13.9	187	14.4	290	16.8	104	10.9	92	8.8
Drop out at trial round	101	8.8	87	6.7	153	8.8	85	8.9	45	4.3
Drop out or failed quiz	138	12.0	109	8.4	91	5.2	193	20.2	113	10.8
Drop out during tasks	347	30.1	557	43.0	558	32.1	146	15.3	187	17.9
Finished	344	29.8	278	21.5	518	29.8	339	35.5	530	50.6
Total	1154	100.0	1295	100.0	1738	100.0	956	100.0	1047	100.0

Panel B: **Payoff (Euro) and Completion Time** *Payoff*, in Euro, is the sum of variable incentivized payoff for the main experiment and the payoffs of the risk-taking and patience tasks. *Total time* is the total time (in minutes) that the subjects spent to finish the experiment. The large number of total time in the row *Max* comes from the subjects who finished the experiment but did not click Finish in the end. The last row summarizes the number of observations where the total time is longer than 65 minutes.

	Koehler *et al* (2015)		Meissner (2016)		Blaufus & Milde (2021)		Anderhub *et al* (2000)		Fatas *et al* (2007)	
	Payoff	Total time	Payoff	Total time	Payoff	Total time	Payoff	Total time	Payoff	Total time
Min	0.00	8.35	0.00	8.07	0.00	12.37	0.00	5.45	0.00	3.28
50^{th}-percentile	1.85	26.91	20.14	22.73	8.38	25.41	3.12	14.55	3.43	9.75
95^{th}-percentile	9.98	62.98	46.44	54.70	25.01	65.62	11.42	47.02	16.36	25.53
Max	24.97	2728.22	101.16	4668.42	74.89	2273.38	27.20	14225.83	57.10	7483.52
(Obs. > 65 min)		(14)		(8)		(26)		(12)		(5)

Table 3: **Subject Characteristics and Treatment Assignments.** Standard errors are in parentheses. *Age* is in years old. *Female* is a dummy indicating a female. *Education* equals 1 if there is no answer, 2 if there is no qualification, 3 if vocational education, 4 if Bachelor degree, 5 if Master degree and 6 if Doctoral degree. *Finance ability* equals 1 if no answer is provided, 2 if a subject has never received financial training or been in finance courses, and 3 if a subject has ever received financial training or been in courses. *Income level* equals 1 if there is no response, 2 if the monthly household income is below €400, 3 if the income is between €400 and €800, and the value increases with the interval of €400 to 12 that indicates the income is more than €4,000. *Risk-taking* is the decision in the risk taking task at the end of the survey where the subjects chose how many percentage points (0-100) of their earnings they would like to put into a lotto. *Patience* is the decision at the end of the survey where the subjects decided how much they were willing to delay the payment to earn interest (no delay, 1 month, 2 months and 3 months).

	Koehler *et al* (2015)		Meissner (2016)		Blaufus & Milde (2021)			Anderhub *et al* (2000)		Fatas *et al* (2007)		
	Long first	Short first	Borrow first	Save first	Deferred	Immediate	Matching	Product	Summation	Annuity	Combined	Lump
Age	41.92	41.05	43.64	42.56	47.19	48.72	48.80	44.57	48.53	48.37	49.59	48.98
	(1.32)	(1.20)	(1.44)	(1.60)	(1.19)	(1.14)	(1.11)	(1.39)	(1.61)	(1.09)	(1.19)	(1.06)
Female	0.57	0.49	0.52	0.40	0.40	0.38	0.40	0.42	0.52	0.42	0.45	0.51
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Education	3.22	3.39	3.56	3.52	3.29	3.43	3.35	3.24	3.40	3.81	3.67	3.61
	(0.07)	(0.08)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Finance ability	2.17	2.24	2.16	2.24	2.15	2.23	2.19	2.20	2.23	2.23	2.27	2.26
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)
Income level	7.31	7.43	7.18	7.63	7.61	8.01	8.11	7.17	7.19	8.18	7.79	7.84
	(0.25)	(0.26)	(0.29)	(0.32)	(0.26)	(0.23)	(0.24)	(0.25)	(0.27)	(0.22)	(0.26)	(0.26)
Risk-taking	24.60	27.42	25.27	25.93	27.52	26.71	26.25	35.52	31.94	31.37	31.98	29.65
	(1.44)	(1.48)	(1.74)	(1.52)	(1.47)	(1.49)	(1.37)	(1.77)	(1.51)	(1.67)	(1.50)	(1.59)
Patience	2.66	2.85	2.82	2.62	2.69	2.74	2.81	2.64	2.51	2.69	2.75	2.40
	(0.10)	(0.10)	(0.11)	(0.12)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Observations	166	178	147	131	161	180	177	176	163	177	170	183

