Essays in Household Finance

Andre Lot

NHH - Norwegian School of Economics

Department of Finance

April 19, 2023

This PhD thesis was submitted to NHH – Norwegian School of Economics in partial fulfilment of the requirements for the Regulations of the philosophiae doctor (PhD) degree.

Preface

This thesis consists of three chapters studying topics related to individual retirement decision-making, within the umbrella of household finance. The most common framework used in research in this domain are life cycle models, enticing and flexible, as they accommodate many dimensions and elements of individual financial decision making in the context of a lifetime.

These decisions are complex. They have a long time span between decision-making and observable outcomes, which limited opportunities for decision-makers to learn from their own mistakes. In some instances, personally impactful decisions (such as choosing to convert pension savings, at the time of retirement, in annuities or withdraw them as a lump sum) are only made once, or at a yearly frequency. The institutional settings in OECD countries also impose many constraints on individual financial decision for retirement and old-age provisions, in the form of mandatory participation in different programs, such as state pension schemes or mandatory occupational pension funds.

In the first chapter, I study how a systematic bias on surival beliefs affects a stylized pension choice, concerning the timing of retirment, in an experimental setting. The main contribution of this chapter is to characterize different components of the previously defined longevity bias, as I decompose it into longevity misinformation (wrong assumptions about human longevity in general) and longevity pessimism (biased positional belief in which the average subject thinks he or she will live shorter than an average person of their same age and sex). I find that longevity pessimism is the largest component of the survival belief bias, and that longevity pessimism drives a choice for an earlier retirement payoff.

The second chapter attempts to replicate four experimental studies on retiremenet decision making, using online samples drawn from the broader population. The features of retirement decision making could raise questions about the validity of certain findings of the earlier experimental literature on this topic, to the extent that the results could be driven by the demographic characteristics of traditional student samples. We replicate most of the main findings of the original studies. Subjects choose to retire later when they earn payoffs as lump sum instead of annuites. Savings are higher under a matching contribution than under a tax rabate scheme. Subjects are debt-averse and make less efficient consumption decisions when they need to borrow from future income instead of saving from current income to smooth their consumption. We do not replicate the original finding that subjects make 'qualitatively correct' adjustments to their spending paths when ambiguity on survival risk is reduced.

In the final chapter, we examine the factors that determine ownership of tax-incentivized retirement savings accounts, and also the opening of this account among household that do not previously invest through them. Using a longitudinal panel of Swiss households, we find that several variables that determine cross-sectional ownership do not explain the decision of households to open that account.

Acknowledgments

Doing scientific research is often a lonely endeavor as we engage in our own thoughts and reasoning. Nonetheless, our interactions with fellow colleagues are a crucial part of this interesting business of *producing* and *disseminating* knowledge.

My journey that now yields this thesis is no exception. I would not have arrived at this point without the many contributions of friends, colleagues, mentors, supporting staff, and family, to whom I am deeply grateful for their intelectual generosity, support, friendship, engagement, feedback and for often going above and beyond their obligations and duties.

Declarations

All authors of the three research papers published in this thesis (Andre Lot, Kremena Bachmann, Xiaogeng Xu, Thorsten Hens, and Francisco Santos) declare no conflict of interest.

The research projects hereby presented have received generous financial support from the Swiss National Science Foundation (SNSF grant #18910).

Contents

Chapter 1

Longevity Pessimism, Misinformation, and Pension Choice

ANDRE LOT

To determine the value of a pension, individuals need to consider their survival risk. In this paper, I first elicit survival probabilities for a broad set of target ages, using a representative panel of the 18-70 year-old Swiss population. I document a systematic survival belief bias, which is the stylized fact that individuals underestimate their survival probabilities (compared to actuarial life tables). Then, I show that incorrect information about longevity in general is a substantial component of this bias. Next, I implement an incentivized experiment that requires subjects to make risky pension choices, in which payoffs are not affected by participants' own longevity. I find that longevity pessimism induces earlier and less risky choices about the timing of pension benefits, under annuity or lump-sum pension schemes. Finally, I show that happiness and satisfaction have an indirect effect on pension choices through the channel of longevity pessimism. [1]

JEL Classification: G51, C90, J26

Keywords: household finance, longevity risk, subjective beliefs, retirement, experiments

[1] I am very grateful for insightful comments received from Francisco Santos, Thorsten Hens, Niels Friewald, Natalia Gerasimova, Svein-Arne Persson, Roberto Riccò, Darya Yuferova, Diego Bonelli, Giovanna Apicella, Claude Fuet, Susan Thorp, Arjen Siegmann, Kremena Bachmann, Paul Karehnke, Samuel Hirshman, Enrico di Giorgio, Xiaogeng Xu, Thomas de Haan; and from audience members at VU Amsterdam, University of Galway, Univeristy Zürich, Université Laval, ESCP Business School, University of Sydney Business School, FAIR, Banco de España and Nordic Finance Network,.

1.1 Introduction

To determine the value of a pension, which only pays off if the pensioner is alive, individuals need to consider their longevity risk, which is their probabilities of not being alive at future pension pay-off dates. Individual longevity variance is large, and driven in part by one's own longevity risk factors (family history, medical diagnoses, endogenous risky behavior), about which subjects have private information (Perozek, 2008). The realization of all individual longevity risk factors for a whole population is precisely the longevity information that actuarial life tables contain.

However, when asked explicitly, individuals consistently report beliefs about their survival probabilities that are lower than unbiased expectations from life tables. In other words, the typical individual thinks that he or she will die sooner than an average person (of the same age and gender). This characterizes a systematic *survival belief bias*. Part of this bias incorporates incorrect assumptions that individuals have about longevity in general (not only about one's own individual survival), which represents *longevity misinformation*. If the longevity misinformation component is removed from the survival beliefs bias, what remains can be defined accordingly as *longevity pessimism*. The latter could in parts explain some household finance puzzles (Heimer, Myrseth, & Schoenle, 2019), such as the 'annuity puzzle'[2] (Yaari, 1965; Peijnenburg, Nijman, & Werker, 2016), the 'under-saving puzzle'[3] (Skinner & Hubbard, 1994), or the 'old-age precautionary savings puzzle'[4] (Lugilde, Bande, & Riveiro, 2019).

In this paper, using experimental methods, I first explore the determinants of survival belief bias. Next, in novel results, I show that longevity misinformation is itself a substantial component of survival belief biases. Then, I evaluate the impact of longevity pessimism (survival belief without its longevity misinformation component) on financial decisions about the timing of pension payoffs. This experimental decision resembles the trade-offs individuals face – in the field – when deciding whether to delay the start of retirement for a few years, in exchange for an increase in pension payoffs, as they would then spend a smaller fraction of their remaining life expectancy collecting pension benefits and a larger fraction making contributions instead. However, in my experimental

[2]"Why people do not buy annuities?"
[3]"Why people invest so little for retirement while earning labor income?"
[4]"Why people withdraw money too slowly from their investment accounts when they are very old?"

setting, the individual longevity risk of participants does not affect – by design – the risks to their payoffs. The results show that longevity pessimism leads to choices for earlier (and less risky) pension payoffs. Finally, I identify that personal happiness and satisfaction have an indirect impact on pension choice, through the channel of longevity pessimism, as happy subjects are less pessimistic about their longevity and choose later (riskier) pension payoffs.

To elicit survival beliefs, I employ an established procedure that asks subjects to assess their chances of being alive at different forward-looking target ages. I use a sample with a broad age range (18-70 years) of residents of Switzerland. Using sets of many survival beliefs for each individual allows the construction of fine-grained and smoother survival curves for each individual, extending the methodology of Dormont et al. (2018) and Wu, Stevens, and Thorp (2015). This also allows for more variation of subject age and thus of survival horizons whose probabilities subjects are asked about. My elicitation procedure contrasts with most studies on the longevity belief literature, which use coarse measures from retirement panels restricted to older subjects, usually eliciting survival beliefs only for nearer horizons (forward ages around 10 or 25 years ahead only).

Subjects consistently underestimate their survival probabilities at younger target ages. For example, the average woman (man) in the sample has an actual probability of living up to 70 years of age of 92.6% (88.2%) according to life tables,[5] but reports beliefs with a subjective probability of only 83.0% (82.3%). However, the seemingly small underestimation of survival probabilities until younger target ages (50 to 70 years) is critical. Because survival in any discrete period (one year) is conditional on having survived from birth until that period, underestimating survival probabilities to younger ages has a large impact on remaining life expectancy, as implied by those distorted probabilities.

In contrast, subjects vastly overestimate their survival to very old ages (beyond 90 years). Both women and men report average subjective probabilities of living up to 100 years of age of 14.2%, while actual unbiased probabilities from life tables are only 3.4% and 1.4%, respectively. Actuarial probabilities of someone living until age 70 are large, but the probabilities of someone living up to age 100 are small. If subjects were to make financial plans for retirement based on life expectancy implied by distorted survival probabilities, in the pattern described above, their savings

[5]This is the average probability considering the age distribution of subjects in the sample at the time of elicitation, not the probabilities at birth.

and investment outcomes would be impacted more by their underestimation of survival during the early phase of retirement than by their overestimation of survival chances until very advanced ages.

These findings contribute to the literature on subjective survival beliefs, adding to the body of evidence that the distortion of survival probabilities strongly depends on target ages. They also strengthen the methodological case that measures of survival beliefs bias that take only focal estimation of longevity (i.e., simply asking subjects until what age they think they will live) conceal strong underestimation of survival probabilities to younger target ages and overestimation to older target ages, which partially compensate each other over the lifetime.

In the following step, I elicit subjects' beliefs about the survival of an average person of their same age and gender. Survival beliefs about oneself incorporate private information subjects have about their own longevity risk factors, but these should not affect survival beliefs about strangers. However, the differences observed between both sets of beliefs (about oneself and the average person) are large. The women (men) in the sample assess that an average Swiss woman (man) has a survival probability of 82.2% (80.2%) up to age 70. In absolute terms, survival probabilities about the average person deviate only 0.8 (2.2) percentage compared to subjective survival probabilities about oneself, but deviate 10.4 (8.0) percentage points from unbiased probabilities (from life tables).

Individuals may have private information on their *own* longevity risk factors. Previous studies found that individuals recognize the impact of salient medical and health events on their own longevity (Bissonnette, Hurd, & Michaud, 2017; Bell, Comerford, & Douglas, 2020; Hurd & McGarry, 2002), but not necessarily of the background impact of their risky endogenous behavior such as smoking (Hurwitz & Sade, 2020). Individuals might even have distorted perceptions about the impact of these individual risk factors on their own life expectancy (Heimer, Myrseth, & Schoenle, 2019), or be generally pessimistic about any risk that affects them personally.

However, these mechanisms should not affect subjective beliefs about the survival of an average person. Therefore, I assign the systematic bias of underestimating the survival probabilities of strangers – the 'average person' – to longevity *misinformation* in a broad sense. This does not concern one's own survival and the longevity risk factors that affect the person individually, but rather the lack of knowledge, skewed perceptions, and/or distorted beliefs about everyone's longevity.

The characterization of longevity misinformation is a contribution of this paper to understanding the formation of individual survival beliefs. Longevity misinformation can be incorporated into any assessment that individuals make about their own survival relative to that of an average person. It may be an additional mechanism that drives heterogeneity in household financial decisions throughout the life cycle, complementing recent studies that analyze household responses to shocks in longevity risk factors (Kvaerner, 2022).

Individual survival belief biases for different target ages can be aggregated into a measure of longevity pessimism. It reflects one's overall attitude regarding his or her own survival with respect to life tables, after accounting for, and partially removing, the impact of longevity misinformation and private longevity risk factors, from the present until a given target age. The effects of longevity pessimism and private longevity information could attenuate each other[6] with respect to their impact on life expectancy. They are also difficult to disentangle from each other in empirical studies of field data. Analysis of financial decisions in the life-cycle (Browning & Crossley, 2001) in the field, with stochastic longevity (Groneck, Ludwig, & Zimper, 2016; Cocco & Gomes, 2012) is further complicated by the possible presence of bequest motives (Ameriks et al., 2011; Kvaerner, 2022; Inkmann, Lopes, & Michaelides, 2011).

Within compulsory-participation pension schemes (found in most OECD countries), neither longevity pessimism nor private information on longevity risk factors matters. When aggregated for large populations (in life tables), the average survival probabilities have little short-term variance.[7] This facilitates actuarial pricing of pensions, aimed at a representative individual of the population involved, while forcing everyone to pool and share their individual longevity risk. The existence of mandatory pension schemes further complicates the empirical analysis of the formation of longevity beliefs inferred from voluntary individual retirement investment decisions. In practice, for most individuals currently living in OECD countries, the vast majority of their retirement savings, investment, and subsequent drawdown is implemented through predetermined mandates

[6]For example, an individual who has a serious known medical condition likely to reduce her life span compared to an average person, yet overestimates her survival probabilities.

[7]In the long-term, a process known as *macrolongevity drift* becomes relevant. It concerns the epoch changes in expected immediate (one-year) survival probabilities for the same chronological ages. For example: a Swiss man of age 60 in 2022 is more likely to survive one additional year than a man of age 60 in 1975 because medical science is better equipped to treat certain diseases now, road safety has improved, and smoking rates have decreased.

prescribing highly regulated schemes.

To address some of these limitations, I investigate the role of longevity pessimism in an experimental task that involves simulated risky pension choices[8] (similar to Fatas, Lacomba, & Lagos, 2007). Subjects make choices about the timing of pension payoffs, for which they need to consider termination probabilities over multiple periods. In my experimental setup, longevity has no impact on the resolution of uncertainty (termination probabilities) for the participants' payoff in the task. Therefore, private information on longevity risk factors cannot improve subjects' assessment of their risk within the task.

My results show that the more pessimistic subjects are about their own longevity, the earlier (and less risky) their pension payoff choice is. Moreover, subjects delay their pension choice when the benefits are paid as lump sum, instead of a fair-priced annuity. Introducing a 'pessimistic annuity', priced as if the actuarial probabilities were weighted according to Tversky and Kahneman (1992), induces earlier pension choices than the fair-priced annuity, but the treatment effect is small.

Such findings offer two different contributions to the literature. I provide evidence that longevity pessimism is associated with the evaluation of risky financial choices on retirement, beyond considerations of whether subjects are informed about longevity in general or about their own individual longevity risk factors in particular. I also contribute to the literature on annuitization puzzles with further evidence that annuities attract less risk taking on pension choices than lump-sum payoffs and that longevity pessimism affects choices under both pension frameworks.

In further results, I also find that idiosyncratic happiness (Becker & Trautmann, 2022) can influence survival beliefs, as unhappy subjects may assume longevity-pessimistic beliefs. In the field, happiness is plausibly affected by many common drivers of longevity, such as health status or self-destructive behaviors, which further bolsters the case for the use of experimental elicitation that can reduce or remove these endogenous factors from affecting the termination risk in a simulated task.

Finally, I examine indirect effects of the individual happiness and satisfaction index on pension

[8]The experimental task is significantly non-contextual, in the sense that it refrains from using terms such as 'pension', 'retirement' or 'benefits' on its interface or instructions.

decision through longevity pessimism. I find that longevity pessimism confounds 52% of the total effect of happiness on pension decisions in a model that also accounts for the effects of change in health status.

Taken together, the results of this paper also have some empirical and policy implications. Because longevity misinformation comprises a significant part of survival belief bias, there may be potential to improve individual decision-making on financial decisions about retirement through better information or financial education of individuals making analogous decisions in the field. The age-dependent patterns of (over)underestimation of survival at younger (older) target ages and the effect of longevity pessimism on the timing choice of pension benefits suggest that removal of institutional constraints in the design of pension schemes should proceed with caution. The underlying mechanisms that make individuals longevity-pessimistic also affect decisions they make regarding risk-taking in pension payoffs, while, as noted, they are also substantially misinformed about longevity in general. In the field, this could result in the promotion of reforms to pension schemes that, inadvertently, exacerbate certain individual inefficient investment behaviors with respect to household welfare (under-saving for retirement by underestimating the financial needs in old age) or moral hazard for societal welfare programs (accelerated decumulation of retirement investments as individuals outlive their savings and subsequently rely on public assistance).

The remainder of the paper is organized as follows. In Section 1.2, I introduce the framework of pension decisions with longevity risk, define survival biases and longevity pessimism, and introduce the experimental setup. The main results are presented in Section 1.3, with additional analysis and robustness checks in Section 1.4. In Section 1.5 I discuss the results and conclude. This experiment was pre-registered with AsPredicted at Wharton Credibility Lab.[9]

1.2 Experimental Setup, Design and Data

In this section, I introduce the standard actuarial model for survival (Subsection 1.2.1), followed by survival beliefs measures, their biases and a model of longevity pessimism (Subsection 1.2.2). I then present the experimental design of the main pension choice task (Subsection 1.2.3), and

[9] AsPredicted #107473

briefly elaborate on the happiness and satisfaction index and its components (Subsection 1.2.4). Then, in Subsection 1.2.5, I explain in detail the experimental procedures that I adopt, and in Subsection 1.2.6 I discuss the simple univariate characteristics of my sample and present information on the recruitment, attrition, and general performance of the participants.

1.2.1 Longevity Beliefs and Biometric Returns

Using discrete measures, an average individual of current age a and sex[10] g has an expected probability $\zeta_{a,g,f}$ of dying between any current or future age $f_t \geqslant a$ and $f_t + 1$.[11] Then, the probabilities that an individual survives between his current age and any target[12] age $t > a$ (the cumulative survival probabilities) between a and t are:

$$\varphi_{a,g,t} = \prod_{f=a}^{t-1} \left(1 - \zeta_{a,g,f}\right) \tag{1.1}$$

The remaining life expectancy (the conditional expected lifespan from t onwards), in years,[13] of an individual of gender g from any target age $t > a$ onwards can be thus computed as:

$$e_{a,g,t} = \sum_{t}^{\bar{T}} \varphi_{a,g,t} \tag{1.2}$$

whereas $\bar{T}$ is the upper absolute limit of human longevity when $\zeta_{a,g,\bar{T}} = 1$, or, in other words, the maximum age a person of his or her gender can reach. The special case of the current remaining life expectancy (when $t = a$) is $e_{a,g} = \sum_{t=a}^{\bar{T}} \varphi_{a,g,t}$.

Let a *pension* be defined as a financial product whose cash flows are contingent on its individual holder being alive at each scheduled payoff date.[14] The present value of this pension must account

[10]The demographic 'life tables' that consolidate aggregate longevity expectations for large population groups are commonly segregated by sex, and usually do not account for non-binary groups (identifying themselves other than males or females) due to small group sizes and lack of historical data.

[11]This implies, for instance, that two women, of current ages 32 and 57, might have different expected one-year probabilities of dying at age 74 due to the process of macrolongevity drift.

[12]For clarity, I henceforth use *target age* to designate a set of future ages, expressed in chronological years (and not as offsets from current age), over which I analyze subjects' probabilities and respective beliefs.

[13]Assuming that each unit of f is also one year.

[14]For simplicity, I assume one payoff per evaluation period. Furthermore, for the purposes of all research questions in this study, it is not relevant whether pension payoffs are nominally fixed, unit-linked or inflation-indexed.

for the probabilities that its holder will not be alive to collect some (or all) of the future payoffs. Because these survival probabilities are always smaller than one – for any target age – the present value of the cash flows of a pension is lower than the present value of a series of zero-coupon bonds with the same maturities as the pension payoff schedule. The cumulative impact of longevity on the present value of a single pension cash flow, between the present and the target age t, can be expressed as total *biometric returns*:

$$\nu_{a,g,t} = \frac{1}{\varphi_{a,g,t}} - 1 \tag{1.3}$$

Furthermore, assuming a constant nominal interest rate r per period, the implicit one-period total return rate r^* for a pension payoff due at t, aggregating both the interest rate and the biometric returns, can be defined as:

$$r^*_{a,g,t} = \left(\underbrace{\left(1 + \nu_{a,g,t}\right)^{\frac{1}{t-a}}}_{\text{annual biometric return}} \times (1 + r) \right) - 1 \tag{1.4}$$

Equation (1.4) shows that the impact of biometric returns on total pension returns, for any given maturity, is considerably affected by the current age of different subjects. As an example, let us consider two Swiss men of current – as of 2021 – ages 44 and 54 years old, and a single pension cash flow with 25-year maturity. Their unbiased cumulative survival probabilities – from life tables – until target ages 69 and 79 (at maturity for each) are 87.9% and 70.8%, and their annualized biometric returns would be 0.52% and 1.39%, respectively. If, instead, the valuation of pension cash flows concerned two Swiss men 10 years younger (34 and 54) with a maturity of 10 years longer (35 years), their annual biometric return would be 0.39% and 1.04%, respectively.