Table 4: **Effects of Individual Characteristics on the Efficiency of Decisions**. The dependent variables are the measurements of the efficiency of decisions: for Koehler, Langstaff, and Liu (2015), the dependent variable is the dummy indicating that there is no unspent money in the last period and no bankruptcy happened; for Meissner (2016), it is the dummy indicating that there is no overspending; for Blaufus and Milde (2021), it is the dummy indicating that the decisions follow the rule of thumb (s.d. of saving decisions is less than 2.5%); for Anderhub et al. (2000), it is the mean of deviations of the observed decisions from the optimal decisions. The results of the first three papers are marginal effects of Logistic estimation, and the results of the last paper are OLS estimation. The round/treatment control covariates include the round number and the treatment dummies. The number of observations equals the number of the decisions of all the rounds and all subjects in the first three papers and the number of subjects in the last paper. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	Koehler *et al* (2015)		Meissner (2016)		Blaufus & Milde (2021)		Anderhub *et al* (2000)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.006***	-0.006***	-0.001	-0.001	0.003**	0.003**	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.050**	-0.050**	-0.050***	-0.041**	-0.128***	-0.125***	0.102**	0.103**
	(0.023)	(0.023)	(0.017)	(0.017)	(0.039)	(0.039)	(0.047)	(0.047)
Education	0.019	0.018	0.004	0.005	-0.001	-0.003	0.046*	0.046*
	(0.012)	(0.012)	(0.009)	(0.009)	(0.023)	(0.023)	(0.026)	(0.026)
Finance Ability	0.044*	0.044*	0.035*	0.027	-0.114***	-0.117***	-0.052	-0.052
	(0.026)	(0.026)	(0.020)	(0.020)	(0.044)	(0.044)	(0.051)	(0.051)
Income	-0.003	-0.003	-0.005*	-0.005**	0.012*	0.012*	-0.012*	-0.012*
	(0.004)	(0.004)	(0.003)	(0.002)	(0.006)	(0.006)	(0.007)	(0.007)
Risk-taking	-0.000	-0.000	-0.002***	-0.002***	-0.004***	-0.004***	0.003***	0.003***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Patience	0.034***	0.034***	-0.011	-0.010	0.027*	0.027*	-0.026	-0.026
	(0.008)	(0.008)	(0.007)	(0.007)	(0.016)	(0.016)	(0.018)	(0.018)
Constant							1.164***	1.158***
							(0.145)	(0.149)
Round/treatment FE	No	Yes	No	Yes	No	Yes	No	Yes
N	1376	1376	1112	1112	972	972	339	339

Table 5: **Anderhub et al. (2000) Reproduction: Consumption and Resolution of the Survival Ambiguity.** Average fraction of initial wealth consumed in the first three periods, according to path of resolution of ambiguity on longevity risk. *1st* and *2nd period removal* are the card decks removed at each period, which eventually eliminates ambiguity of the actual survival probabilities subjects will face (the remaining deck color of red, green or yellow being then used to determine survival after periods 3, 4 and 5). Sequences are ranked in descending order of optimal consumption fraction (in parentheses).

Sequence Rank	1st period removal	2nd period removal	Product	Summation
1	¬green	¬yellow	0.67 (0.80)	0.72 (0.89)
2	¬yellow	¬green	0.69 (0.76)	0.69 (0.88)
3	¬green	¬red	0.70 (0.66)	0.71 (0.81)
4	¬red	¬green	0.72 (0.59)	0.69 (0.79)
5	¬yellow	¬red	0.69 (0.58)	0.71 (0.71)
6	¬red	¬yellow	0.73 (0.56)	0.72 (0.70)

Table 6: **Anderhub et al. (2000) Reproduction: Average initial consumption shares within the two groups.** *Round* is the sequence of round. *Group 1* includes the three sequences that have higher initial consumption share for each treatment, and *group 2* includes the three sequences that have lower initial consumption share for each treatment. Based on initial consumption of each treatment in Table 5, group 1 for treatment Product includes sequences 3, 4 and 6, and group 2 includes sequences 1, 2 and 5. Group 1 for treatment Summation includes sequences 1, 3 and 6, and group 2 includes sequences 2, 4 and 5.