The impact of biometric returns on pension valuation is most important for middle-aged individuals and pension maturities around the turn of the first decade of typical retirement. Then, the discount horizon is short enough not to dilute the total biometric returns when capitalized on annualized rates, making the biometric returns relatively more important with respect to interest rates r in terms of discounting pension cash flows. Simultaneously, for middle-aged individuals, the correspondent cumulative survival probabilities are still high enough that survival is more likely than death, for subjects to actually collect their pension payoffs.

In this study, the Swiss life table from the Swiss Federal Statistics Office (SFSO) for 2021, compiled by the Human Mortality Database (Max Planck Institute for Demographic Research, University of California, & French Institute for Demographic Studies, 2022), as parameters of the expected actuarial (unbiased) longevity and probabilities of survival and mortality.

Although life tables offer pretty accurate estimates of longevity of large groups representative of their populations, subjects hold individual beliefs on their own survival probabilities that are different from the actuarial expectations (Bissonnette, Hurd, & Michaud, 2017; Wu, Stevens, & Thorp, 2015). These differences can arise from private information about one's own longevity risk factors (such as family history or personal health status), from idiosyncratic over- or underestimation of longevity, from misinformation about the distribution of survival probabilities, and from personal biases on how the subject assess risky prospects in general.

Individual (subjective) longevity belief measures comprise subjective *survival beliefs* and *mortality beliefs*, measured as probabilities; and subjective *life expectancy*, measured in years. Payne et al. (2013) show that a 'live until' framing of longevity – which elicits survival probabilities – reduces inconsistencies on belief elicitation, compared to a 'die by' alternative, which yields mortality probabilities.[15]

To elicit survival beliefs, I extend the mechanism proposed by Wu, Stevens, and Thorp (2015) to incorporate a wider span of chronological age of subjects (18-70 years old), and elicit more precise measurement of survival beliefs (on a scale with 99 discrete points) in order to build subjective survival curves less affected by coarse measurements of individual beliefs. Each subject i of current age a_i is asked "*What are your chances of being alive at age...*" as the prompt to input survival probabilities $\tilde{\varphi}_{i,t}$ for a set $\mathbf{F}$ of target ages that span five-year intervals:

$$\mathbf{F}_i = \{t_n \in (t_{n=1} = 50, t_{n-1} + 5, \ldots, 105) \,|\, t_n > a_i\} \tag{1.5}$$

Subjects younger than 50 input estimated survival probabilities for 12 target ages. Those older than 50 are elicited on fewer target ages, starting with the first target age that is higher than their current

[15]This result is consistent is the premise that eliciting the less salient state – surviving another number of years – is less likely to attract probability distortions on reported beliefs than eliciting the salient event – dying.

age. The set of target ages is fixed with respect to specific chronological ages (50, 55, $\cdots$), instead of offsets from current age ($age+5, age+10, \cdots$) as in some previous studies. This means that stepwise implicit probabilities between target ages after the first are directly comparable between subjects. If all subjects are specifically asked about the probabilities of living from their current ages up to 80 and 85 years, it is trivial to calculate the implicit survival probabilities between ages 80 and 85 as $\frac{\tilde{\varphi}_{i,85}}{\tilde{\varphi}_{i,80}}$. This procedure also avoids heterogeneous elicitation sets where some subjects are asked beliefs about salient ages (e.g. 60, 65, 70 years) and others are not (e.g. 57, 62, 67, 72 years).

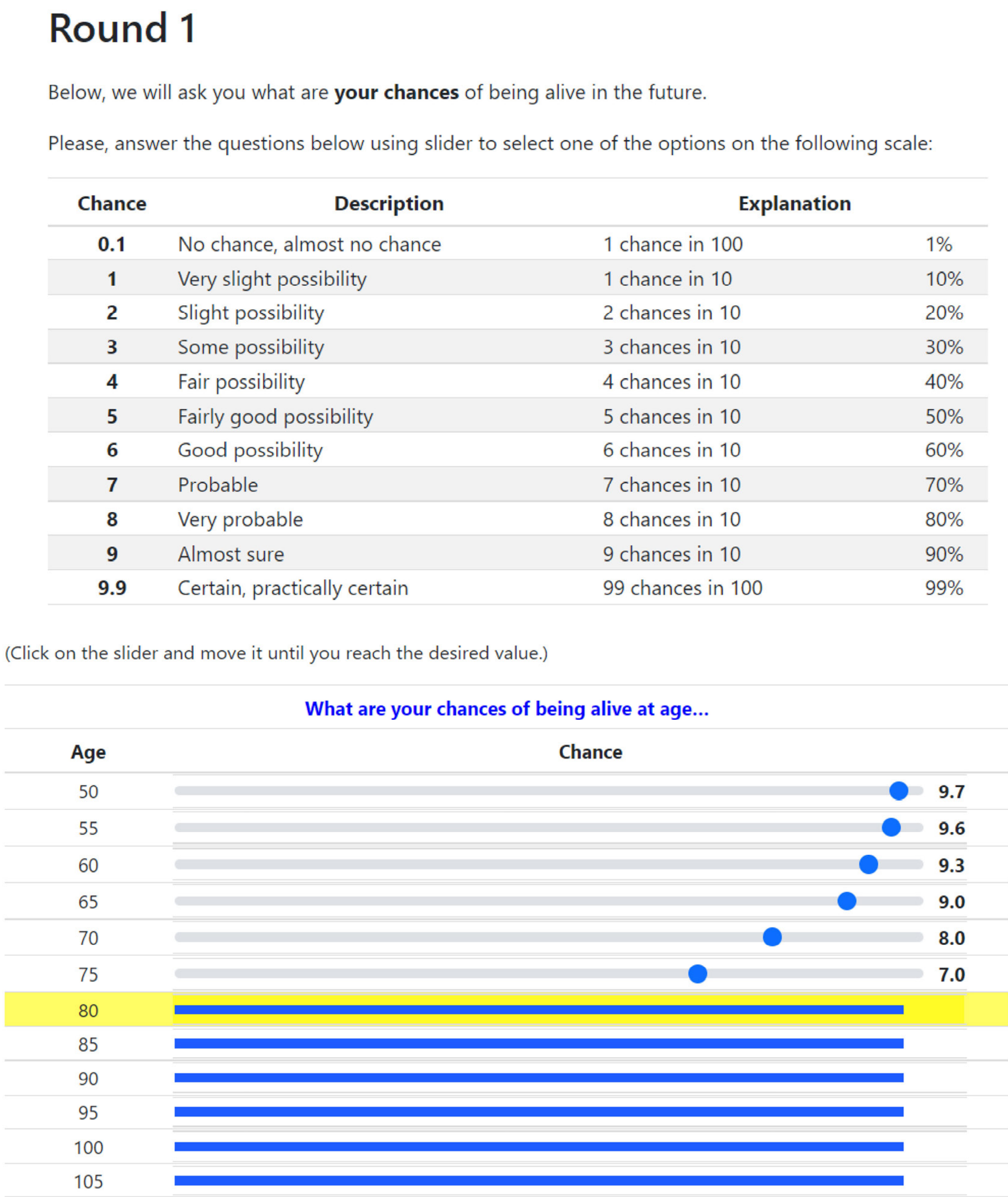
Survival probabilities are elicited on a 0.1-9.9 scale with 0.1 discrete increments. An information table explaining the scale is available on the same screen as subjects input their beliefs. Subjects choose the probabilities, using an interactive slider, for each target age, without defaults or preset values. In this way, this study uses a finer discrete scale (as in Dormont et al., 2018), instead of the usual coarse target age vectors from most previous studies. This reduces the potential impact of truncation and partial identification of probabilities (Bissonnette & de Bresser, 2018; Imbens & Manski, 2004; Kleinjans & Soest, 2014; de Bresser, 2019), in particular at younger target ages. Nonetheless, subjects might still input survival probabilities that are within the rounding interval to their actuarial expectations, when $(\varphi_{a_i,g_i,t}-0.005) \leqslant \tilde{\varphi}_{i,t} < (\varphi_{a_i,g_i,t}+0.005)$. In such cases, the input of the survival belief is replaced by the actual probability from the life table. Figure 1.1 shows a screenshot of the English-translated online elicitation interface.

Although most of the previous literature on survival belief bias considers only subjective beliefs about oneself $(\mathbf{F}_i^{own})$ vis-a-vis their actuarial expectations from life tables, I also elicit two additional different sets of beliefs with different subject or object, for a total of three sets of probabilities $\mathbf{F}_i^j$ per subject, as explained below.

In the first additional set, subjects input their survival beliefs about an average person of the same age and gender[16] $(\mathbf{F}_i^{pop})$. Deviations between this measure and those from life tables indicate misinformation about longevity risk in general, regardless of its source. Any private information that subjects may possess about their own longevity risk factors should not affect their assessment of the survival of an average person.[17] On average, these survival probability estimates for an

[16]For example, a prompt reads '*'What are the chances of a typical 23 years old Swiss woman still being alive at age ...*"

[17]Eliciting probabilities of an archetype of same age and gender reduces the cognitive burden on subjects and,



Chance	Description	Explanation	
0.1	No chance, almost no chance	1 chance in 100	1%
1	Very slight possibility	1 chance in 10	10%
2	Slight possibility	2 chances in 10	20%
3	Some possibility	3 chances in 10	30%
4	Fair possibility	4 chances in 10	40%
5	Fairly good possibility	5 chances in 10	50%
6	Good possibility	6 chances in 10	60%
7	Probable	7 chances in 10	70%
8	Very probable	8 chances in 10	80%
9	Almost sure	9 chances in 10	90%
9.9	Certain, practically certain	99 chances in 100	99%

Figure 1.1 – Interface for elicitation of longevity beliefs. Screenshots of the (translated) online interface used to elicit subjective longevity beliefs. An yellow highlight hovers with the mouse, and subjects determine the starting point the slider for any target age by clicking anywhere on any blue bar.

average person should match the parameters of the life table.

The final set of beliefs concerns the survival beliefs of a subject's family and close friends about the subject's survival $\left(\mathbf{F}_i^{fam}\right)$, according to the subjects' expectations of them. Family and friends might be partially informed about the subject's longevity risk factors, such as family longevity history

arguably, limits the potential impacts of any secondary bias from gender and/or age differences when subjects assess *relative* survival probabilities of other people.

(how old did deceased relatives live or presence of hereditary diseases), endogenous risk behavior (whether the subject smokes) or health status (medical diagnoses well known to close associates of the subject). These beliefs provide a useful double-comparison reference point with respect to both the subject's survival beliefs about oneself and about an average person.

For simplicity, the sets of survival beliefs are hereafter simply referred to as *oneself (own)*, *average person (pop)* and *family (fam)*, respectively.

In the last step of elicitation of survival beliefs, subjects also provide a single focal point estimate of life expectancy, that is, their estimated age at death ("*To what age do you think you will live (in years)?")*, for the three sets of beliefs. This simpler elicitiation mechanism provides an alternative measure of longevity to be compared with life tables. Such focal subjective estimations are nonetheless unstable, as subjects tend to cluster their estimations around 'round' and 'salient' numbers, generating beliefs clustered at these salient ages. For this reason, when using life expectancy estimations, I take the implied values from survival beliefs instead.

1.2.2 Longevity Belief Bias Measurement

Taking the three sets of beliefs elicited on survival probabilities for each subject, I first calculate the survival biases between each set of beliefs, from their counterfactual probabilities of the Swiss life tables. As the probabilities are numerically bounded within $[0.01 - 0.99]$, while their actuarial expectations (from life tables) also vary substantially between the target ages, it is necessary to scale the deviations between subjective and actuarial parameters. Using the correspondent mortality probabilities from the beliefs $j = \{own, pop, fam\}$ elicited for each subject and target age, the survival belief scaling factors $\iota_{i,t}$ are obtained:

$$\iota_{i,t}^{j} = \frac{1 - \tilde{\varphi}_{i,t}^{j}}{1 - \varphi_{a_i,g_i,t}} \tag{1.6}$$

Each of these factors is the ratio of the implicit subjective mortality probability to the unbiased expectation from the life table. Hence, target ages other than the last (105 years) have elicited overlapping beliefs. For example, a woman of current age 45 can only survive up to age 70 if she first survives until 65 years old. Then, her belief in her survival probability up to age 70 $\left(\tilde{\varphi}_{i,70}^{own}\right)$ also contains expectation about her survival up to age 65 $\left(\tilde{\varphi}_{i,65}^{own}\right)$.

To obtain a comparable survival bias measure for each set and subject, accumulated until each target age, it is necessary to aggregate the implicit biases for each subject, to the extent that the scaling factors $\iota_{i,t}^{j}$ are not constant across target ages t within subject i.

Of particular interest is the fact that survival probabilities across $\mathbf{F}_i$ range from very high (in younger t) to very low (in very old t), while also including t for which $\varphi_{a,g,t}$ is moderate between both tails. The scaling factors of survival belief $\iota_{i,t}^{j}$ for younger target ages are sensitive to small absolute deviations, as $\left(1-\varphi_{a_i,g_i,t}\right)$ is small, but propagate over a long remaining life span, greatly impacting subjective remaining life expectancy. On the other hand, biased beliefs for very old target ages have only limited effects on remaining life expectancy, because subjects are unlikely to survive – for instance – up to age 100 anyhow. In addition, a comparable individual bias measure must account for the fact that subjects older than 50 (the first target age) have a variable number of target ages in their belief sets, and that the scaling factors are also sensitive to the subject's current age.[18]

Therefore, using these survival belief scaling factors, I calculate, for each subject, belief set $j = \{own, pop, fam\}$ and target age t, the natural logarithm of the average scaled survival belief factor for the target ages up to t, weighted by unbiased actuarial probabilities, and define the three corresponding individual survival belief bias measures:

$$q_{i,t_n}^{j} = \ln\left[\frac{\sum_{t}^{t_n} \iota_{i,t}^{j} \times \left(1-\varphi_{a_i,g_i,t}\right)}{\sum_{t}^{t_n} 1-\varphi_{a_i,g_i,t}}\right] \tag{1.7}$$

Subjects with $q_i^j > 0$ are *pessimistic* about the survival belief j (oneself, average person or family), with respect to actuarial unbiased probabilities. Likewise, $q_i^j < 0$ indicates survival *optimism* at the individual level. Differences of q^j between subjects indicate their relative ratios of pessimism or optimism.

The measure q_i^{own} is analogous to the most common *longevity bias* as defined and analyzed by

[18]For example, a female subject of current age 48 has higher probability or surviving up to age 73 than a woman with current age 20, because the risk of that this individual dies between ages 20 and 48 (irrelevant for the older individual who already reached that age) is embedded in the cumulative survival probabilities between ages 20 and 73.

the existing literature, comparing the beliefs of subjects about their own longevity to the survival probabilities of life tables. As mentioned previously, this *oneself* bias comprises private information on factors that affect longevity and potentially a term that incorporates pessimism and optimism of the subject about his or her own longevity.

Alternatively, q_i^{pop} cannot incorporate any private information on subjects' own longevity risk factors. It measures the bias between a subject's survival belief of an average person of the same age and gender and the survival probability for this average person from the life tables. Consequently, the measure q_i^{pop} characterizes *longevity misinformation*, or incorrect assumptions that subjects have about longevity in a broad sense, not only about their own survival.

Following, q_i^{fam} can be assumed to embed partial information on one's own longevity risk factors, as previously discussed. If a subject believes that family and friends, who know the subject well, are as pessimistic as him or herself and assume that they have partial information on negative longevity factors about the subject, then it could be expected that $q_i^{own} > q_i^{fam} > q_i^{pop}$. Otherwise, q_i^{fam} will also incorporate differences on the expected pessimism of family and friends and the subject's pessimism about his or her survival.

Individual survival beliefs about oneself can also be scaled with respect to the subject's family and average person beliefs, allowing for comparison of these relative biases between subjects. For that purpose, I define the two additional survival bias measures, average-weighted by the actuarial unbiased probabilities as in Equation (1.7):

$$q_{i,t_n}^{own:pop} = \ln \left[\frac{\sum_t^{t_n} \frac{1-\tilde{\varphi}_{i,t}^{own}}{1-\tilde{\varphi}_{i,t}^{pop}} \times \left(1-\varphi_{a_i,g_i,t}\right)}{\sum_t^{t_n} 1-\varphi_{a_i,g_i,t}} \right] \tag{1.8a}$$

$$q_{i,t_n}^{own:fam} = \ln \left[\frac{\sum_t^{t_n} \frac{1-\tilde{\varphi}_{i,t}^{own}}{1-\tilde{\varphi}_{i,t}^{fam}} \times \left(1-\varphi_{a_i,g_i,t}\right)}{\sum_t^{t_n} 1-\varphi_{a_i,g_i,t}} \right] \tag{1.8b}$$

To quantify longevity *pessimism* as a comparable measure across subjects, regardless of their current age, I first regress the relative bias of oneself to family beliefs on the relative bias of oneself to the average person, longevity misinformation, age, target age, and gender, as follows:

$$q_{i,t}^{own:fam} = \alpha + \beta_1 q_{i,t}^{own:pop} + \beta_2 q_{i,t}^{pop} + \gamma_1 a_i + \gamma_2 t + \gamma_3 g_i + \mu_i + \varepsilon_{i,t} \tag{1.9}$$

and then use its predicted values for each subject and target age as the measure of longevity pessimism $\psi_{i,t} = \widehat{q_{i,t}^{own:fam}}$.

Finally, bias on survival beliefs could also be measured in terms of differences in implied partial life expectancy between the current age and each target age. Partial life expectancy $\widetilde{ex}_{i,t}^{j}$ is how many years the subject is expected to live from the present up to a given target age. Because individuals always have survival probabilities smaller than one between the present and any target age, in expectation they will accumulate fewer years lived between a_i and t than $t - a_i$.

From the life tables, the partial unbiased life expectancy $\overline{ex}_{a_i,g_i,t}$ is extracted from the probability mass function of individual survival. From the elicited beliefs on survival probabilities, the expected partial life expectancy $\widetilde{ex}_{i,t}^{j}$, for each subject, until any target age, for the belief sets $j = \{own, pop, fam\}$, is given by:

$$\widetilde{ex}_{i,t_n}^{j} = \begin{cases} \tilde{\varphi}_{i,t}^{j} \times (t - a_i) & \text{if } n = 1 \\ \widetilde{ex}_{i,t_{n-1}}^{j} + 5\tilde{\varphi}_{i_i,t}^{j} & \text{if } n > 1 \end{cases} \tag{1.10}$$

Then, I take three life expectancy bias measures $kex_{i,t}^{j}$ as the simple numerical difference between partial life expectancy (implicit from the probabilities elicited for $j = \{own, pop, fam\}$ and the unbiased parameter from the life table, as:

$$kex_{i,t}^{j} = \widetilde{ex}_{i,t}^{j} - \overline{ex}_{a_i,g_i,t} \tag{1.11}$$

Then, I also calculate the relative life expectancy bias measures, analogous to those of equations

(1.8a) and (1.8b):

$$kex_{i,t}^{own:fam} = \widetilde{ex}_{i,t}^{own} - \widetilde{ex}_{i,t}^{fam} \tag{1.12a}$$

$$kex_{i,t}^{own:pop} = \widetilde{ex}_{i,t}^{own} - \widetilde{ex}_{i,t}^{pop} \tag{1.12b}$$

These life expectancy bias measures will be used for robustness checks. Like the $q_{i,t}^{j}$, $q_{i,t}^{own:pop}$ and $q_{i,t}^{own:fam}$ survival belif biases, life expectancy bias measures aggregate, at the individual level, different survival beliefs relative to benchmarks (from life table or a different set of beliefs between oneself, average person, and family). Contrary to the former, nonetheless, life expectancy bias does not weight distortion on beliefs that are measured, implicitly, serveral times for future target ages that have partially overlapping chronological spans – as previously noted.

1.2.3 Pension Payoff Choice

The main decision-making task of the experiment is the choice of a period (1-20) of an experimental life (round) to collect, start to collect or start paying pension payments. After answering questions on their longevity beliefs as described, subjects face experimental risk (termination probabilities), and there is no interest rate. Subjects make one choice per round, at its start. This task expands the design and treatment conditions used by Fatas, Lacomba, and Lagos (2007).

The termination probabilities are given by a random draw without replacement of virtual cards. Subjects start a round with a deck consisting of 19 green cards and one red card. At each period, a card is drawn: if the red card is selected, the round is terminated immediately; otherwise, the subject advances to the next period. This mechanism implies that a round cannot go past 20 periods (when the only remaining card would be the red one), the average experimental longevity is 10.5 periods (at the start of a round), the distribution of termination periods for subjects in a round is uniform, the one-period termination probabilities increase at each period (a process that mirrors the longevity dynamic of senescence),[19] and the marginal increase in termination probabilities across periods is monotonically positive.

[19]At older ages, as a person gets one year older, his or her probability of surviving another 12 months decrease.

There are four conditions on the treatment of the payoff structure. Their environmental parameters are shown in Table 1.1. In the baseline condition *Fair*, subjects decide when (period) to start collecting payoffs (in points). They keep collecting fixed payoffs every period until termination (i.e., until they draw the red card). For example, a subject that chooses period 11 will earn zero if terminated before period 11. Otherwise, the subject earns 303 points per period until a red card is drawn.

This structure resembles a life annuity pension, whose nominal payoff per period increases the longer the subject postpones the beginning of retirement (pension choice for later payoffs). As cumulative survival probabilities decrease in later periods, biometric returns increase substantially, so nominal payoffs become quite high, although the subject is not very likely to reach such periods before termination.

At the start of a round, nevertheless, the expected value of the option for any period is 1000 points. Therefore, subjects are making a risky choice with the same underlying expected value on 20 different prospects whose conditional termination risk realizations at their overlapping end-tails are identical.

The *Pessimistic* condition has annuity payoff mechanics identical to those of *Fair*, but with distorted payoff values. I recalculate the actuarial probabilities as if subjects were engaged in probability weighting according to the probability weighting function of Tversky and Kahneman (1992). I use the average standard probability weighting coefficient for Swiss survey participants from Rieger, Wang, and Hens (2017). The weighted experimental cumulative survival probabilities s for each period p then become:

$$w\left[s_p\right] = \frac{\left(s_p\right)^{0.54}}{\left({s_p}^{0.54} + \left(1 - s_p\right)^{0.54}\right)^{\frac{1}{0.54}}} \tag{1.13}$$

which are used to define the expected payoffs for this treatment condition. As seen in Table 1.1, the payoffs are higher than in the *Fair* condition until pension choice in period 10, and lower afterward. The expected values (discounted by unbiased probabilities) are now different between periods, being the highest at period 2 (1257 points), the lowest at period 20 (169 points), and higher than 1100 points (a 10% increase from the other treatments) for all periods 1 to 8.

Table 1.1 – Pension Decision Task Parameters. The table summarizes the experimental parameters for the four treatment conditions. *Cumulative survival* is the probability that a subject survives until the period. *Termination* is the probability that the experimental life ends at the period, conditional on surviving until the period. *Biometric returns* are the compound implicit one-period biometric returns from the start of the round to the period. In the *Fair* and *Pessimistic* treatment conditions, payoffs are the fixed amount of points subjects get, per period, starting at the chosen period, until termination. In *Lump-sum*, subjects get a single payoff in the chosen period. In *Reverse*, subjects receive an endowment of 2000 points at the start of a round and pay the specific amount from the chosen period until termination.

Period	Probabilities		Biometric ret.	Payoffs (points)			
	cml. survival	termination		Fair	Pessimistic	Lump-sum	Reverse
1	1.000	0.050	0.00 %	95	117	1 000	-95
2	0.950	0.053	2.60 %	105	132	1 053	-105
3	0.900	0.056	3.57 %	117	146	1 111	-117
4	0.850	0.059	4.15 %	131	161	1 176	-131
5	0.800	0.063	4.56 %	147	178	1 250	-147
6	0.750	0.067	4.91 %	167	197	1 333	-167
7	0.700	0.071	5.23 %	190	219	1 429	-190
8	0.650	0.077	5.53 %	220	245	1 538	-220
9	0.600	0.083	5.84 %	256	275	1 667	-256
10	0.550	0.091	6.16 %	303	311	1 818	-303
11	0.500	0.100	6.50 %	364	356	2 000	-364
12	0.450	0.111	6.88 %	444	411	2 222	-444
13	0.400	0.125	7.30 %	556	483	2 500	-556
14	0.350	0.143	7.79 %	714	578	2 857	-714
15	0.300	0.167	8.36 %	952	710	3 333	-952
16	0.250	0.200	9.05 %	1 333	903	4 000	-1 333
17	0.200	0.250	9.93 %	2 000	1 207	5 000	-2 000
18	0.150	0.333	11.12 %	3 333	1 746	6 667	-3 333
19	0.100	0.500	12.88 %	6 667	2 909	10 000	-6 667
20	0.050	1.000	16.16 %	20 000	3 377	20 000	-20 000

If subjects are underweighting their high termination probabilities in the first periods, this modified set of payoffs should attract, on average, earlier pension choice. As well, on this condition the payoffs for low-probability very late periods are also substantially reduced (3377 points in period 20, instead of 20000 in the *Fair* condition).

In the *Lump-sum* condition, subjects earn a single payoff at their chosen period, as long as they have not been terminated before. Further realization of experimental survival after that period is irrelevant to his or her payoff in that round. Concentrated pension payoffs in lump sums can lead to a delay in pension choice (Fatas, Lacomba, & Lagos, 2007), as the cognitive burden of integrating a stream of uncertain payoffs is reduced. Furthermore, the salience of a large amount paid could attract subjects to take more risk when the realization is not contingent on the aggregation of present values that include later periods when survival probabilities are low. The expected values of the payoff of these conditions are identical to those of the *Fair* condition, that is, 1000 points for the choice of any period.

Finally, in the *Reverse* condition, the the subjects are given an initial endowment at the beginning of each round. They then need to make a stream of payments out of that endowment from their chosen period until termination, as if they were the issuers (instead of holders) of a life-annuity pension. The endowment (2000 points) is equal to twice the expected value of the payments, so the expected payoff value in all periods is the same as in *Fair* and *Lump-sum* (1000 points, after the expected payment of 1000 points from the endowment is made). In this condition, a subject becomes bankrupt (earning no variable payoff in that round) if the total payments he/she needs to make exceed the initial endowment. Bankruptcy is possible for all pension choices, except for period 1, if termination occurs too late. For instance, a subject whose reverse pension choice is period 10, with eventual termination at period 18, will have made 9 payments of 303 points each: a total of 2727 points that exceeds the initial endowment by 727. Only in a choice for period 1 or 2 would prevent bankruptcy in all possible cases (a subject that survives until the last period will have paid in total 1900 and 1995 points if he or she made a choice for period 1 or 2, respectively).

To the extent that subjects are loss-averse and treat payments out of their endowment as losses, but do not distort the implicit probabilities, they should on average make earlier pension decisions than in other treatments. Biometric returns, similar to those under *Fair* condition, should be less effective in inducing choices in later periods. Subjects can earn a maximum of 2000 points in *Reverse*, which is equivalent to the maximum payoff of the pension choice in period 1 for the condition *Fair*.

The termination probabilities in the task are completely unrelated to the subject's own longevity risk factors, as uncertainty on the payoffs of the pension choice task is resolved within a short experimental session. Even if the impact of these factors is weighted and distorted in terms of their probabilities (Heimer, Myrseth, & Schoenle, 2019), there should be no significant impact on pension choices in this task. If, however, subjects are longevity pessimistic for reasons unrelated to their expected information on longevity risk factors, and not entirely due to wrong information on longevity in general, then longevity pessimism could affect their pension choices in the task.

The treatment conditions on payoff structure present the same underlying decision problem: assessing cumulative survival probabilities in a risky prospect, and deciding a period for payoffs structured according to each treatment. This decision is analogous to the individual deciding whether to postpone or anticipate the start of retirement or the schedule of voluntary annuities. The conditions *Fair* and *Pessimistic* conditions require a subjective assessment of experimental survival probabilities for the maximum possible duration of a round (20 periods), since the expected payoffs lasts until the subject faces termination. The *Reverse* condition inverts the gain frame from accruing payoffs over multiple periods to a loss frame of spending down (possibly going bankrupt) from an endowment that is already the maximum possible payoff a subject can attain. The *Lump-sum* condition offers simple independent prospects in each period, which require a simpler assessment of survival probabilities only until the chosen period.

Importantly, this study is not primarily concerned with the treatment effects of each of these conditions. Instead, it focuses on whether the effect of longevity pessimism in pension decisions is robust to different payoff structures, that resemble different underlying optimization problems faced by individuals making voluntary pension decisions in the field.

1.2.4 Happiness and Satisfaction

Happiness, broadly defined (Frey & Stutzer, 2002), is correlated with several factors that drive longevity. It has an U-shaped pattern (Becker & Trautmann, 2022): higher at young and old age, lowest in middle age and could drive subjective longevity beliefs (Gimenez, Gil-Lacruz, & Gil-Lacruz, 2021). In summation, happiness can be a determinant of both longevity beliefs, while also being correlated with individual preferences that influence choice under risk, as in the pension choice task.

I ask the subjects five questions on happiness and life satisfaction, combining questions from the European Values and Satisfaction Survey by Sortheix and Lönnqvist (2014), and relevant theme questions from the Swiss Household Panel (FORS - Swiss Centre of Expertise in the Social Sciences, 2022). The questions concern overall happiness; and satisfaction with current life, personal life history, finances, and health.[20] All questions are measured on a scale of 1-10, at 0.1 intervals, and input with a slider with no default value.

As the expected correlation with these measures is relatively high, I use the first factor in a principal component analysis of the answers to these five questions (all on the same scale) as a *health and satisfaction index*. This reduced index is then used to evaluate how happiness and satisfaction, in a broader sense, could affect pension decisions, directly or indirectly, through the channel of longevity pessimism.

1.2.5 Experimental Procedure

The experiment consists of several self-contained individual tasks that elicit survival beliefs, the choice of pension payoffs, and a few sets of individual characteristics. Figure 1.2 illustrates the sequence of tasks in the experiment.

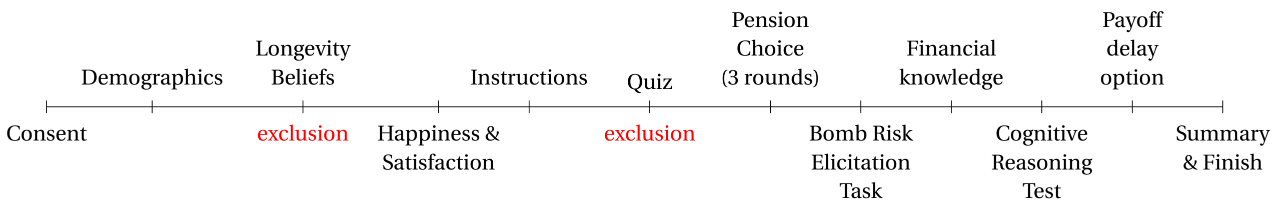


Figure 1.2 – Timeline of Experimental Session. Subjects are dropped, during the experiment, if they repeatedly violate the monotonicity of elicited cumulative survival beliefs, or fail a simple quiz on the mechanics of the main task.

After giving consent and providing demographic information, survival probabilities are elicited for each set of beliefs j (oneself, average person, and family) are elicited on single screens, which are identical except for changes in text that identify the relevant set.

The monotonicity of the beliefs in each $\mathbf{F}_i^j$ is enforced. Otherwise, a participant who informs a higher survival probability for an older target age than a younger one would imply survival probabilities greater than one between those ages. A participant receives an error message and a practical example at the first violation, and is excluded after a second.

An additional single screen asks participants to give focal (years) life expectancy estimations, in terms of age at death, for all sets.

[20]A separate question inquires subjects on recent changes to their health.

The participants then answered questions about happiness and satisfaction and read the relevant instruction screens for their treatment. Instructions remain available throughout the rest of the main pension decision task, as clickable tabs at the bottom of each screen.

Before proceeding to the main decision, participants must pass a quiz of four questions on the basic mechanics and rules of the pension decision tasks. The instructions remain available for consultation during the quiz. Participants who do not correctly answer all quiz questions after two attempts are excluded from the experiment.

After passing the quiz, participants complete three rounds of the pension decision task. All the outcomes and randomization of the termination period in each round are independent of each other. In another study, a similar pension decision task with repeated rounds showed significant learning effects, especially from participants who are terminated before their chosen pension period (Bachmann et al., 2022). For this reason, this experiment does not have a standard trial round, relying instead on the quiz and subsequent exclusion criteria to ensure that participants know how the task works.

At the beginning of a round, participants make their pension decision, selecting one period out of 20 from a slider that, when moved, automatically adjusts feedback information on expected payoffs conditional on outcomes of the draw of red and green cards. Then, on the same screen, participants navigate through the draw period by period, until termination. A period-by-period recursive table is populated with payoffs accumulated in that round, if any. Once the round reaches termination (red card is drawn), a brief intermission screen is shown and participants move on to the next round.

One of the three rounds is selected for compensation, which participants do not know until the very last step of the session. However, they will know the payoffs of each round, based on their termination and pension decision. Therefore, the results of the next decisions could be potentially affected by the expected payoffs from the main task.

Then, participants complete the *"bomb" risk elicitation task* (BRET) – designed by Crosetto and Filippin (2013) – as implemented by Holzmeister and Pfurtscheller (2016), using a 8×8 matrix setup in a one-shot procedure. From its results, I extract the CRRA coefficients through numerical simulation. The BRET was selected for simplicity, the limited time required to complete it and for being more distinct – in structure and interface – from the pension decision task than the multiple choice lists (Holt & Laury, 2005) or risky investment allocation (Gneezy & Potters, 1997) task. This facilitates partial obfuscation to participants of the preferences and attitudes that I am eliciting from them, to the extent possible.

In the following steps, participants answer a three-question financial literacy quiz (from Lusardi & Mitchell, 2014), and a five-question cognitive reflection test (CRT) using adapted questions from Frederick (2005) and Thomson and Oppenheimer (2016). They receive additional compensation for each block (financial literacy and CRT) if they answer all questions correctly (without the opportunity for a second attempt). For the analysis, the sum of correct answers on both blocks is used and defined as *knowledge score*.

The last decision that the participants make is the time to receive their variable compensation. They can choose to receive it immediately or in 1-4 months with 5% interest per month added to their compensation. The fixed show-up fee is paid separately. The number of months chosen for the compensation delay is defined as *patience*.

Finally, participants navigate to a screen that shows a summary of all their incentivized tasks, the realization of the random choice of the pension decision round for compensation, and the conversion of points of their total compensation into Swiss Francs.

1.2.6 Participants and Incentives

Participants were recruited online, in 2022, from the Swiss panel (German-speaking subjects only) of the commercial market research vendor Bilendi. The panel is a heterogeneous sample of the adult (18-70) population of Switzerland, instead of the most common samples in the longevity beliefs literature that only include older individuals. Bilendi sent e-mail invitations with a brief description of the experiment and compensation. It also handled all payments to participants afterwards, comprising a fixed show-up fee and any variable incentive. The experiment was implemented on oTree (Chen, Schonger, & Wickens, 2016).

Table 1.2 – Sample Characteristics. Statistics per treatment condition and for all subjects in the sample. *Patience* is the subject choice (months) to delay compensation (0-4) for 5% monthly interest. *Knowledge score* is sum of correct answers on the CRT (five) and financial literacy (three) questions. *Happiness and satisfaction index* is the first PCA component of five questions on overall happiness and satisfaction. *Change in health, overall happiness* and the four satisfaction variables are measured on a scale of 0-10. Sample with all subjects who finished decisions at least for the first round of the pension choice task.

	Treatment Condition				
	Fair	Pessimistic	Lump-sum	Reverse	(all)
percentage of observations					
gender: male	42.7	49.9	46.7	43.4	45.8
financial training in school: yes	38.0	31.4	32.0	31.4	33.1
has third-pillar account: yes	71.8	69.9	68.5	64.5	68.7
income– < CHF 3000	24.2	20.3	20.1	20.9	21.4
CHF 3000-3999	11.6	11.1	13.3	11.1	11.8
CHF 4000-4999	14.8	18.7	17.6	15.2	16.6
CHF 5000-5999	13.8	15.0	12.4	11.7	13.3
CHF 6000-6999	10.7	10.3	14.9	13.9	12.4
CHF 7000-7999	10.7	8.1	7.7	12.3	9.7
⩾ CHF 8000	14.2	16.4	13.9	14.9	14.9
education– compulsory schooling	2.8	4.9	2.8	3.1	3.5
vocational high school	34.5	38.6	41.7	43.1	39.5
academic high school	16.8	14.0	12.8	14.6	14.5
technical/prof. school	16.5	13.0	13.3	14.6	14.3
university/post-grad.	29.3	29.5	29.4	24.6	28.3
employment– active, full time	52.4	55.8	55.0	48.7	53.1
active, part time	23.9	22.4	22.5	27.7	24.1
outside workforce	5.7	6.9	3.3	5.9	5.5
retired	5.7	6.6	8.9	7.8	7.3
student	10.0	6.9	8.9	7.3	8.2
unemployed	2.3	1.5	1.4	2.5	1.9
mean					
age	41.387	40.430	41.961	41.896	41.386
CRRA	0.847	0.784	0.841	0.708	0.803
patience	2.616	2.331	2.428	2.627	2.480
knowledge score	4.636	4.664	4.715	4.525	4.648
recent (12mo.) change in health	5.646	5.791	5.954	5.964	5.839
happiness and satisfaction index	-0.061	0.081	0.059	0.013	0.025
overall happiness	7.566	7.740	7.667	7.608	7.649
satisfaction with present life	7.504	7.689	7.682	7.669	7.638
satisfaction with life history	7.194	7.367	7.301	7.279	7.288
satisfaction with finances	6.299	6.357	6.358	6.313	6.333
satisfaction with health	7.401	7.412	7.486	7.401	7.425
median					
total incentivized payoff (CHF)	5.99	6.41	8.84	11.40	7.92
completion time (seconds)	1177	1160	1119	1207	1159
N	351	407	360	357	1475

A total of 2370 participants clicked on invitation links and consented to participate.[21] Among them, 221 violated the monotonicity of survival beliefs and were dropped. Another 362 were dropped after failing the instruction quiz.[22] In total, 1,475 participants made a decision in the first round on the pension choice task.[23] Some descriptive statistics of the characteristics and decisions of the participants (other than survival beliefs and pension choice) for this group are shown in Table 1.2. Of these participants, 155 voluntarily quit or abandoned the experiment (with non-completions concentrated in the *Reverse* treatment condition), and 1340 completed all tasks and earned compensation.

After data collection ended, eight participants were excluded from the sample for reporting gender other than male or female, because life tables are not available for non-binary genders. Another 12 participants were removed for assigning, for any set of survival beliefs, the same survival probabilities for all target ages.