Round	1	2	3	4	5	6
Product, group 1	0.70	0.71	0.71	0.73	0.72	0.71
Product, group 2	0.67	0.68	0.70	0.68	0.68	0.69
Summation, group 1	0.72	0.74	0.71	0.71	0.72	0.71
Summation, group 2	0.69	0.69	0.70	0.73	0.69	0.71

Table 7: **Anderhub et al. (2000) Reproduction: Facing an uncertain future.** *Cases* is the number of decisions, all of the decisions of all the subjects. $T \geq k$ (k=4, 5, 6) means that the subject reaches at least period k. x_k (k=4, 5, 6) is the consumption decision in period k.

Panel A: **Treatment Product**

	Cases	%		Cases	%		Cases	%
$T \geq 4$	729	100.0	$T \geq 5$	477	100.0	$T = 6$	264	100.0
$x_3 > x_4$	525	72.0	$x_3 > x_4 > x_5$	241	50.5	$x_3 > x_4 > x_5 > x_6$	94	35.6
$x_3 \geq x_4$	569	78.1	$x_3 \geq x_4 \geq x_5$	281	58.9	$x_3 \geq x_4 \geq x_5 \geq x_6$	126	47.7
$T \geq 5$	477	100.0	$T = 6$	264	100.0			
$x_4 > x_5$	333	69.8	$x_4 > x_5 > x_6$	125	47.3			
$x_4 \geq x_5$	368	77.1	$x_4 \geq x_5 \geq x_6$	157	59.5			
$T = 6$	264	100.0						
$x_5 > x_6$	180	68.2						
$x_5 \geq x_6$	209	79.2						

Panel B: **Treatment Summation**

	Cases	%		Cases	%		Cases	%
$T \geq 4$	655	100.0	$T \geq 5$	422	100.0	$T = 6$	258	100.0
$x_3 > x_4$	482	73.6	$x_3 > x_4 > x_5$	213	50.5	$x_3 > x_4 > x_5 > x_6$	91	35.3
$x_3 \geq x_4$	514	78.5	$x_3 \geq x_4 \geq x_5$	256	60.7	$x_3 \geq x_4 \geq x_5 \geq x_6$	132	51.2
$T \geq 5$	422	100.0	$T = 6$	258	100.0			
$x_4 > x_5$	289	68.5	$x_4 > x_5 > x_6$	112	43.4			
$x_4 \geq x_5$	328	77.7	$x_4 \geq x_5 \geq x_6$	155	60.1			
$T = 6$	258	100.0						
$x_5 > x_6$	161	62.4						
$x_5 \geq x_6$	197	76.4						

Table 8: **Fatas, Lacomba, and Lagos (2007) Reproduction: Timing of Retirement Treatment Effects.** The dependent variable is the mean retirement period chosen in the three lives. *Annuity* and *Combined* are dummies for subjects assigned to such treatment conditions; *Lump*-sum is the baseline. *Risk-taking* is the decision in the risk taking task at the end of the survey where the subjects chose how many percentage points (0-100) of their earnings they would like to put into a lotto. *Patience* is the decision at the end of the survey where the subjects decided how much they were willing to delay the payment to earn interest (no delay, 1 month, 2 months and 3 months). Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)
Risk-taking	0.028***		0.026***	0.026***	0.028***
	(0.005)		(0.005)	(0.005)	(0.005)
Patience		0.255***	0.184**	0.193**	
		(0.080)	(0.079)	(0.079)	
Annuity			-0.779***	-0.916***	-0.863***
			(0.215)	(0.247)	(0.247)
Combined				-0.281	-0.218
				(0.250)	(0.250)
Constant	5.136***	5.336***	4.980***	5.087***	5.481***
	(0.182)	(0.234)	(0.259)	(0.276)	(0.224)
(Annuity−Combined)				-0.635**	-0.645**
				(0.250)	(0.251)
R2	0.059	0.019	0.090	0.092	0.082
Prob. >F	0.000	0.002	0.000	0.000	0.000
Observations	530	530	530	530	530

Table 9: **Fatas, Lacomba, and Lagos (2007) Further Analysis: Effects of Experienced Termination Period on Later Decisions**. The dependent variable is the decision of retirement starting period in round 2 in columns (1-4) and the decision of retirement starting period in round 3 in columns (5-8). *End period round 1 (2)* is the termination period in round 1 (2). *Annuity* and *Combined* are dummies for subjects assigned to such treatment conditions; *Lump*-sum is the baseline. *If paid round 1 (2)* is a dummy indicating whether a subject gets a non-zero payoff in round 1 (2). *Risk-taking* is the decision in the risk taking task at the end of the survey where the subjects chose how many percentage points (0-100) of their earning they would like put into a lotto. *Patience* is the decision at the end of the survey where the subjects decided how much they were willing to delay the payment to earn interest (no delay, 1 month, 2 months, and 3 months). Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
End period round 1	0.060**	0.066**	0.060**	0.277***	0.077***	0.083***	0.078***	0.254***
	(0.029)	(0.029)	(0.028)	(0.037)	(0.030)	(0.029)	(0.029)	(0.037)
End period round 2					0.051*	0.050*	0.062**	0.191***
					(0.030)	(0.029)	(0.029)	(0.037)
Annuity		-0.733**	-0.832***	-0.788***		-1.033***	-1.139***	-0.966***
		(0.304)	(0.298)	(0.280)		(0.309)	(0.300)	(0.280)
Combined		0.177	0.047	-0.054		-0.362	-0.497	-0.545*
		(0.307)	(0.301)	(0.283)		(0.312)	(0.304)	(0.282)
If paid round 1				2.782***				2.290***
				(0.332)				(0.335)
If paid round 2								1.763***
								(0.340)
Risk-taking			0.027***	0.022***			0.030***	0.023***
			(0.006)	(0.006)			(0.006)	(0.006)
Patience			0.188*	0.152*			0.201**	0.118
			(0.096)	(0.090)			(0.097)	(0.090)
Constant	5.888***	6.029***	4.811***	2.329***	5.193***	5.614***	4.181***	0.666
	(0.264)	(0.310)	(0.390)	(0.471)	(0.362)	(0.401)	(0.471)	(0.578)
Observations	530	530	530	530	530	530	530	530

Table 10: **Koehler, Langstaff, and Liu (2015) Reproduction: Decision Constraints and Outcomes.** Treatment sequencing is the subjects' condition in first two rounds. *Lifetime income*, *lifetime expenses* and *retirement expenses* are environmental variables. *Lifetime spending* is the sum of all decisions in all periods. *Retirement savings* is the savings balance after the last period of the working phase. *Saving rates* are the fraction saved from income of discretionary budget at each period. *Bankruptcy rate* is fraction of subjects who did not save enough to cover mandatory expenses in retirement, and *Undersaving/deficit* is the sum of expenses that exceeds retirement savings in all retirement periods for this group of subjects. *Lost saving prevalence* is fraction of subjects who had unspent savings at the end, for whom *Lost savings* is savings left after last period. *Spending variability* is standard deviation of spending. *Difference savings to previous period* is average change in retirement savings from the previous round.

Condition	long	long	long	long	short	short	short	short
Round	1	2	3	4	1	2	3	4
Treatment sequencing	long first	long first	short first	short first	short first	short first	long first	long first
Lifetime income	1620	1620	1620	1620	1620	1620	1620	1620
Lifetime expenses	720	720	720	720	720	720	720	720
Lifetime spending	650	639	688	693	676	678	591	616
Retirement expenses	360	362	367	364	183	180	183	181
Retirement savings	912	969	952	949	609	608	742	707
Saving rate (from income)	0.527	0.561	0.548	0.546	0.369	0.368	0.444	0.424
Saving rate (from budget)	0.720	0.760	0.738	0.738	0.571	0.569	0.684	0.653
Bankruptcy prevalence	0.139	0.072	0.096	0.039	0.09	0.067	0.024	0.018
Undersaving / deficit	-96	-102	-43	-106	-54	-39	-99	-77
Lost savings prevalence	0.687	0.675	0.646	0.663	0.702	0.657	0.717	0.717
Lost savings	383	398	335	319	326	342	434	398
Spending variability	32.5	29.83	30.10	28.10	30.93	29.58	32.43	29.76
Difference savings to previous round		58	344	-3		0	-227	-34
N	166	166	178	178	178	178	166	166