Treatment cells are reasonably balanced in most characteristics. Differences in CRRA coefficients between treatment conditions could be related to wealth-dependent behavior arising from the possible realization of payoffs from the pension decision rounds. Overall, there is a slight skew towards female participants. The happiness and age measures are very similar in all treatment conditions. The variable incentive compensation earned in the experiment ranged from zero to CHF 58.14, with a median of CHF 7.92. The median completion time for the entire session was 19.3 minutes.

[21]The participants were equally split across the fourt treatment conditions at the start of the experiment. Data collection was carried out in several short periods that attracted many simultaneous connections, which, together with the very high number of queries required by our interface design, made it technically unfeasible to dynamically rebalance treatment cells based on participant responses.

[22]Participants were alerted at the welcome and consent screen to the exclusion conditions, which meant they would also not receive any compensation (fixed or variable). The quiz, as implemented, also serves as an attention check.

[23]This extended sample is used in some estimations, where noted.

1.3 Results

1.3.1 Longevity Biases

A first examination of the average cumulative subjective survival probabilities at all target ages, summarized in Table 1.3, shows the usual pattern of overestimation (underestimation) of mortality (survival) at earlier target ages, and vice versa at old target ages,[24] (in line with Wu, Stevens, & Thorp, 2015; Heimer, Myrseth, & Schoenle, 2019) for beliefs about oneself. Women underestimate their own survival up to the target age 90 and men up to age 80.

Absolute deviations between actual and subjective probabilities are high at typical ages of the first decade of retirement: women assess an average probability of survival up to age 75 of only 76.9%, while the actual probability (from the life table) is 87.5%. At very old target ages, the overestimation of longevity is also large: men assess their probability of living up to age 95 to be 25.1%, while the actual probability is 9.4%.

Interestingly, there is also a pervasive bias in survival beliefs about an average person of the same age and gender. For most target ages, the survival beliefs for the average person are closer to the beliefs about oneself than to the actuarial neutral probabilities from the Swiss life table.

The survival beliefs about an average person do not incorporate private longevity information nor relative pessimistic attitudes a subject might hold about risks concerning only him or herself. This suggests the possibility that individuals might be misinformed about human longevity in general. Misinformation, in this context, does not necessarily mean a lack of factual knowledge about human longevity. It could as well arise from cognitive editing processes on risk assessment, such as probability weighting (Prelec, 1998).

Figure 1.3 shows the averages of the (subjective) survival probabilities in the upper graphs. In the lower graphs, the plots are for the average survival biases for oneself, average person and family $\left(q_{i,t}^{j}\right)$, and longevity pessimism $\left(\psi_{i,t}\right)$, across target ages. Higher values indicate more pessimistic

[24] The smaller number of observations for target ages younger than 50 are due to the presence in the sample of subjects with ages between 50-70, whose beliefs are only elicit at a smaller set of target ages.

Table 1.3 – Longevity Beliefs and Biases per Target Age. Longevity beliefs are cumulative survival probabilities between current age and each future target age (rows). *Life table* are actuarial unbiased probabilities from the Swiss life table for 2021. *Oneself* are beliefs of the subject about his/her longevity. *Average person* are beliefs of the subjects about the longevity of an average Swiss person of the same current age and gender. *Family* are subject's expectation of the longevity beliefs of family and friends about the subject. The three $\iota^{j}_{i,t}$ variables measure scaled survival biases for oneself, average person and family, respectively, until each target age. *Pessimism* is the cumulative modeled index of longevity pessimism from the first target age (higher values indicate more pessimism).

	Longevity Beliefs				Survival Bias (scaled)				
target age	life table	oneself	average person	family	$\iota^{own}_{i,t}$	$\iota^{pop}_{i,t}$	$\iota^{fam}_{i,t}$	longevity pessimism	Obs.
female									
50	0.991	0.932	0.935	0.948	8.266	8.033	6.431	0.342	546
55	0.985	0.919	0.921	0.937	5.759	5.473	4.328	0.348	592
60	0.973	0.898	0.898	0.920	3.951	3.836	3.004	0.323	636
65	0.955	0.877	0.869	0.899	2.704	2.899	2.223	0.284	675
70	0.926	0.830	0.822	0.861	2.395	2.473	1.961	0.290	711
75	0.875	0.769	0.763	0.804	1.850	1.885	1.557	0.298	712
80	0.789	0.672	0.677	0.718	1.556	1.531	1.335	0.324	712
85	0.647	0.566	0.583	0.619	1.231	1.180	1.078	0.334	712
90	0.425	0.405	0.432	0.463	1.036	0.987	0.934	0.351	712
95	0.174	0.274	0.305	0.317	0.879	0.842	0.827	0.352	712
100	0.034	0.142	0.168	0.176	0.888	0.861	0.853	0.369	712
105	0.002	0.077	0.095	0.104	0.925	0.907	0.898	0.384	712
male									
50	0.986	0.931	0.925	0.931	5.849	6.463	5.798	0.190	375
55	0.976	0.918	0.903	0.912	3.705	4.607	4.197	0.140	432
60	0.959	0.898	0.882	0.892	2.673	2.958	2.627	0.104	496
65	0.927	0.869	0.851	0.864	1.913	2.060	1.867	0.064	539
70	0.882	0.823	0.802	0.821	1.588	1.759	1.567	0.058	606
75	0.799	0.751	0.730	0.756	1.240	1.336	1.205	0.067	608
80	0.681	0.649	0.631	0.658	1.099	1.151	1.064	0.101	608
85	0.509	0.533	0.514	0.542	0.950	0.989	0.930	0.132	608
90	0.285	0.379	0.362	0.379	0.868	0.891	0.867	0.160	608
95	0.094	0.251	0.243	0.251	0.826	0.836	0.827	0.189	608
100	0.014	0.142	0.138	0.139	0.870	0.874	0.872	0.219	608
105	0.001	0.080	0.079	0.084	0.921	0.922	0.917	0.248	608

beliefs, and zero indicates neutral (not pessimistic or optimistic) beliefs. Cumulative[25] longevity pessimism is present for both genders, but a lower level is present for males in all target ages. Survival bias decreases on target ages for all sets of beliefs of both genders.

Noticeably, women are more pessimistic about their own survival than they think their family and relatives are, at all target ages. Men, on the other hand, are on average consistently more optimistic

[25]Accumulated since the subject's current age, not only at a specific target age as in the upper graphs

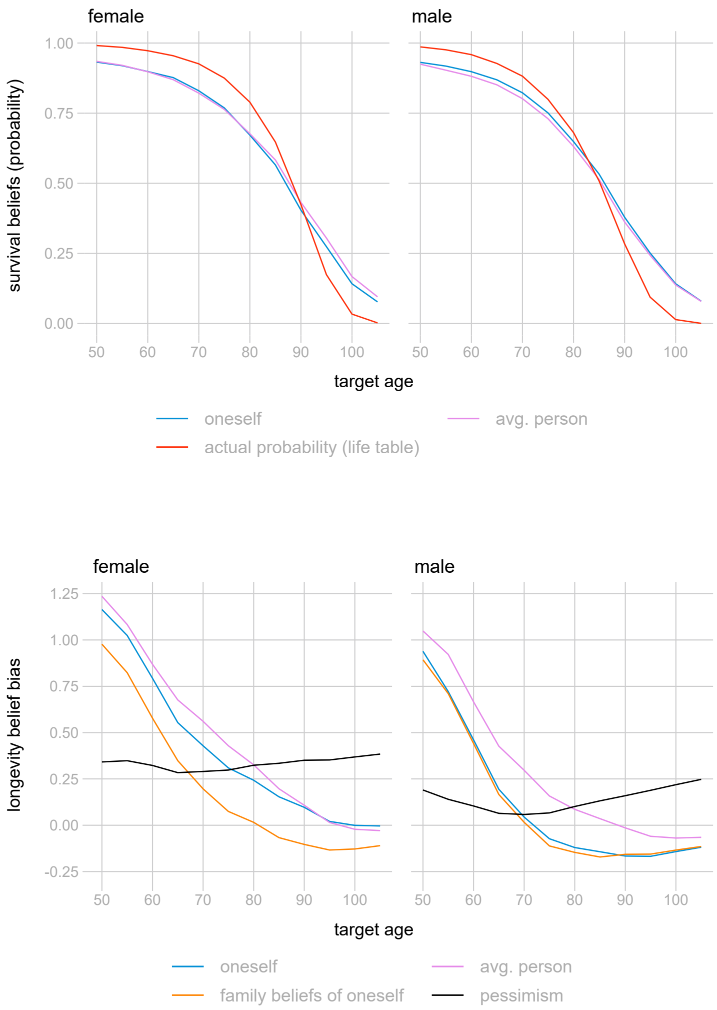


Figure 1.3 – Bias and Pessimism of Longevity Beliefs. The upper graphs show the sample averages of implicit survival probabilities between *current age* and future *target age* for each subject's belief about oneself and an average Swiss person of the same age and gender. Actual probabilities are from the 2021 Swiss life table. The bottom graphs shows averages for cumulative survival bias $\left(\psi_{i,t}\right)$, for survival beliefs about oneself, about an average Swiss person of same *age* and *gender*, and for one's family and friends' belief about the subject's longevity. Longevity *pessimism* $\left(q^{j}_{i,t}\right)$ is modeled at the individual level for every span between current *age* and *target age* relevant for each subject (the higher the values for the four measures, the more pessimistic a person is, zero implies neutral beliefs).

about themselves than for survival an average Swiss man of their age. Men become optimistic about their own survival $\left(q^{own}_{i,t} < 0\right)$ after target age 70, whereas women become neutral only at target age 95.

Of particular interest is the comparison of survival beliefs between oneself and the average person. They reflect subjects' relative assessment of their longevity, compared to peers of same age and gender that do not share the subject's own idiosyncratic longevity risk factors – such as medical

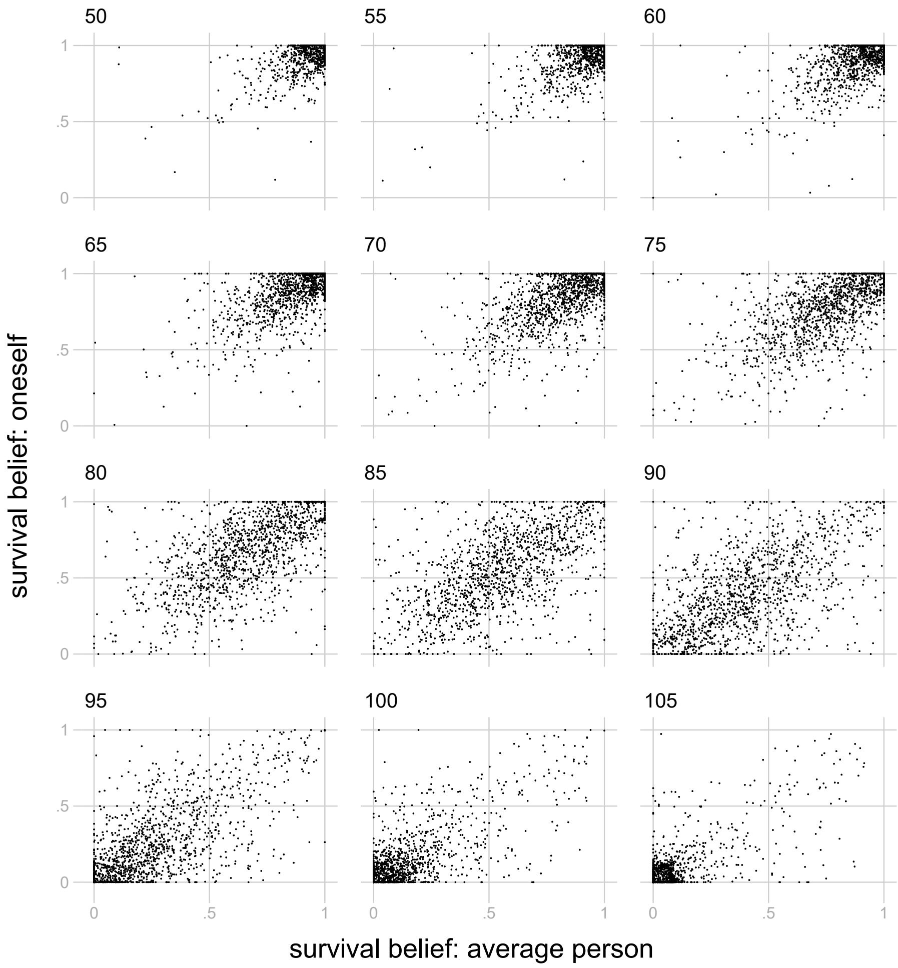


Figure 1.4 – Survival Beliefs about Oneself and Average Person. For every *target age* plot, each dot is the pair of each subject's reported survival probabilities for oneself and for an average Swiss person of the same current age and gender.

diagnosis or family history –, while sharing cohort longevity risks. In Figure 1.4, each dot represents a pair of one subject's survival beliefs for oneself and for an average person, at different target ages. From their joint distribution, probabilities coalesce at their extremes for both the highest and the lowest target ages. There is more dispersion and outliers at later target ages.

The dispersion of beliefs is at its highest for target ages 80 to 90, which is also the age range where the senescence effect (the marginal increase on one-year mortality risk) is particularly important for remaining life expectancy. From the density mass of the joint distribution plots, subjects convey a basic understanding of the 'very high' and 'very low' survival probabilities at both tails of the target age sets $\mathbf{F}_i$. There is much more dispersion in target ages that represents the transition between low mortality risk in middle age and high mortality risk in advanced old age. Hence, the covariance of survival beliefs about oneself and average person is higher when survival probabilities are moderate.

Proceeding further, I examine the factors that drive longevity pessimism and survival bias measures, regressing them on subject characteristics. Results are displayed in Table 1.4; Higher values for all dependent variables indicate more pessimistic subjects. Happiness is negatively and significantly associated with longevity and survival pessimism: happier subjects are less pessimistic and biased about their longevity, except for the relative comparison between beliefs about oneself and from family and friends about oneself. The effect size is moderate. As pessimism $\psi_{i,t=105}$ is a logarithmic transformation of ratios, each additional unit of the happiness and satisfaction index is associated with 2.42% less distortion of weighted-average mortality probabilities, as aggregated up to target age 105.

Recent *change in health* is also negatively associated with longevity pessimism and with smaller survival biases measured against actuarial unbiased probabilities. Subjects whose health has improved more within the last year have less negatively distorted assessments of probabilities of oneself, average person and family and friends. Although a change in personal health is a significant (and trivial) driver of actuarial or subjective survival in general (Heimer, Myrseth, & Schoenle, 2019), the results suggest that it also affects the subjects' perceptions of longevity of an average person. This could not be explained by any incorporation of private longevity information that is only relevant for the subjects' own survival.

Otherwise, apart from the stylized gender difference in longevity pessimism (women are more pessimistic than men) firmly established in the literature, no other individual characteristic significantly and consistently affects longevity pessimism and survival bias.

Table 1.4 – Formation of Longevity Belief Biases and Pessimism. OLS regressions of *pessimism* and other survival bias measures. *Employment: k* are indicators that equal one for each category *k* of employment status, and zero otherwise. See Table 1.5 for the definition of other variables.

	(1)	(2)	(3)	(4)	(5)	(6)
dependent variable:	$\psi_{i,t=105}$	$q_{i,t=105}^{own}$	$q_{i,t=105}^{own:pop}$	$q_{i,t=105}^{own:fam}$	$q_{i,t=105}^{pop}$	$q_{i,t=105}^{fam}$
happiness and satisfaction	−0.0245***	−0.0450***	−0.0278***	0.0030	−0.0173**	−0.0384***
	[0.0051]	[0.0072]	[0.0076]	[0.0094]	[0.0069]	[0.0076]
change in health	−0.0083**	−0.0446***	0.0048	−0.0013	−0.0447***	−0.0407***
	[0.0042]	[0.0084]	[0.0061]	[0.0092]	[0.0082]	[0.0082]
CRRA	0.0073	−0.0201	0.0165	0.0056	−0.0172	−0.0178
	[0.0166]	[0.0126]	[0.0241]	[0.0133]	[0.0109]	[0.0146]
patience	−0.0003	−0.0048	−0.0003	0.0035	−0.0005	−0.0056
	[0.0045]	[0.0072]	[0.0066]	[0.0089]	[0.0070]	[0.0073]
financial training: yes	−0.0070	−0.0134	−0.0095	0.0015	−0.0008	0.0011
	[0.0183]	[0.0293]	[0.0272]	[0.0374]	[0.0282]	[0.0303]
knowledge score	−0.0086*	0.0098	−0.0181***	−0.0013	0.0168**	0.0013
	[0.0047]	[0.0077]	[0.0069]	[0.0104]	[0.0078]	[0.0087]
gender: male	−0.1422***	−0.0974***	−0.0786***	−0.1570***	−0.0328	0.0298
	[0.0170]	[0.0274]	[0.0251]	[0.0342]	[0.0268]	[0.0284]
age	0.0032***	−0.0016	−0.0010	0.0010	−0.0010	−0.0025**
	[0.0008]	[0.0011]	[0.0011]	[0.0016]	[0.0010]	[0.0011]
education	−0.0062	−0.0017	−0.0111	−0.0039	0.0066	−0.0008
	[0.0062]	[0.0112]	[0.0090]	[0.0132]	[0.0108]	[0.0115]
employment: active, part-time	0.0062	0.0080	0.0025	−0.0560	0.0166	0.0526
	[0.0218]	[0.0332]	[0.0327]	[0.0398]	[0.0328]	[0.0336]
employment: outside workforce	−0.0082	0.0331	−0.0173	−0.0205	0.0160	0.0467
	[0.0333]	[0.0543]	[0.0484]	[0.0932]	[0.0509]	[0.0636]
employment: retired	0.0423	0.0750	0.0529	0.0588	0.0168	0.0532
	[0.0398]	[0.0533]	[0.0562]	[0.0876]	[0.0434]	[0.0586]
employment: student	−0.0063	−0.0177	0.0006	0.0680	−0.0258	−0.0852
	[0.0267]	[0.0444]	[0.0415]	[0.0589]	[0.0447]	[0.0521]
employment: unemployed	0.0028	0.0747	−0.0415	0.2427**	0.1239	−0.0076
	[0.0415]	[0.0759]	[0.0691]	[0.1216]	[0.0977]	[0.0918]
constant	0.3547***	0.2949***	0.2217***	0.2530**	0.1784**	0.2279**
	[0.0496]	[0.0897]	[0.0726]	[0.1122]	[0.0847]	[0.0913]
Adjusted R^2	0.106	0.100	0.032	0.012	0.058	0.065
N	1276	1276	1276	1276	1276	1276

Heterokedasticity-robust standard errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.3.2 Pension Decisions

In the main task, subjects decide in which period they want to collect (*Lump-sum* treatment condition), start collecting (*Fair* and *Pessimistic*) or start paying (*Reverse*) pension benefits. Figure 1.5 shows the distribution of the average (in all rounds) of the pension choice for benefits payoff per subject, per treatment.

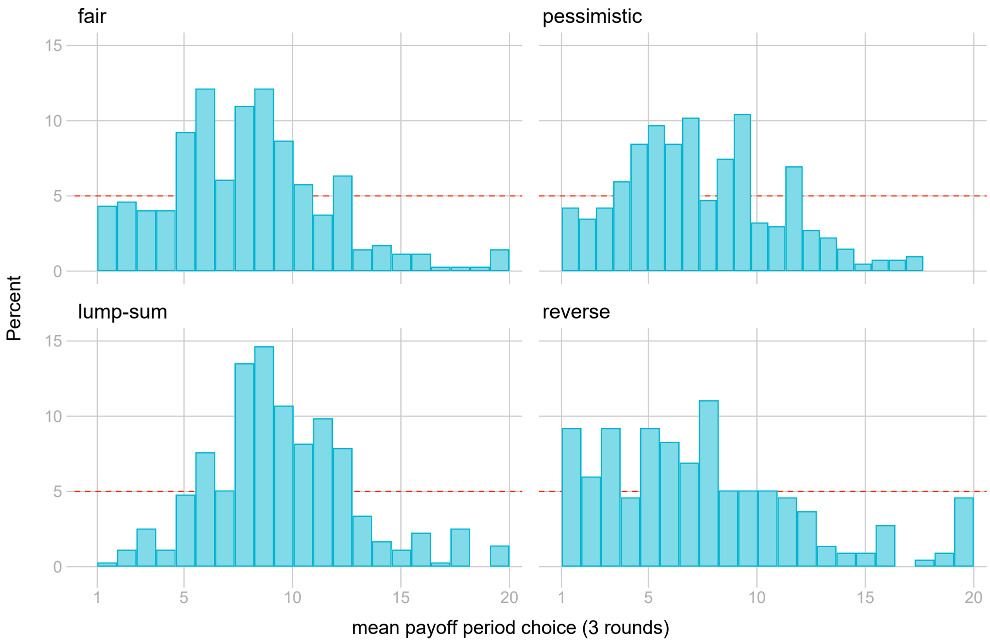


Figure 1.5 – Pension Choice and Treatments. Graphs show the distribution of mean *pension decision* of payoff period chosen by subjects across three rounds, per treatment condition. On *Fair* and *Pessimistic* conditions, the decision is when to start receiving payoffs. On *Lump-sum*, the decision is on the timing of the single payoff. On *Reverse*, the decision is when to start making payments. Uniform distribution highlighted in red.

In the first round, subjects made an average pension choice for payoff at 8.71 periods in *Fair*, 8.37 in *Pessimistic*, 10.02 in *Lump-sum* and 6.37 in *Reverse*. The dispersion of choices is very similar across all conditions, with a standard deviation of choice ranging from 4.37 to 4.67. Taking into account the average of all rounds, subjects chose pension payoffs on average at period 7.95 (*Fair*), 7.36 (*Pessimistic*), 9.53 (*Lump-sum*) and 7.84 (*Reverse*). The general results for the *Fair* and *Lump-sum* decisions are similar to those of Fatas, Lacomba, and Lagos (2007) and reproduced by Bachmann et al. (2022), which used the difference between these two conditions as their main treatment effect.

Fewer subjects chose early *Lump-sum* payouts, compared to the fraction of subjects who chose to start receiving annuity payments earlier in *Fair* and *Pessimistic*. The expected value of the payoffs of the latter is higher than for the other treatments in earlier periods, as discussed in Subsection 1.2.3, yet a cursory inspection of the pension choice distribution does not show an obvious right-skew that the favorable distortion of present values in earlier periods should attract from risk-neutral or risk-averse subjects.

As discussed in Subsection 1.2.6, the cases of subjects who quit the experiment after the beginning of the pension decision task were concentrated on *Reverse* treatment. Therefore, results of this treatment should be interpreted with some caution on the possibility of an endogenous treatment effect on subjects who quit the experiment. Notwithstanding, a larger fraction of subjects in this condition chose to start making payments earlier (out of their endowment, specific to this condition) than subjects chose to (start) earning payoffs in other conditions. This could indicate a preference to avoid bankruptcy risks associated with a delay in start of payments (out of the subject endowment) to intermediate periods with moderate survival probabilities.

A particular concern with the pension choice task is that it allows risk-seeking participants to gamble for very high payoffs (upwards of CHF 150 when converted to monetary compensation under three of the conditions) with low probability (5%) by choosing the last period (20) in all rounds. There are 19 such extreme cases (out 1475 observations in the extended sample) of subjects with pension choice at period 20 in all rounds: 5 in *Fair*, 5 in *Lump-sum* and 9 in *Reverse*. At the opposite extreme, 33 subjects always made the pension choice for the first period in all rounds.

Otherwise, the pension choice is somehow sticky: 1102 subjects made identical decisions in all three rounds. Excluding those cases, 146 subjects repeated their decision for the first and second rounds only, and 220 for the second and third rounds only.

The effects of longevity pessimism on these pension decisions are summarized in Table 1.5. Pessimism at target age 80 $(\psi_{i,t=80})$ is only weakly associated with pension choice. Since the variable is a logarithmic of ratios and increasing values indicate increasing pessimism, each additional unit of pessimism reduces, on average, the choice payoff period by 0.52 and 0.48 periods in specifications (1) and (2), respectively.

This effect is not significant in (3), when I introduce *change in health* as a control. There, each unit of recent positive change in health (on a scale 0-10) delays the pension choice by 0.19 periods. Change in health is also relevant in additional specifications that include more subject covariates (4) or restrict the sample to decisions in the first round (6).

Treatment effects are significant for the condition *Lump-sum* on the *Fair* baseline in all specifications, associated with a delay in pension choice of 1.35 to 1.56 periods. Adding additional controls for subject characteristics and preferences does not substantially change the coefficient of the

Table 1.5 – Longevity Pessimism and Pension Choice. OLS regressions of the pension choice (payoff period). In (1-4) the dependent variable is the average of periods chosen on 3 round. In (5,6) it is the pension choice of the first round only. *Longevity pessimism* measured at target age 80 ($\psi_{i,t=80}$) for each subject; higher values indicate more pessimist subjects, zero indicates neutral (unbiased) beliefs. *Pessimistic*, *Lump-sum* and *Reverse* are indicators that equal one for the treatment condition and zero otherwise (*Fair* is the baseline condition). *Gender: male* and *financial training: yes* are indicators that equal one if for the respective categories, and zero otherwise.*Age* measured in years and *education* on as levels 1-5. *Happiness and satisfaction* is an index equal to the first factor a PCA analysis on 5 measures of overall happiness and satisfaction. Recent (1yr.) *change in health* is measured on a scale 0-10. *Knowledge score* is the number of correct answers (0-8) on a financial literacy quiz and CRT, combined. *CRRA* is the risk-aversion coefficent from a power utility model extracted from the "bomb" risk elicitation task (BRET). *Patience* is the delay choice, in months, (0-4) of subject compensation in exchange of interest.

dep. variable: pension choice	(1)	(2)	(3)	(4)	(5)	(6)
rounds:	all rounds	all rounds	all rounds	all rounds	1st round	1st round
longevity pessimism	−0.737*	−0.649*	−0.549	−0.522	−0.900**	−0.706
	[0.383]	[0.384]	[0.381]	[0.385]	[0.431]	[0.432]
treatment: Pessimistic	−0.515*	−0.495*	−0.507*	−0.496*	−0.208	−0.152
	[0.299]	[0.297]	[0.297]	[0.301]	[0.365]	[0.370]
treatment: Lump-sum	1.560***	1.554***	1.497***	1.482***	1.360***	1.354***
	[0.291]	[0.291]	[0.289]	[0.286]	[0.360]	[0.359]
treatment: Reverse	−0.471	−0.460	−0.496	−0.497	−2.368***	−2.374***
	[0.427]	[0.429]	[0.429]	[0.434]	[0.478]	[0.487]
Pessimistic × longevity pessimism	0.252	0.204	0.160	0.219	0.143	0.233
	[0.485]	[0.485]	[0.482]	[0.495]	[0.576]	[0.585]
Lump-sum × longevity pessimism	0.403	0.350	0.336	0.325	−0.032	−0.105
	[0.463]	[0.462]	[0.456]	[0.456]	[0.573]	[0.564]
Reverse × longevity pessimism	0.757	0.708	0.655	0.525	0.705	0.520
	[0.753]	[0.757]	[0.750]	[0.745]	[0.810]	[0.801]
gender: male		0.200	0.188	0.291		0.340
		[0.223]	[0.222]	[0.223]		[0.270]
age		0.021***	0.023***	0.022***		0.022**
		[0.008]	[0.008]	[0.008]		[0.009]
education		0.084	0.098	0.112		0.169
		[0.081]	[0.081]	[0.084]		[0.105]
happiness and satisfaction			−0.024	−0.005		−0.035
			[0.060]	[0.060]		[0.076]
change in health			0.194***	0.175***		0.234***
			[0.059]	[0.059]		[0.071]
financial training: yes				−0.277		−0.166
				[0.229]		[0.277]
knowledge score				−0.110*		−0.122
				[0.063]		[0.079]
CRRA				0.308***		0.249***
				[0.056]		[0.062]
patience				0.150***		0.205***
				[0.057]		[0.071]
constant	8.005***	6.746***	5.521***	5.546***	8.812***	5.673***
	[0.226]	[0.435]	[0.554]	[0.659]	[0.267]	[0.826]
Adjusted R^2	0.050	0.057	0.064	0.082	0.063	0.088
N	1320	1320	1320	1276	1320	1276

Heterokedasticity-robust standard errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

treatment indicator. Meanwhile, the treatment effect for *Reverse* is sizeable and significant only for the decisions in the first round. Finally, *Pessimistic* treatment effects are small and weakly significant only, regardless of the additional control variables added to the main specification in (2-4).

All interactions of *treatment* indicators and longevity pessimism are not significant in all specifications. This shows that to the limited extent that pessimism about longevity in general affects experimental pension choice decisions, this does not occur at significantly different margins for any of the treatments compared to the baseline *Fair*.

Some additional personal characteristics and preferences impact pension choice on their own. *Age* has a highly significant but small effect, delaying the pension choice by 0.02 period for each additional year of chronological age in several specifications. *Patience* concerning delay of monetary compensation is positively associated with a later pension choice of payoffs.

The less risk-averse subjects are, as measured by the *CRRA* coefficient[26] from the "bomb" risk elicitation task (Crosetto & Filippin, 2013), the more they delay pension choice as well. Since this risk-aversion elicitation task is presented to participants after the main pension choice task, its results could still be affected by the realization of termination periods over its tree rounds, even if participants will only be informed on which round will be used for monetary compensation at the end of the experiment.

The findings remain qualitatively unchanged when using several cumulative survival bias measures, instead of the modeled *longevity pessimism*. In Table 1.6, survival bias for beliefs about oneself with respect to actual probabilities (1,2) and the belief of family and friends about oneself (5,6) are significant drivers of pension choice.

The more pessimistic the subject in those two measures is (the higher $q^j_{i,t=80}$), the earlier their pension payoff choice is. The bias implicit in the subject's survival probabilities with respect to those about an average person is not significant for the pension decision in specifications (3,4).

Similarly to the main results, the treatment effects for *Lump-sum* are strong and significant for all

[26]The higher the coefficient of a classic CRRA power utility function, the less risk-averse the subject is.

Table 1.6 – Survival Belief Bias and Pension Choice. OLS regressions of the pension choice (payoff period) on different survival bias measures (accumulated until target age 80) of survival beliefs about oneself: from actual (life table) probabilities (1,2), from average Swiss person of same age and gender (2,3); from the belief of family and friends about the subject's survival (5,6). See Table 1.5 for the definition of other variables.

dep. variable: pension choice	(1)	(2)	(3)	(4)	(5)	(6)
measure of longev. belief bias:	$q^{own}_{i,t=80}$	$q^{own}_{i,t=80}$	$q^{own:pop}_{i,t=80}$	$q^{own:pop}_{i,t=80}$	$q^{own:fam}_{i,t=80}$	$q^{own:fam}_{i,t=80}$
survival belief bias	-0.396^{**}	-0.324^{*}	-0.267	-0.249	-0.439^{**}	-0.425^{**}
	[0.193]	[0.194]	[0.260]	[0.257]	[0.191]	[0.189]
treatment: Pessimistic	-0.466^{*}	-0.477^{*}	-0.451^{*}	-0.474^{*}	-0.533^{*}	-0.551^{**}
	[0.268]	[0.268]	[0.264]	[0.263]	[0.276]	[0.276]
treatment: Lump-sum	1.581^{***}	1.536^{***}	1.640^{***}	1.577^{***}	1.511^{***}	1.453^{***}
	[0.271]	[0.270]	[0.268]	[0.266]	[0.280]	[0.279]
treatment: Reverse	-0.333	-0.372	-0.283	-0.331	-0.516	-0.576
	[0.392]	[0.394]	[0.384]	[0.385]	[0.393]	[0.391]
Pessimistic × *longevity belief*	0.015	-0.022	0.076	0.071	0.108	0.101
	[0.247]	[0.246]	[0.339]	[0.336]	[0.271]	[0.271]
Lump-sum × *longevity belief*	0.148	0.116	0.231	0.249	0.293	0.283
	[0.240]	[0.239]	[0.308]	[0.304]	[0.253]	[0.250]
Reverse × *longevity belief*	0.286	0.253	0.553	0.549	0.586	0.624
	[0.351]	[0.350]	[0.561]	[0.555]	[0.398]	[0.399]
gender: male	0.201	0.189	0.267	0.240	0.228	0.202
	[0.219]	[0.218]	[0.220]	[0.218]	[0.219]	[0.218]
age	0.018^{**}	0.021^{***}	0.021^{***}	0.022^{***}	0.022^{***}	0.023^{***}
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
education	0.090	0.106	0.085	0.096	0.093	0.105
	[0.080]	[0.080]	[0.081]	[0.081]	[0.081]	[0.081]
happiness and satisfaction		-0.043		-0.009		-0.012
		[0.060]		[0.060]		[0.059]
change in health		0.181^{***}		0.201^{***}		0.202^{***}
		[0.059]		[0.059]		[0.058]
constant	6.759^{***}	5.576^{***}	6.582^{***}	5.364^{***}	6.691^{***}	5.457^{***}
	[0.426]	[0.548]	[0.422]	[0.543]	[0.425]	[0.546]
Adjusted R^2	0.061	0.067	0.054	0.063	0.058	0.067
N	1320	1320	1320	1320	1320	1320

Heterokedasticity-robust standard errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

bias measures, the treatment effects for *Pessimistic* are limited and weakly significant, and none of the treatment interactions and the three survival biases are significant.

1.3.3 Happiness Indirect Effects on Pension Decision

Happiness and satisfaction can affect longevity pessimism itself, while also having a direct impact on pension choice. Changes in health can also influence longevity pessimism, directly through changes in subjective survival beliefs that, in this case, are related to the subject receiving new private information that directly affects their longevity (Hurd & McGarry, 2002; Kvaerner, 2022).

I attempt to unravel these direct and indirect effects.[27] The three-way relationship between longevity pessimism, change in health and happiness and satisfaction is plotted in Figure 1.6. The heat map splits all individual observations at the subject level into many bins, according to their joint distribution of happiness and satisfaction and longevity pessimism. The area of the circles is the number of observations in each bin, and its color is the average change in health within that bin. The dashed red lines divide the plot into four quadrants, whose subjects I characterize as sad and pessimist (top left), happy and pessimist (top right), happy and optimist (bottom right), and sad and optimist (bottom left).[28]

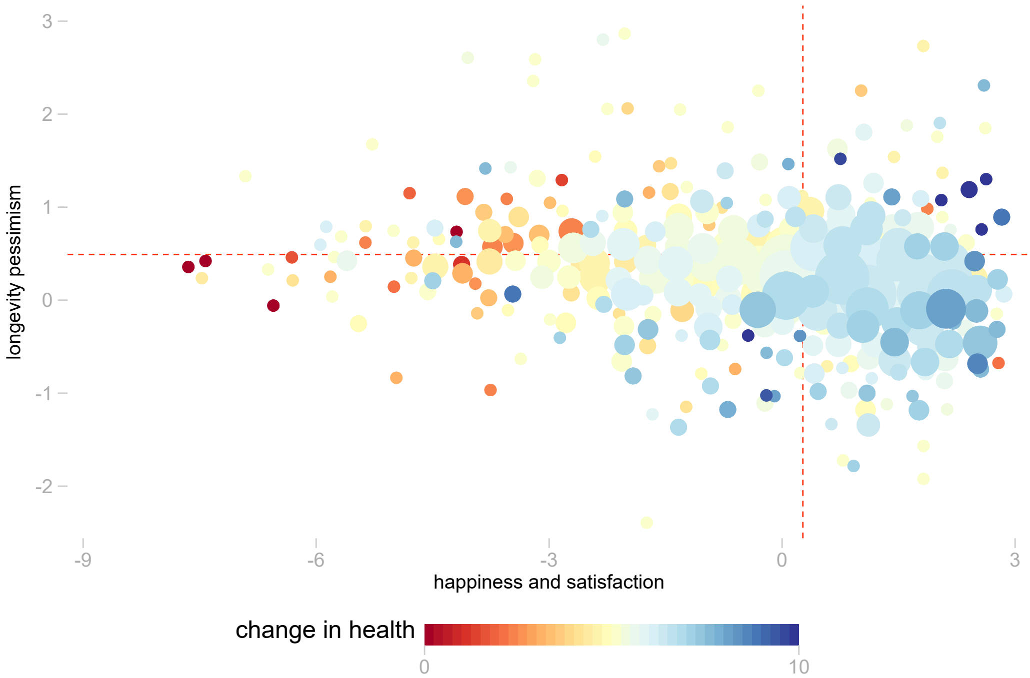


Figure 1.6 – Longevity Pessimism, Change in Health and Happiness. Average *change in health* (colored circles) shown according to groups of the joint distribution of subjects across longevity *pessimism* (the higher its value, the more pessimistic a subject is) and *happiness* indexes. *Change in health* is the reported twelve-month change in the health status (5 implying no change). *Happiness and satisfaction* is the first PCA component of a set of five questions on overall happiness and satisfaction with present life, life history, finances and current health.Longevity *pessimism* is modeled at the individual level for every span between current *age* and all *target age* relevant for each subject (the higher its value, the more pessimistic a person is). Sample medians highlighted in red. The area of the circles are proportional to the number of subjects within each group.

The relationship between change in health and happiness is clear as it is trivial: people who had more negative changes in recent health also have a lower happiness and satisfaction index, reflecting the negative impact of receiving bad medical news or perceiving a deterioration of one's own health.

[27] In the pre-registration of this study, a full moderated-mediation model was proposed. However, considering effect size of longevity beliefs on pension decisions, any expected indirect effect is also limited ex ante.

[28] Higher values on *longevity pessimism* indicate more pessimistic subjects.

In Apicella and De Giorgi (2022), bad health news leads to changes in sentiment that affect subjective longevity beliefs and actual survival probabilities. In this study, the relationship is less clear with respect to the effect of change in health on longevity pessimism. Some of the bins with highest average change of health – implying an improvement on the subject's health status – are in the quadrant 'happy and pessimist'. Also, most of the bins 'happy and optimist' have higher (positive) changes in health. Bins with the lowest reported change in health are weakly skewed towards the 'sad and pessimist' quadrant.

Given these broad distribution patterns – which seem to indicate nonlinearity of the relationship of these three variables – and the presence of categorical concomitants, a simple decomposition of effects of happiness or change in health could be biased. To address this concern, I use a linearized form of the KHB-decomposition (Breen, Karlson, & Holm, 2013, 2021) to identify these indirect effects. Table 1.7 summarizes its results. For this analysis, the observations are at the subject×target age level.

For each of the key variables, *reduced* is the coefficient of this variable with respect to *pension choice* when *pessimism* is not included as a regressor, *full* is the coefficient when *pension choice* is included, and *indirect effect* is the difference between the coefficients and its significance, indicating the how much of the effect of variables on pension choice is absorbed and confounded through the effects of longevity pessimism on pension choice.

Indirect effects through longevity pessimism are relevant for all specifications and variables. In specification (1), longevity pessimism confounds 52.1% of the effect of happiness and satisfaction on pension choice, 4.0% of the effect of change in health, and 12.4% of the effect of the gender indicator.

The results should still be taken with the caveat that the coefficients of change in health and happiness are, in general, relatively small and that both variables are partially correlated $\left(\rho = 0.329\right)$ through a plausibly casual relationship. For this reason, I also investigate the decomposition of the effects of both variables separately from each other. In specification (2), longevity pessimism confounds 23.0% of the effect of happiness and satisfaction on pension decision. In specification (3), longevity pessimism confounds 6.45% of the effect of change in health on pension decision.

Table 1.7 – Indirect Effects of Happiness through Longevity Beliefs. The table shows the results of a linearized KHB-decomposition of the direct and indirect effects of *happiness and satisfaction* index, *change in health*, and *gender*, on the average pension choice of payoff period. '*Reduced*' rows the coefficients of a regression excluding the control variable *pessimism*. '*Full*' are the coefficients of a specificiation including the control. '*Indirect Effect*' is the partial effect of the variables on *pension choice* through their own effects on longevity *pessimism*. *Happiness and satisfaction* is an index equal to the first factor a PCA analysis on 5 measures of overall happiness and satisfaction. Recent (1yr.) *change in health* is measured on a scale 0-10. *Gender: male* is an indicator that equals one for male, and zero for female subjects. Concomitant factors (not shown) include *education*, *financial training*, *knowledge score*, *CRRA*, *patience*, *target age* and indicator variables for treatments. See Table 1.5 for other variables' definition. Observations are subject×target age.

dep. var.: pension choice	(1)	(2)	(3)
happiness and satisfaction			
Reduced	0.028	0.083***	
	[0.017]	[0.016]	
Full	0.013	0.064***	
	[0.017]	[0.017]	
Indirect Effect	0.014***	0.019***	
	[0.004]	[0.005]	
change in health			
Reduced	0.166***		0.174***
	[0.017]		[0.016]
Full	0.159***		0.163***
	[0.017]		[0.016]
Indirect Effect	0.007**		0.011***
	[0.003]		[0.004]
gender: male			
Reduced	0.453***	0.469***	0.455***
	[0.064]	[0.064]	[0.064]
Full	0.397***	0.404***	0.397***
	[0.064]	[0.064]	[0.064]
Indirect Effect	0.056***	0.065***	0.059***
	[0.012]	[0.012]	[0.012]
N	14355	14355	14355

Robust standard errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the results suggest a complex structural relationship in which longevity pessimism confounds a larger fraction of the effect of change in health and happiness on pension choice, when both casual factors are simultaneously considered.

1.4 Robustness Checks and Alternative Analysis

1.4.1 Deviations of Realized Longevity

The sets of survival beliefs j allow the calculation of the partial life expectancy, that is, the number of years expected to live within the time period from age to target age, according to equation (1.10). Comparing the differences in partial life expectancy in time (years), up to the target age of 105 years, allows an alternative investigation of the factors that drive longevity and survival biases.

Table 1.8 shows the results of a regression of expected life expectancy on subject characteristics. In this setting, the dependent variables are simple differences, in years, between the partial life expectancy and a benchmark.

The results are similar to those of the main models, which use longevity pessimism and survival biases (see Table 1.4). More happiness is significantly associated with longer partial life expectancy in all measures. In specification (1) each additional unit of the happiness and satisfaction index increases one own's partial life expectancy, relative to actuarial expectations form the life table up to age 105, by 1.04 years. This effect is smaller (0.41 additional years per unit of the index) when the benchmark is of the expected realized longevity of an average person of the same age and gender as the subject (specification 4).

The significance (or lack thereof) of other personal characteristics is similar to those of the main analyses using the pessimism and bias measures.

1.4.2 Pessimism, Longevity Bias and Savings Behavior in the Field

I examine whether longevity pessimism or survival belief biases affect some decisions that subjects make in the field. In particular, I look at the impact of those measures on participation in a tax-incentivized 'third-pillar' individual retirement savings scheme that exists in Switzerland.[29]

The scheme offers, up to a cap, labor income tax deductions for deposits into long-term savings managed accounts, which can then be invested into vetted eligible products and securities. Upon

[29] so-called *pillar 3a accounts*

Table 1.8 – Implicit Subjective Life Expectancy. OLS regression of measures of expected realized longevity at age 105, relative to different benchmarks. *Employment: k* are indicators that equal one for each category k of employment status, and zero otherwise. See Table 1.5 for the definition of other variables.

	(1)	(2)	(3)	(4)	(5)
dependent variable (yrs):	$kex^{own}_{i,t=105}$	$kex^{own:pop}_{i,t=105}$	$kex^{own:fam}_{i,t=105}$	$kex^{pop}_{i,t=105}$	$kex^{fam}_{i,t=105}$
happiness and satisfaction	1.0447***	0.6397***	0.2917**	0.4050***	0.7530***
	[0.1502]	[0.1291]	[0.1376]	[0.1406]	[0.1480]
change in health	0.7464***	0.0812	0.0621	0.6652***	0.6844***
	[0.1386]	[0.1014]	[0.1183]	[0.1318]	[0.1462]
CRRA	0.3398*	0.1002	0.1906	0.2396	0.1492
	[0.2058]	[0.1498]	[0.1417]	[0.1883]	[0.1994]
patience	0.1000	0.0063	0.0445	0.0937	0.0555
	[0.1337]	[0.1039]	[0.1157]	[0.1299]	[0.1335]
financial training: yes	0.4047	0.2203	0.3344	0.1844	0.0703
	[0.5352]	[0.4262]	[0.4683]	[0.5188]	[0.5316]
knowledge score	−0.1678	−0.0244	−0.4051***	−0.1435	0.2373
	[0.1457]	[0.1118]	[0.1262]	[0.1524]	[0.1569]
gender: male	2.6055***	1.3540***	2.6101***	1.2515**	−0.0046
	[0.5285]	[0.4142]	[0.4429]	[0.5272]	[0.5271]
age	0.0607***	−0.0073	−0.0269	0.0681***	0.0876***
	[0.0209]	[0.0158]	[0.0183]	[0.0201]	[0.0205]
education	0.0227	0.1985	0.0212	−0.1758	0.0015
	[0.1993]	[0.1486]	[0.1672]	[0.1916]	[0.2041]
employment: active, full time	0.1702	0.3393	0.8698*	−0.1691	−0.6996
	[0.6266]	[0.4643]	[0.4935]	[0.6208]	[0.6166]
employment: outside workforce	−0.3026	−0.4805	0.9497	0.1779	−1.2522
	[1.1443]	[0.8691]	[0.9500]	[1.0237]	[1.1934]
employment: retired	−1.1400	−1.2894**	−0.1436	0.1494	−0.9964
	[0.9177]	[0.6568]	[0.7104]	[0.7694]	[0.8460]
employment: student	1.0241	−0.4165	−1.4565	1.4406	2.4806**
	[1.1096]	[0.9677]	[1.0922]	[1.0018]	[1.0522]
employment: unemployed	−1.1698	2.3006	−2.1778	−3.4704	1.0080
	[2.2206]	[1.8567]	[2.0914]	[2.7777]	[2.0012]
constant	−9.1023***	−1.4081	−0.2593	−7.6942***	−8.8430***
	[1.7682]	[1.3077]	[1.5885]	[1.6676]	[1.9132]
Adjusted R^2	0.125	0.044	0.034	0.060	0.075
N	1276	1276	1276	1276	1276

Heterokedasticity-robust errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

retirement, the balances are withdrawn within ten years and taxed, in part, at half-rate[30] as usual income. The balance can also be used to purchase a family residence or (within conditions) as seed capital for a new business.

[30]There are substantial cantonal differences in the relative net tax incentives embedded in the scheme.

I evaluate two outcomes: whether the subject has an active third-pillar account (regardless of when it was opened) and, conditional on a positive answer, whether a new deposit in this account was made within the last 12 months.

Table 1.9 – Longevity Beliefs and Long-Term Saving Schemes. The table reports the odds' ratio of a logistic regression of indicator variables on third-pillar accounts. In (1-3), the dependent variable is one if the subjects owns a third-pillar tax-incentivized retirement savings account. In (4-6), the dependent variable is one if the subject – conditional on having an account – made a qualified deposit within the last year. *Longevity pessimism* measured at target age 80 ($\psi_{i,t=80}$) for each subject; higher values indicate more pessimist subjects, zero indicates neutral (unbiased) beliefs. *Income* is defined in levels 1-8. See Table 1.5 for the definition of other variables.

3rd-pillar scheme:	(1) account	(2) account	(3) account	(4) deposit	(5) deposit	(6) deposit
longevity pessimism	0.126			0.082		
	[0.128]			[0.209]		
survival bias $q^{own:pop}_{i,t=80}$		0.053			−0.009	
		[0.086]			[0.139]	
survival bias $q^{own}_{i,t=80}$			0.064			0.130
			[0.068]			[0.106]
gender: male	-0.024	−0.045	−0.031	0.103	0.085	0.110
	[0.155]	[0.153]	[0.154]	[0.252]	[0.251]	[0.250]
age	-0.005	−0.005	−0.004	-0.058^{***}	-0.058^{***}	-0.057^{***}
	[0.006]	[0.005]	[0.006]	[0.010]	[0.010]	[0.010]
education	0.009	0.010	0.007	−0.106	−0.108	−0.106
	[0.062]	[0.062]	[0.062]	[0.101]	[0.101]	[0.101]
happiness and satisfaction	0.128^{***}	0.123^{***}	0.128^{***}	0.175^{***}	0.168^{***}	0.188^{***}
	[0.042]	[0.042]	[0.043]	[0.066]	[0.065]	[0.066]
change in health	-0.085^{**}	-0.087^{**}	-0.082^{**}	−0.082	−0.082	−0.078
	[0.040]	[0.040]	[0.040]	[0.064]	[0.064]	[0.064]
income	0.434^{***}	0.434^{***}	0.432^{***}	0.249^{***}	0.249^{***}	0.248^{***}
	[0.046]	[0.046]	[0.046]	[0.068]	[0.068]	[0.068]
financial training: yes	0.407^{**}	0.404^{**}	0.406^{**}	0.053	0.053	0.048
	[0.162]	[0.162]	[0.162]	[0.246]	[0.245]	[0.247]
knowledge score	0.046	0.047	0.045	−0.015	−0.015	−0.017
	[0.043]	[0.043]	[0.043]	[0.075]	[0.075]	[0.075]
CRRA	0.019	0.018	0.021	0.242^{*}	0.239^{*}	0.257^{*}
	[0.041]	[0.042]	[0.042]	[0.130]	[0.129]	[0.134]
patience	0.115^{***}	0.115^{***}	0.115^{***}	0.015	0.016	0.017
	[0.040]	[0.040]	[0.040]	[0.061]	[0.061]	[0.061]
constant	-0.596	−0.557	−0.609	4.101^{***}	4.119^{***}	4.036^{***}
	[0.465]	[0.464]	[0.465]	[0.806]	[0.799]	[0.805]
N	1124	1124	1124	777	777	777

Heterokedasticity-robust standard errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In the logistic regressions shown in Table 1.9, longevity pessimism and survival biases do not affect the odds of a subject owning or making a deposit into a third-pillar retirement account. *Income* is positively associated with participation in the scheme, which is expected because higher effective personal income tax rates make participation in the program more attractive to individuals. *Financial training* increase the odds of a subjective owning an account, which might be reflected in better knowledge of the taxation and/or pension system. *Patience* is positively associated with ownership of a third-pillar account, but not with recent deposits in the accounts.

Happiness and satisfaction affect participation in the savings scheme, while a positive change in health reduces the odds of ownership of an account while having no influence on the offs of a recent deposit. The latter result is, to some extent, puzzling: a recent change in health could affect recent deposit more than the status on whether a subject opened an account possibly many years earlier.

1.4.3 Survival and Pension Decisions – Additional Controls

In the discussion of the main results (Subsection 1.3.2), I presented results of the effect of survival bias measures on pension decision.

In Table 1.10, I show additional regressions that expand those models, adding more demographic controls.

The results are qualitatively unchanged with respect to the effects of the main variables of interest – the survival bias measures and treatment effects –. The additional variables on demographic characteristics and preferences follow mainly the patterns observed with respect to models using longevity pessimism instead of survival bias (Table 1.5).

Change in health is positively associated with a delay in pension choice of payoff period. Subjects who are less risk-averse (higher *CRRA* score) delay their pension choice. Higher patience in receiving monetary compensation is also associated with delayed pension choice. *Financial training* and *knowledge score* are not relevant casual factors driving pension choice.

1.4.4 Determinants of Happiness and Satisfaction

The *happiness and satisfaction* index itself is a factor of a principal components analysis on other measures or overall happiness and satisfaction with present life, life history, finances, and health.

Table 1.10 – Additional Analysis on Alternative Measures of Survival Bias. OLS regressions of pension choice (payoff period) on different survival bias measures (accumulated up to target age 80) of longevity beliefs about oneself: from actual (life table) probabilities (1), from average Swiss person of same age and gender (2); from the belief of family and friends about the subject's longevity (3). See Table 1.5 for the definition of other variables.

dep. variable: pension choice *measure of longev. belief bias:*	(1) $q^{own}_{i,t=80}$	(2) $q^{own:pop}_{i,t=80}$	(3) $q^{own:fam}_{i,t=80}$
survival belief bias	−0.281	−0.267	−0.401**
	[0.194]	[0.258]	[0.191]
treatment: Pessimistic	−0.454*	−0.447*	−0.535*
	[0.271]	[0.267]	[0.281]
treatment: Lump-sum	1.523***	1.560***	1.441***
	[0.267]	[0.264]	[0.277]
treatment: Reverse	−0.413	−0.367	−0.593
	[0.393]	[0.386]	[0.399]
Pessimistic × *survival belief*	−0.005	0.121	0.128
	[0.250]	[0.345]	[0.277]
Lump-sum × *survival belief*	0.105	0.251	0.245
	[0.237]	[0.304]	[0.250]
Reverse × *survival belief*	0.328	0.381	0.569
	[0.353]	[0.540]	[0.391]
gender: male	0.297	0.339	0.300
	[0.222]	[0.221]	[0.220]
age	0.020**	0.021***	0.022***
	[0.008]	[0.008]	[0.008]
education	0.121	0.107	0.113
	[0.084]	[0.084]	[0.084]
happiness and satisfaction	−0.016	0.007	0.006
	[0.061]	[0.060]	[0.059]
change in health	0.167***	0.182***	0.183***
	[0.059]	[0.059]	[0.058]
financial training: yes	−0.289	−0.265	−0.258
	[0.230]	[0.229]	[0.228]
knowledge score	−0.103	−0.113*	−0.103
	[0.063]	[0.063]	[0.063]
CRRA	0.301***	0.313***	0.311***
	[0.057]	[0.056]	[0.054]
patience	0.151***	0.151***	0.152***
	[0.056]	[0.056]	[0.056]
constant	5.516***	5.417***	5.455***
	[0.650]	[0.645]	[0.646]
Adjusted R^2	0.084	0.081	0.085
Observations	1276	1276	1276

Heterokedasticity-robust errors in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The index itself captures 70.9% of the variance of its components.

In Table 1.11, I examine how the relationship of this index with other personal characteristics. Change in health has a positive and significant effect on the happiness and satisfaction index. Age

Table 1.11 – Determinants of Happiness and Satisfaction. OLS regressions for *happiness and satisfaction* index (the first PCA factor on 5 measures of overall happiness and satisfaction with present life, life history, health and finances). *Gender: male* and *financial training: yes* are indicators that equal one if for the respective categories, and zero otherwise.*Age* measured in years, *education* as levels 1-5 and *income* as levels 1-8. Recent (1yr.) *change in health* is measured on a scale 0-10. *Knowledge score* is the number of correct answers (0-8) on a financial literacy quiz and CRT, combined. *CRRA* is the risk-aversion coefficient from a power utility model extracted from the "bomb" risk elicitation task (BRET). *Patience* is the delay choice, in months, (0-4) of participant compensation in exchange of interest.

	(1)	(2)	(3)	(4)
change in health	0.307***	0.308***	0.303***	0.299***
	[0.026]	[0.026]	[0.027]	[0.027]
gender: male		−0.159	−0.151	−0.138
		[0.104]	[0.106]	[0.109]
age		0.019***	0.019***	0.020***
		[0.003]	[0.004]	[0.004]
income		0.168***	0.155***	0.152***
		[0.026]	[0.027]	[0.027]
education		0.137***	0.111**	0.113***
		[0.041]	[0.043]	[0.044]
financial training: yes			0.192*	0.196*
			[0.105]	[0.107]
knowledge score			0.015	0.012
			[0.029]	[0.030]
CRRA				−0.019
				[0.033]
patience				0.017
				[0.028]
constant	−1.762***	−3.558***	−3.524***	−3.528***
	[0.173]	[0.263]	[0.305]	[0.315]
R^2	0.106	0.195	0.192	0.187
N	1320	1184	1168	1146

Heterokedasticity-robust standard in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

has a significant but small effect. Income and education have sizable and significant positive effects on happiness and satisfaction.

1.5 Discussion and Conclusion

In line with previous studies using samples from older adults, I also find that individuals, on average, underestimate their survival probabilities in relation to actuarial probabilities from life tables. Survival belief biases could arise from different sources. Individuals might be misinformed about longevity, in the sense that they lack the knowledge of proper human survival probabilities at different ages. A typical actuarial survival curve – skewed inverse S-shaped – presents especial cognitive challenges for individuals intuitively integrating its probability mass function. The marginal decrease in the one-year survival probability as a person ages (senescence) can impact the formation of subjective beliefs (Elder, 2013). General attitudes toward risky prospects, in particular probability weighting (Prelec, 1998; Tversky & Kahneman, 1992), could also play a role in how individuals make retirement financial decisions under stochastic survival risk.

Previous studies have shown, for example, that subjects *react* to longevity-impacting events both in terms of reported subjective probabilities (Hurd & McGarry, 2002; Apicella & De Giorgi, 2022) or observed financial behavior in administrative data (Kvaerner, 2022). Usually, such longevity shock-event models will allow for an idiosyncratic term that captures baseline survival over- or underestimation at the individual level. In the exploratory part of this study, the results suggest that longevity misinformation represents a substantial component of this idiosyncratic deviation between subjective and actual probabilities. This partially contradicts the conclusion of Post and Hanewald (2013) about how much of the dispersion of subjective longevity is not explained by awareness of individual longevity risk factors.

The results show that, in fact, subjects not only underestimate their own survival probabilities with respect to life tables (survival belief bias), but also severely underestimate the survival of an average person of their same age and gender (longevity misinformation). Subtle in principle, the distinction of the elicitation object (one's own probabilities or those of an average person) is important. Biases on survival beliefs about oneself can originate from private information on longevity risk factors, such as health status, risky behavior, or family history. However, one's beliefs about survival of an average person of a large population should not be impacted by any long-term risk factor or short-term shocks affecting one's own longevity. Also, any pessimism that is intrinsic to how one assesses his or her own longevity risk, relative to others, should still not affect his/her beliefs about longevity of an average person.

I show evidence of longevity pessimism, accumulated over the lifetime, is high at target ages typical of the first decade of retirement in contemporary societies, when retirees are mostly healthy and when year-on-year mortality probabilities are low (if non-negligible and increasing on age in the senescence dynamic).

To investigate the potential casual impact of longevity pessimism on financial decisions, I deploy a simulated task where subjects make a choice in terms of their pension benefit payoff. This pension choice task offers identical (except in one treatment condition) expected value payoffs in a stochastic survival environment. Its between-subject treatment concerns the pension payoff structure (*Fair* annuity, *Pessimistic* annuity, *Reverse* annuity or *Lump-sum*), and the decision on the timing (and thus survival risk) of payoffs as the outcome, instead of varying the payoff structure within-subjects over multiple rounds.

Because the experimental survival risk is resolved within a very short time, actual survival beliefs, even from a strict bounded-rationality perspective, should not have any impact on how subjects assess their stochastic termination risks in the experimental task. Yet, I find that longevity pessimism, to some extent, affects how much payoff risk subjects undertake. Recent health changes for subjects also affect their pension payoff choices, with those reporting recent health improvement delaying their chosen payoff period, i.e., shifting it to a more risky option. These results further suggest that private longevity information cannot account for the full deviation of reported subjective beliefs from actual unbiased probabilities, which could facilitate the contextualization of results of the previously cited contemporary studies that use shocks to longevity risk factors as an identification mechanism.

Finally, I explore whether longevity pessimism could be a mediator of the effects of happiness and satisfaction (which itself is influenced by recent health changes) on pension choice. Happiness and satisfaction do have significant – if moderate in size – indirect effects on pension choice through longevity beliefs. The presence of indirect effects suggests that the non-misinformation component of longevity pessimism could be related to general predispositions of the subject with respect to risk assessment in a broader sense (not only in the financial risk-taking domain).

In terms of the potential to improve financial decision making in the field, given these findings, longevity misinformation is a better candidate than longevity pessimism for financial literacy interventions (Behrman et al., 2012). Ex ante, the effects of wrong information on longevity could be mitigated with the provision of correct actionable information at the time of decision-making. This would be facilitated, in the field, by the fact that individuals make actual pension decisions infrequently (such as when choosing whether to withdraw lifetime pension savings as lump sum or convert them into annuities). Mitigation of possible longevity pessimism is more challenging: it concerns how people assess a very specific form of idiosyncratic risk (one's own longevity), for which misperceptions of actual risk factors might contribute (Heimer, Myrseth, & Schoenle, 2019).

These findings have implications for policymakers that consider implementing pension reforms that are, in principle, actuarially-neutral (as in Fatas, Lacomba, & Lagos, 2007). Actuarial neturality within changes to pension decision architecture drives policy expectations that subjects would react according to conditional survival probabilities of life tables, adjusting their decisions accordingly. However, my results corroborate concerns that biased subjective beliefs could produce biased decisions with important consequences for the long-term financial well-being of individuals during retirement.

Overall, I conclude that misinformation about human longevity is an important component of individual survival belief bias. In turn, longevity pessimism affects subjects' pension choice, driving them to make earlier and less risky pension payoff choices. There is also an indirect effect of happiness and satisfaction on these pension choices through its impact on longevity pessimism.

References

Ameriks, J., Caplin, A., Laufer, S., & Van Nieuwerburgh, S. (2011). The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives. *Journal of Finance*, *66*(2), 519–561. https://doi.org/10.1111/j.1540-6261.2010.01641.x

Apicella, G., & De Giorgi, E. G. (2022). A Behavioural Gap in Survival Beliefs. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3821595

Bachmann, K., Xu, X., Hens, T., & Lot, A. (2022). Experimental Research on Retirement Decision-Making: Evidence from Reproductions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4217816

Becker, C. K., & Trautmann, S. T. (2022). Does happiness increase in old age? Longitudinal evidence from 20 European countries. *Journal of Happiness Studies*, *23*(7), 3625–3654. https://doi.org/10.1007/s10902-022-00569-4

Behrman, J. R., Mitchell, O. S., Soo, C. K., & Bravo, D. (2012). How Financial Literacy Affects Household Wealth Accumulation. *American Economic Review*, *102*(3), 300–304. https://doi.org/10.1257/aer.102.3.300

Bell, D. N., Comerford, D. A., & Douglas, E. (2020). How do subjective life expectancies compare with mortality tables? Similarities and differences in three national samples. *Journal of the Economics of Ageing*, *16*, 100241. https://doi.org/10.1016/j.jeoa.2020.100241

Bissonnette, L., & de Bresser, J. (2018). Eliciting Subjective Survival Curves: Lessons from Partial Identification. *Journal of Business and Economic Statistics*, *36*(3), 505–515. https://doi.org/10.1080/07350015.2016.1213635

Bissonnette, L., Hurd, M. D., & Michaud, P.-C. (2017). Individual survival curves comparing subjective and observed mortality risks. *Health Economics*, *26*(12), e285–e303. https://doi.org/10.1002/hec.3506

Breen, R., Karlson, K. B., & Holm, A. (2013). Total, Direct, and Indirect Effects in Logit and Probit Models. *Sociological Methods and Research*, *42*(2), 164–191. https://doi.org/10.1177/0049124113494572

Breen, R., Karlson, K. B., & Holm, A. (2021). A Note on a Reformulation of the KHB Method. *Sociological Methods and Research*, *50*(2), 901–912. https://doi.org/10.1177/0049124118789717

Browning, M., & Crossley, T. F. (2001). The Life-Cycle Model of Consumption and Saving. *Journal of Economic Perspectives*, *15*(3), 3–22. https://doi.org/10.1257/jep.15.3.3

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. https://doi.org/10.1016/J.JBEF.2015.12.001

Cocco, J. F., & Gomes, F. J. (2012). Longevity risk, retirement savings, and financial innovation. *Journal of Financial Economics*, *103*(3), 507–529. https://doi.org/10.1016/j.jfineco.2011.10.002

Crosetto, P., & Filippin, A. (2013). The “Bomb” Risk Elicitation Task. *Journal of Risk and Uncertainty*, *47*(1), 31–65. https://doi.org/10.1007/s11166-013-9170-z

de Bresser, J. (2019). Measuring subjective survival expectations – Do response scales matter? *Journal of Economic Behavior and Organization*, *165*, 136–156. https://doi.org/10.1016/j.jebo.2019.06.018

Dormont, B., Samson, A.-L., Fleurbaey, M., Luchini, S., & Schokkaert, E. (2018). Individual Uncertainty About Longevity. *Demography*, *55*(5), 1829–1854. https://doi.org/10.1007/s13524-018-0713-4

Elder, T. E. (2013). The Predictive Validity of Subjective Mortality Expectations: Evidence From the Health and Retirement Study. *Demography*, *50*(2), 569–589. https://doi.org/10.1007/s13524-012-0164-2

Fatas, E., Lacomba, J. A., & Lagos, F. (2007). An experimental test on retirement decisions. *Economic Inquiry*, *45*(3), 602–614. https://doi.org/10.1111/j.1465-7295.2007.00027.x

FORS - Swiss Centre of Expertise in the Social Sciences. (2022). SHP Group, Living in Switzerland waves 1-22 + Beta version wave 23 + Covid 19 data [Dataset]. https://doi.org/doi.org/10.48573/vaxa-3w42

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, *40*(2), 402–435. https://doi.org/10.1257/002205102320161320

Gimenez, G., Gil-Lacruz, A. I., & Gil-Lacruz, M. (2021). Is Happiness Linked to Subjective Life Expectancy? A Study of Chilean Senior Citizens. *Mathematics*, *9*(17), 2050. https://doi.org/10.3390/math9172050

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645. https://doi.org/10.1162/003355397555217

Groneck, M., Ludwig, A., & Zimper, A. (2016). A life-cycle model with ambiguous survival beliefs. *Journal of Economic Theory*, *162*, 137–180. https://doi.org/10.1016/j.jet.2015.12.008

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. https://doi.org/10.1111/jofi.12828

Holt, C. A., & Laury, S. K. (2005). Risk Aversion and Incentive Effects: New Data without Order Effects. *American Economic Review*, *95*(3), 902–904. https://doi.org/10.1257/0002828054201459

Holzmeister, F., & Pfurtscheller, A. (2016). oTree: The "bomb" Risk Elicitation Task. *Journal of Behavioral and Experimental Finance*, *10*, 105–108. https://doi.org/10.1016/j.jbef.2016.03.004

Hurd, M. D., & McGarry, K. (2002). The Predictive Validity of Subjective Probabilities of Survival. *Economic Journal*, *112*(482), 966–985. https://doi.org/10.1111/1468-0297.00065

Hurwitz, A., & Sade, O. (2020). An Investigation of Time Preferences, Life Expectancy, and Annuity versus Lump Sum Choices: Can Smoking Harm Long-Term Saving Decisions? *Journal of Economic Behavior and Organization*, *180*, 812–825. https://doi.org/10.1016/j.jebo.2019.05.011

Imbens, G. W., & Manski, C. F. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica*, *72*(6), 1845–1857. https://doi.org/10.1111/j.1468-0262.2004.00555.x

Inkmann, J., Lopes, P., & Michaelides, A. (2011). How Deep Is the Annuity Market Participation Puzzle? *Review of Financial Studies*, *24*(1), 279–319. https://doi.org/10.1093/rfs/hhq080

Kleinjans, K. J., & Soest, A. V. (2014). Rounding, focal point answers and nonresponse to subjective probability questions. *Journal of Applied Econometrics*, *29*(4), 567–585. https://doi.org/10.1002/jae.2321

Kvaerner, J. (2022). How Large Are Bequest Motives? Evidence Based on Shocks to Mortality. *SSRN Electronic Journal*, (November). https://doi.org/10.2139/ssrn.2985465

Lugilde, A., Bande, R., & Riveiro, D. (2019). Precautionary Saving: a Review of the Empirical Literature. *Journal of Economic Surveys*, *33*(2), 481–515. https://doi.org/10.1111/joes.12284

Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, *52*(1), 5–44. https://doi.org/10.1257/jel.52.1.5

Max Planck Institute for Demographic Research, University of California, & French Institute for Demographic Studies. (2022). Human Mortality Database. www.mortality.org

Payne, J. W., Sagara, N., Shu, S. B., Appelt, K. C., & Johnson, E. J. (2013). Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect. *Journal of Risk and Uncertainty*, *46*(1), 27–50. https://doi.org/10.1007/s11166-012-9158-0

Peijnenburg, K., Nijman, T., & Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, *70*, 18–35. https://doi.org/10.1016/j.jedc.2016.05.023

Perozek, M. (2008). Using subjective expectations to forecast longevity: do survey respondents know something we don't know? *Demography*, *45*(1), 95–113. https://doi.org/10.1353/dem.2008.0010

Post, T., & Hanewald, K. (2013). Longevity risk, subjective survival expectations, and individual saving behavior. *Journal of Economic Behavior and Organization*, *86*, 200–220. https://doi.org/10.1016/j.jebo.2012.11.012

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497. https://doi.org/10.2307/2998573

Rieger, M. O., Wang, M., & Hens, T. (2017). Estimating cumulative prospect theory parameters from an international survey. *Theory and Decision*, *82*(4), 567–596. https://doi.org/https://doi.org/10.1007/s11238-016-9582-8

Skinner, J., & Hubbard, R. (1994). Expanding the life-cycle model: Precautionary saving and public policy. *American Economic Review*, *84*(2), 174–179. https://doi.org/10.2307/2117824

Sortheix, F. M., & Lönnqvist, J.-E. (2014). Personal Value Priorities and Life Satisfaction in Europe. *Journal of Cross-Cultural Psychology*, *45*(2), 282–299. https://doi.org/10.1177/0022022113504621

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an Alternate Form of the Cognitive Reflection Test. *Judgment and Decision Making*, *11*(1), 99–113. https://doaj.org/article/e596f711e83b42d68afbee3e1faff075

Tversky, A., & Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/Doi10.1007/Bf00122574

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. https://doi.org/10.1016/j.jedc.2015.03.003

Yaari, M. E. . (1965). Uncertain Lifetime , Life Insurance, and the Theory of the Consumer. *Review of Economic Studies*, *32*(2), 137–150. http://www.jstor.org/stable/2296058

Chapter 2

Experimental Research on Retirement Decision-Making: Evidence from Replications

Andre Lot **Kremena Bachmann**[1] **Xiaogeng Xu**[2] **Thorsten Hens**[3]

We adapt the design of four experimental studies on retirement decision-making and conduct replications with a larger online sample from the broader population. We replicate most of the main effects of the original studies. In particular, we confirm that consumption decisions are less efficient when subjects need to borrow from the future than when they need to save from the present. When subjects collect retirement benefits as lump sum instead of annuities, they choose to retire later, as suggested by the original study. We also confirm that savings are higher when they are incentivized with matching contributions than when incentivized with tax rebates. However, when faced with varying survival risks, subjects in our replication make only partial adjustments to spending paths when ambiguity is reduced. We also propose a further experimental research agenda in related topics and discuss practical issues on subject recruitment, attrition, and redesign of complex tasks.[4]

JEL Classification: C91, D15, G51, J26 | **Online Repository** [click] with supplementary materials.

Keywords: retirement, savings, annuities, life-cycle optimization, income smoothing, experiments, replications

[1]Zürich University of Applied Sciences, University Zürich

[2]Hanken School of Economics

[3]University Zürich, NHH – Norwegian School of Economics, University Luzern

[4]This paper has been accepted for publication at the Journal of Banking and Finance at the time of submission of this thesis. We are grateful for the comments and suggestions of three anonymous referees, the special issue editors (Michael Kirchler and Utz Weitzel); Christian König gen. Kersting, Thomas de Haan, and Alexander Ziegler; and the audiences at Financial Economics Colloquium at University of Zurich (2021), Experimental Finance conference in Bonn (2022), and FAIR Midway conference in Tromsø (2022).

2.1 Introduction

Retirement financial decisions over the life cycle exhibit puzzling patterns in the field, such as subjects not converting savings into life annuities, saving too little before retirement, or 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. The long spans between decisions and observable outcomes, as well as a low decision frequency, limit the ability to learn from the own experience. Normative institutional settings and strong social norms around these decisions impose further challenges for researchers seeking to identify the underlying drivers of observed behaviour.

Experimental studies on retirement decisions have addressed the empirical challenges associated with these decisions in the field. However, many of these studies have relied on student samples that do not vary with respect to characteristics that can be related to the studied treatment effects. In general, such homogeneity with respect to those characteristics may hinder conclusions on whether the observed causal relationship has external validity.

In the context of retirement decision-making, using student samples can be problematic since students are more likely than individuals from the general population to use hyperbolic discounting (Carbone, 2006), which can cause differences in the behaviour when dealing with life cycle optimization problems. Higher cognitive abilities within student samples could also conceal the limitations faced by the representative agent in the population making retirement decisions motivated by myopic planning (Ballinger et al., 2011). Students' lack of experience with long-term debt management could also plausibly explain certain suboptimal life-cycle optimization results observed in student samples, such as those in Meissner (2016). Such individual-level characteristics of different samples can potentially moderate the treatment effects in studies on retirement decision-making, which calls the generalizability of the reported causal effects into question.

This paper aims to evaluate the external validity of some main findings in the experimental literature on retirement decision-making by using online samples from the general population. We selected four experimental studies addressing different aspects of the retirement decision-making problem, in which the observed effects can potentially depend on individual characteristics such as experience with specific decisions and general ability to deal with complex decision problems. By using samples from the general population that differ with respect to these characteristics than the samples used in the original studies, we evaluate which findings of the original studies can be replicated. All our replication studies have been preregistered.

We successfully replicated most of the main effects of the selected experimental studies. In particular, we found that although subjects behaved less optimally than the subjects in the original study, their decisions were still better when they needed to save for the future than when they needed to borrow from the future (as in Meissner, 2016). In addition, we find that the impact of this debt aversion remained after considering individual differences in patience and risk preferences. In the face of survival risk, the subjects in our replication were also more likely to delay the timing of retirement when collecting benefits as lump sum than as annuities (like those in Fatas, Lacomba, & Lagos, 2007). In addition, we find that these timing decisions were affected by the survival risk that subjects experienced in previous rounds of the experiment. When incentives to save were offered as matching contributions rather than tax rebates, the effective savings rates were higher (in line with the observations of Blaufus & Milde, 2021). Finally, when facing varying survival risk, subjects in our sample adjusted their spending (as observed by Anderhub et al., 2000). However, in our sample, the response of spending to changes in the resolution of ambiguity of varying survival risk was insufficient and weaker than in the original study.

In addition to testing the replicability of the original studies, we document evidence of substantial suboptimal decision-making behaviour. Subjects consistently under-consumed their lifetime income, or consistently did not save enough, going bankrupt when needing to fund mandatory expenses. Such inefficiencies remain hidden in experiments with enforced lifetime budgets.

At last, we present and discuss some important methodological challenges and practical issues concerning the modification of original tasks, the implementation of such experiments with online panels from the general population, and the efficiency of decision-making within the tasks. We then 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 replicating the main effects of several experimental studies on retirement decision-making using larger and more heterogeneous samples than the original studies, our paper mainly contributes to the discussion of whether the experimental findings on this topic are externally valid. Although student samples can be generally criticized as they are on average more homogeneous than non-student samples (Peterson, 2001) and show different personal and attitudinal characteristics (Hanel & Vione, 2016), research has been sensitive enough to note that the usefulness of a sample should be judged upon having variance on relevant moderators (Druckman & Kam, 2009). The usefulness of student samples has been studied in various areas of research. In political science research, Krupnikov and Levine (2014) found that both student and diverse national adult samples behave consistently and in line with theoretical predictions once relevant moderators are taken into account. In economics, Horton, Rand, and Zeckhauser (2011) found that the main effects of common experiments in economics (such as prisoner's dilemma, priming, and framing effect in risk-taking) also hold true among Amazon Mechanical Turk workers. In retirement decision-making, Carbone (2005) found that differences in demographic characteristics do not affect the strategies used to solve the life-cycle optimization problem. However, Carbone (2006) found that people from the general population have a shorter planning horizon than students, and students are more likely to discount hyperbolically. Our study contributes to this discussion by showing that such differences between different samples have only a limited impact on the main effects of experimental research studying the behaviour in the context of financial retirement decision-making.

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. Previous research has shown that the complexity of decisions can motivate myopic planning (Ballinger et al., 2011) or the use of heuristics, which could potentially lead to suboptimal decisions. With respect to the implications of heuristics, Winter, Schlafmann, and Rodepeter (2012) found that the outcome of such heuristics does not need to be different from the outcomes of the underlying life-cycle dynamic optimization problems. Our research contributes to this discussion by showing that the complexity of the decision problem may lead to suboptimal behaviour, as it can potentially motivate decisions that are not sensitive enough to changes in the characteristics of the decision problem.

Finally, our findings have implications for policymakers who consider pension reforms that allow more discretion in retirement decisions or relax compulsory mandates. In this context, policymakers often assume that individuals would make retirement financial decisions in line with their individual preferences and economic constraints. Our results suggest that the financial retirement planning might be too complex for individuals to respond optimally to changes in the decision environment, and the suboptimal decision behaviour may impose a restriction on the efficacy of policy reforms.

In Section 2.2, we present an overview of the relevant experimental literature. Then, in Section 2.3, we introduce the original studies and present the results of our replications. We discuss the implications of our results and propose a future research agenda on this topic in Section 2.4, and conclude in Section 2.5. Additional experimental materials, original data and analysis code are available in the Online Repository.

2.2 Experimental Literature on Retirement Decision Making

Experiments on individual retirement decision making have investigated 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 behaviour. In the second subsection, we then discuss in more detail the most common experimental task features that distinguish experiments in this domain. Table 2.1 summarizes the studies in terms of their main findings and distinguishing features with respect to the experimental design.

Table 2.1 – An Overview of Experimental Studies on Retirement Decision-Making

Study	Focus	Main Dependant Variable(s)	Main Finding(s)	N	MP	LR	InU	ME	IoS	ELB	SWRP	IU	Sample
Agnew, Anderson, and Szykman (2015)	behavioural 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.**	**100**	**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)
Besheats 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)	**behavioural biases (framing)**	**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)
A. L. Brown, Chua, and Camerer (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)
J. R. 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)	behavioural biases	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)**

– *Continued on next page*

Table 1. An Overview of Experimental Studies on Retirement Decision-Making – *Continued from previous page*

Study	**Focus**	**Main Dependant Variable(s)**	**Main Finding(s)**	**N**	**MP**	**LR**	**InU**	**ME**	**IoS**	**ELB**	**SWRP**	**IU**	**Sample**
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)	behavioural biases	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 behaviour differs significantly from optimal behaviour; the comparative static implications of actual behaviour 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 savings.**	**149**	**X**			**X**		**X**	**X**		**M-Turk**
Levy and Tasoff (2020)	biases	overconsumption	Observed behaviour consistent with behaviour 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)

Notes: 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 rows in bold font are the five studies that are included in this replication study.

2.2.1 Drivers of Retirement Decision Making Behaviour

One strand of experimental studies investigates how specific features of the decision problem affect people's decision behaviour. Carbone and Hey (2004) investigated how people adjust their consumption behaviour to the possibility of unemployment, and found that people overreact to the risk of unemployment. In a study, in which they varied the length of the retirement phase, Koehler, Langstaff, and Liu (2015) found that most participants responded sensibly by saving more of their current income when faced with a long compared to a short retirement phase. Meissner (2016) studied optimal consumption on an increasing and decreasing income path and found that when people are required to borrow to smooth consumption (i.e., when their income path is increasing), deviations from optimal behaviour are more likely. Anderhub et al. (2000) relaxed the assumption in most experiments that the survival probabilities are constant and found that the average subject reacts in a qualitatively correct way to "good" and "bad" news concerning survival risk. While most studies have considered decisions under income distribution risk, Carbone and Infante (2014) studied decision making under risk and ambiguity and found that behaviour under ambiguity is characterized by a significant pattern of under-consumption compared to behaviour under risk. In terms of the quality of the general decision behaviour of the subjects, Hey and Dardanoni (1988) found 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) studied the introduction of automatic savings schemes and found that individuals save less with such schemes, but the reduction is only partial in that the total lifetime consumption measures are higher. Duffy and Li (2019) considered different pension replacement rates and found that subjects achieve the highest experimental payoff when offered a constant life-cycle endowment profile (100% pension replacement rate). Hurwitz, Sade, and Winter (2020) investigated the benefits of implementing a minimum annuity rule and found that this does not guarantee an increase in the demand for annuities, and may even reduce it. Beshears et al. (2020) evaluated the benefits of introducing higher withdrawal penalties in retirement savings schemes and found that higher early withdrawal penalties attract more commitment account deposits. Fatas, Lacomba, and Lagos (2007) examined whether the pension benefits scheme (lump-sum payments or annuities) affects retirement decisions in the face of longevity risk and found that concentrating payments (shifting from annuity to 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. A. L. Brown, Chua, and Camerer (2009) found 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) found 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) found 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) allowed for interpersonal comparison and found that providing this information leads to worse outcomes in the form of more under-saving and lower money earnings. In contrast, Ballinger, Palumbo, and Wilcox (2003) analyzed learning effects using an intergenerational structure and found that subsequent generations perform significantly better in terms of savings than previous generations.

Few studies have analyzed the effect of specific behavioural biases on retirement financial decisions. Levy and Tasoff (2020) found that the subjects' decision behaviour is affected by the exponential growth bias. Agnew et al. (2008) found that an excessive extrapolation of the past performance of the financial market influences the demand for annuities. Blaufus and Milde (2021) found that different frames of tax-related pension incentives can influence retirement savings, while J. R. Brown et al. (2008) also found that the use of different frames can affect the demand for annuities. Several experiments have reported evidence that subjects behave myopically (Carbone & Hey, 2004; Ballinger, Palumbo, & Wilcox, 2003; Carbone, 2005, 2006) and have dynamically inconsistent preferences (A. L. Brown, Chua, & Camerer, 2009). In terms of general decision-making behaviour, Carbone (2005) found that subjects apply common rules of thumb to solve the optimization problem. Subjects also exhibit preferences for building wealth, even if it is not optimal to do so (Gechert & Siebert, 2020).

Finally, some studies have aimed to explain the heterogeneity in behaviour based on dynamic decision-making tasks. Ballinger et al. (2011) found that cognitive abilities (but not personality measures) are good predictors of heterogeneity in saving behaviour observed as a result of using shorter than optimal planning horizons. Carbone (2005) concluded 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) found that hyperbolic discounting affects the behaviour 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.

2.2.2 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 facing the subjects (Agnew, Anderson, & Szykman, 2015; Anderhub et al., 2000; Fatas, Lacomba, & Lagos, 2007; Hey & Dardanoni, 1988).

Another source of uncertainty in the optimization problem that can be introduced is stochastic income. This type of uncertainty can be used in different ways. It can be linked to the probability of becoming unemployed or later re-employed (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 analysing 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 have used deterministic income paths (e.g., Duffy & Li, 2019).

In some experiments, subjects are required to cover some mandatory expenses during the simulated life-cycle in order to incentivize savings (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 studies reviewed, savings were 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 lead to suboptimal decisions.

Some studies have introduced 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, meaning that 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 have enforced a lifetime budget, whereby any wealth left at the last period is automatically spent (Blaufus & Milde, 2021; Bohr, Holt, & Schubert, 2019; A. L. Brown, Chua, & Camerer, 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 suboptimal behaviour or misunderstanding of the experimental tasks.

Finally, to motivate subjects to optimize their consumption paths, most studies have linked subjects' consumption choices to their payoffs. Some studies have specified the link by using a particular (induced) utility function. When there is no interest earned on savings, and payoffs are based on lifetime outcomes, inducing a utility function is essential. Otherwise, subjects might just assign most of their lifetime consumption to 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 the hope that this period is eventually selected for payoff.

2.3 Replications of Adapted Experimental Designs

Taking into account the existing body of previous experimental studies on retirement decision-making (see Table 2.1), we selected four experiments that spanned a heterogeneous set of research topics and experimental design features. In terms of research topics, we selected two studies investigating the impact of different decision characteristics, such as ambiguous survival probabilities (Anderhub et al., 2000), and different income paths (Meissner, 2016) on the consumption behaviour over time. Dealing with these characteristics of the decision problem requires a certain level of cognitive abilities and experiences, such as experience with debt management. The variability with respect to these characteristics is usually low in traditional student samples. The third study evaluated the relevance of institutional features related to the design of retirement benefits on the decision when to retire in the presence of survival risk (Fatas, Lacomba, & Lagos, 2007). The decision problem requires dealing with probabilities of survival, which could be a cognitively demanding task, with important implications for policymakers designing the form of retirement benefits. The fourth study addressed the relevance of behavioural effects, and specifically framing effects, on the decision of how much to save for retirement (Blaufus & Milde, 2021). Depending on the task, older people might not be subject to framing effects as observed by Pu, Peng, and Xia (2017).

The selected experiments also differ with respect to the task features summarized in Table 2.1. We consider diversity in the task features as a selection criterion because these features might cause different levels of inefficiencies in decisions between the original samples and our replications. These features also correspond to the many flavours of life-cycle models (for an overview, see Browning & Crossley, 2001), and could not be plausibly investigated in a single experimental study that simultaneously considers all these decision features using a single parsimonious model.

Finally, our selection of the studies was motivated by the technical feasibility (or lack thereof) of certain experimental designs features using online unassisted samples. Under this consideration, some experimental designs, such as the design used by A. L. Brown, Chua, and Camerer (2009), could not be deployed.

Subjects for all replications were recruited from the Germany recruitment pool of the market research company Bilendi. Since this pool of subjects is not very well known among experimental researchers, we also replicated the study of Koehler, Langstaff, and Liu (2015), which uses a simplified retirement decision-making task with Amazon Mechanical Turk workers to evaluate how income availability over the life-cycle affects consumption behaviour. The main goal of this replication is to see whether our online pool of subjects from the general population can manage such experimental tasks, and whether they respond to financial incentives, which we introduced in addition and which were part of the other replications.

Replications and, in some cases, additional analyses of individual experiments were pre-registered on AsPredicted.[5] Each of these studies addresses a different research question; hence, we do not propose any joint analysis of individual replication results with respect to their original hypotheses. While discussing some replications, in light of the results we found, we offer some additional non-preregistered analyses that are clearly noted as such.

In the replications, we focused on one or two main effects of each study. We intended to replicate the studies using subjects from the general population, who would perform the tasks online without any assistance from experimenters at hand. For this purpose, we modified the original experimental designs and adjusted their tasks as needed. We drew the subjects from the same large pool, and used the same deployment method, quality control mechanism, and common design and interface features in all replications to avoid differences in the results between the studies arising from such differences in the implementation.

In addition to replicating the main effects of the original studies, in the Appendix we present (non-preregistered) analyses of the main effects broken down by subsamples based on socio-demographic characteristics of the subjects (age, gender, income, education, and financial training).

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

The main effects are not always statistically significant in all subsamples, and the significance of the main effects across the subsamples differs between the studies. However, across all studies, the main effects held true in the three subsamples: the subjects who are older than 50 years, those who do not have higher education, and those who had not participated in any financial training. The subsamples with these characteristics clearly do not overlap with the student subsamples used in the original studies.

In the following subsections, we first discuss the approach and procedures we used to modify and adjust the experimental designs and their tasks, and the general engagement and performance metrics of subject participation. We then discuss the specific replication 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.

2.3.1 Redesign and Adaptation of Experimental Tasks

The original experimental sessions included extensive subject education and training. In addition, some experiments had a very complex set of instructions, including direct mathematical formulae presented to subjects to explain the induced utility and complex payoff mechanisms. These features of the original studies would make any attempt to closely replicate all the original experiments unfeasible. To address this challenge, while aiming to preserve the main mechanisms we wanted to replicate, we modified and redesigned the experimental tasks to varying degrees.

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 as much of an issue in our replications as it would have been in experiments that rely on stochastic environmental variables that persist over many periods (such as in the first task of A. L. Brown, Chua, & Camerer, 2009).[6]

Three experiments originally used numerous 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 replications, 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).[7] 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 text fields. Changing the decision slider(s) would also reveal the simple accounting mechanics on savings and cash balances, where relevant, and 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, substantially reducing 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 task. Although at the beginning of the session we showed the instructions and asked subjects to read them, we let the subjects 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 replaced explicit complex mathematical formulae (such as the induced utility in Meissner, 2016) with graphs that showed, more intuitively, the relevant functional

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

[7] All screenshots for all treatments of the replications are available in the online repository.

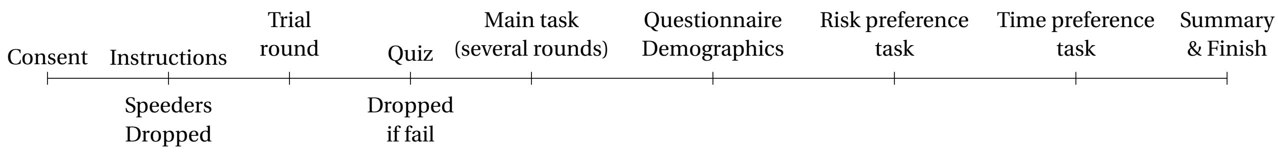


Figure 2.1 – Sequence of the steps for each replication.

relationship between variables.

The session flow in all replications is illustrated in Figure 2.1. Once the subjects completed reading the instructions, they started a trial round.[8] This allowed them to learn by doing the main experimental task, with ready access to the instructions in tabs at the bottom of the screen.[9] 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, gender, 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 elicited their time preferences (patience) as their willingness to delay their variable payoff by 1, 2 or 3 months for 5% monthly interest.[10] 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.

The experiments were deployed in German, which was the default interface language. Less than 2% of the subjects decided to use English, which was offered as an alternative language. The experiments were programmed in oTree (Chen, Schonger, & Wickens, 2016). Power analyses were computed with GPower (Erdfelder et al., 2009).

[8]In experiments adopting a within-subject treatment, the trial round was always identical to the treatment the subjects would undergo in the first live round.

[9]The trial round was not relevant for the payoff.

[10]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 being 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.

2.3.2 Subject Engagement, Quality Control, and Decision Efficiency

The experimental sessions were conducted in individual batches for each experiment between September 2021 and March 2022. A total of 6,213 subjects clicked on e-mail invitations sent[11] from the market research panel.[12]

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). During the quiz, the subjects were dropped if they answered more than two wrong questions on a first attempt or gave any wrong answer in a second attempt.[13] They were also automatically removed from the experiment if they did not finish the session more than 60 minutes after the quiz had been completed.[14]

Panel A of Table 2.2 details the attrition at each step for all the replications. The completion rate ranges from 21.5% to 50.6% of invitation clicks,[15] but the completion rate is not significantly different between the treatments within each replication. We tested the equality of completion rate between treatments for each replication with a proportion test (ANOVA analysis) if the experiment had 2 (3) treatment groups. The p-value is 0.26 for the replication of Anderhub et al. (2000), 0.07 for Fatas, Lacomba, and Lagos (2007), 0.73 for Koehler, Langstaff, and Liu (2015), 0.77 for Meissner (2016), and 0.81 for Blaufus and Milde (2021). The p-values remain the same when running logistic regressions and testing if the treatment indicators are equal to zero.

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

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

[13]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.

[14]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.

[15]Data collection for the reproduction of Blaufus and Milde (2021) was affected by 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.

Table 2.2 – Overview of Attrition, Payoff and Completion Time

Panel A: Subject Attrition

	Anderhub *et al.* (2000)		Fatas *et al.* (2007)		Koehler *et al.* (2015)		Meissner (2016)		Blaufus & Milde (2021)	
	Obs	%	Obs	%	Obs	%	Obs	%	Obs	%
No consent	89	9.3	80	7.6	64	5.5	77	6.0	130	7.5
Drop out at instructions	104	10.9	92	8.8	160	13.9	189	14.6	290	16.7
Drop out at trial round	85	8.9	45	4.3	101	8.8	87	6.7	149	8.6
Drop out / failed quiz	193	20.2	113	10.8	138	12.0	109	8.4	90	5.2
Drop out during tasks	146	15.3	187	17.9	347	30.1	554	42.8	556	32.0
Finished	339	35.5	530	50.6	344	29.8	278	21.5	522	30.1
Total	956	100.0	1047	100.0	1154	100.0	1294	100.0	1737	100.0

Notes: Subject participation according to their furthest stage reached per experiment. *Dropped out at instructions* include the subjects who were rejected for having gone through instruction screens too fast (10s to 60s threshold depending on experiment). *Dropped out at quiz* include the subjects who were rejected for failing to answer a quiz with five or six multiple-choice questions about the experimental instructions, after the trial round. The summary of attrition includes all the subjects who clicked the invitation link and landed on the first web-page of the experiment.

Panel B: Payoff (Euro) and Completion Time

	Anderhub *et al.* (2000)		Fatas *et al.* (2007)		Koehler *et al.* (2015)		Meissner (2016)		Blaufus & Milde (2021)	
	Payoff	Total time	Payoff	Total time	Payoff	Total time	Payoff	Total time	Payoff	Total time
Min	0.00	5.45	0.00	3.28	0.00	8.35	0.00	8.07	0.00	12.37
50^{th}-percentile	3.12	14.55	3.43	9.75	1.85	26.91	20.14	22.73	8.38	25.41
95^{th}-percentile	11.42	47.02	16.36	25.53	9.98	62.98	46.44	54.70	25.01	65.62
Max	27.20	14224.83	57.10	7483.52	24.97	2728.22	101.16	4668.42	74.89	2273.38
(Obs. > 65 min)		(12)		(5)		(14)		(8)		(26)

Notes: Payoff, in Euro, is the sum of variable incentive payoff for the main experiment and the payoffs of the risk-taking and patience tasks, and it does not include the non-variable fee of €4.76 for completing the experiment. *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. The summary of payoff and completion time includes only the subjects who completed the experiment.

Table 2.3 – Subject Characteristics and Treatment Assignments

	Anderhub *et al.* (2000)		Fatas *et al.* (2007)			Koehler *et al.* (2015)		Meissner (2016)		Blaufus & Milde (2021)		
	Product	Summation	Annuity	Combined	Lump	Long first	Short first	Borrow first	Save first	Deferred	Immediate	Matching
Age	44.57	48.53	48.37	49.59	48.98	41.92	41.05	43.64	42.56	47.60	48.98	48.88
	(1.39)	(1.61)	(1.09)	(1.19)	(1.06)	(1.32)	(1.20)	(1.44)	(1.60)	(1.20)	(1.11)	(1.12)
Observations	176	163	177	170	183	166	178	147	131	162	178	182
Test statistic	-1.88 ($p=0.06$)		0.30 ($p=0.74$)			0.48 ($p=0.63$)		0.50 ($p=0.62$)		0.44 ($p=0.65$)		
Female	0.42	0.52	0.42	0.45	0.51	0.56	0.49	0.56	0.49	0.41	0.40	0.38
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	176	163	177	170	182	165	178	144	130	162	178	182
Test statistic	-1.75 ($p=0.08$)		1.28 ($p=0.28$)			1.28 ($p=0.20$)		1.68 ($p=0.09$)		0.11 ($p=0.89$)		
Education	2.34	2.45	2.83	2.67	2.64	2.25	2.49	2.25	2.49	2.36	2.37	2.49
	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Observations	169	160	176	170	181	164	171	144	130	157	177	179
Test statistic	-1.12 ($p=0.26$)		1.92 ($p=0.15$)			**-2.25 ($p=0.03$)**		0.66 ($p=0.51$)		1.21 ($p=0.30$)		
Financial training	0.24	0.26	0.23	0.29	0.28	0.20	0.25	0.20	0.25	0.19	0.21	0.24
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Observations	171	159	177	168	181	163	176	143	129	157	175	179
Test statistic	-0.38 ($p=0.70$)		0.75 ($p=0.47$)			-1.18 ($p=0.24$)		-1.33 ($p=0.18$)		0.61 ($p=0.55$)		
Income level	6.54	6.64	7.40	7.26	7.45	6.72	6.94	6.72	6.94	7.28	7.40	7.37
	(0.24)	(0.25)	(0.21)	(0.24)	(0.22)	(0.24)	(0.24)	(0.24)	(0.24)	(0.23)	(0.22)	(0.21)
Observations	166	152	172	159	168	156	165	133	119	148	171	174
Test statistic	-0.28 ($p=0.78$)		0.19 ($p=0.82$)			-0.66 ($p=0.51$)		-1.51 ($p=0.13$)		0.07 ($p=0.93$)		
Risk-taking	35.52	31.94	31.37	31.98	29.65	24.60	27.42	24.60	27.42	27.66	26.17	26.55
	(1.77)	(1.51)	(1.67)	(1.50)	(1.59)	(1.44)	(1.48)	(1.44)	(1.48)	(1.46)	(1.37)	(1.47)
Observations	176	163	177	170	183	166	178	147	131	162	178	182
Test statistic	0.52 ($p=0.13$)		0.58 ($p=0.56$)			-1.36 ($p=0.17$)		-0.29 ($p=0.78$)		0.28 ($p=0.76$)		
Patience	2.64	2.51	2.69	2.75	2.40	2.66	2.85	2.66	2.85	2.70	2.80	2.73
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)
Observations	176	163	177	170	183	166	178	147	131	162	178	182
Test statistic	0.88 ($p=0.38$)		**3.80 ($p=0.02$)**			-1.37 ($p=0.17$)		1.28 ($p=0.20$)		0.31 ($p=0.73$)		

Notes: Standard errors are in parentheses. *Test statistic* is from the test that checks the equality of means between the treatments. For the replications with 2 treatment groups, *Test statistic* is t-statistic of t-test if the variable is non-binary and z-statistic of proportion test if the variable is binary. For the replications with 3 treatment groups, *Test statistic* is F-statistic of ANOVA analysis. The p-value is in the parenthesis. *Age* is in years old. *Female* is an indicator for female. *Education* equals 1 if the subjects have no qualification, 2 if vocational education, 3 if Bachelor degree, 4 if Master degree and 5 if Doctoral degree. *Financial training* is a dummy indicating that subjects state that they had participated in courses on financial decision making. *Income level* equals 1 if the monthly household disposable income is below €400, 2 if the income is between €400 and €800, and the value increases with the interval of €400 to 11 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 and equal to 1, 2, 3 and 4 for the choice of no delay, 1 month, 2 months and 3 months, respectively. The observations who chose to give no answer to the questions of gender, education, income or financial training are not included. The cells in bold and blue font are with $p < 0.05$.

The payment to a subject includes a payment for finishing the study and an incentive payoff based on the outcome of the replication.[16] Panel B of Table 2.2 summarizes the incentive component of the payoff of the subject and the completion time for the experiments. All replications could produce zero incentive payoff for the subjects, and the largest incentive payoff was €101.16. The panel also summarizes the completion time of the subjects who answered all questions.[17]

The randomization of subjects to treatment cells in all experiments seems satisfactory with respect to the demographics of the subjects, as seen in Table 2.3. For most characteristics and treatments of each experiment, there are no significant differences within each experiment at the 5% level, except a few stances. The means of *education* are different ($p = 0.03$) between the treatments of the replication of Koehler, Langstaff, and Liu