Table 11: **Meissner (2016) Reproduction: Sub-Optimal Consumption Paths**. Deviations and absolute deviations from conditional optimal consumption, following the original study's *m1* and *m2*, respectively; and utility losses from deviations from unconditional optimal consumption *(m3)* at the subject X round level. *BF* and *SF* are Borrowing First and Saving First treatment conditions. *P-values* are calculated for Mann-Whitney-U tests of difference of means between both treatments. $N = 278$.

		round 1	round 2	round 3	round 4
median (m1)	BF	303.35	311.87	-107.55	-93.18
	SF	-120.10	-99.11	342.19	314.18
mean (m1)	BF	152.53	190.96	-95.24	-63.49
	SF	-135.78	-104.80	307.66	316.12
p-value		<0.001	<0.001	<0.001	<0.001
median (m2)	BF	387.54	389.32	235.53	240.37
	SF	192.50	200.64	372.84	334.04
mean (m2)	BF	514.69	520.90	362.06	348.55
	SF	323.59	327.94	444.65	405.65
p-value		<0.001	<0.001	<0.001	<0.001
median (m3)	BF	252.73	265.50	195.66	179.39
	SF	118.27	151.92	238.23	203.78
mean (m3)	BF	$>100,000$	$>100,000$	$>100,000$	$>100,000$
	SF	$>100,000$	$>100,000$	$>100,000$	$>100,000$
p-value		<0.001	<0.001	0.1164	0.9488

Table 12: **Blaufus and Milde (2021) Reproduction: Drivers of Saving Behavior.** The table presents regression results of random-effects models explaining subject's (effective) saving rates. The saving rate (SR) is defined as the saving amount in a given period divided by the income in this period. The effective saving rate (ESR) is the savings rates multiplied by $(1 - tax\ rate)$. *Deferred* is a dummy variable equal to one if the observation belongs to the deferred-tax treatment. *Matching* is a dummy variable equal to one if the observation belongs to the matching treatment. *Male* is a dummy variable equal to one if the subject is male. *High age* and *High risk-taking* is a dummy variable taking the value of one if the subject's answer to the underlying question is above the 75th percentile of all observation. *Period* is a time variable equal to decision period in each sequence (from 1 to 10). *Financial training* is a dummy variables taking the value of one if subjects state that they had participated in courses on financial decision making. Standard errors clustered at subject level are reported in parentheses. ***, **, and * indicate significance of the estimated coefficients at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	SR	SR	SR	ESR	ESR	SR	SR	SR	ESR	ESR
sequence	1	1	2	1	2	1	1	2	1	2
Deferred	0.098***	0.086***	0.095***	-0.090***	-0.086***					
	(0.026)	(0.024)	(0.025)	(0.019)	(0.019)					
Matching						0.057**	0.058***	0.078***	0.056**	0.076***
						(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Period	0.003***	0.003***	0.002**	0.003***	0.002*	0.003***	0.003***	0.002**	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High age		0.110***	0.039	0.082***	0.034		0.043*	0.032	0.043*	0.032
		(0.028)	(0.029)	(0.022)	(0.022)		(0.025)	(0.025)	(0.025)	(0.025)
Male		-0.053**	-0.048*	-0.041**	-0.034*		-0.013	-0.009	-0.013	-0.009
		(0.026)	(0.026)	(0.021)	(0.021)		(0.024)	(0.024)	(0.024)	(0.024)
Financial training		-0.017	0.012	-0.005	0.015		0.001	0.017	0.001	0.017
		(0.028)	(0.029)	(0.023)	(0.023)		(0.025)	(0.027)	(0.025)	(0.027)
High risk-taking		0.113***	0.096***	0.087***	0.074***		0.091***	0.099***	0.091***	0.099***
		(0.028)	(0.028)	(0.023)	(0.022)		(0.027)	(0.026)	(0.027)	(0.026)
Constant	0.332***	0.316***	0.336***	0.323***	0.338***	0.331***	0.307***	0.314***	0.307***	0.315***
	(0.016)	(0.025)	(0.025)	(0.022)	(0.022)	(0.016)	(0.023)	(0.023)	(0.023)	(0.023)
Observations	3,200	3,200	3,200	3,200	3,200	3,380	3,380	3,380	3,380	3,380
Subjects	320	320	320	320	320	338	338	338	338	338
R2	0.0411	0.135	0.0887	0.129	0.0891	0.0183	0.0568	0.0701	0.0561	0.0691

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch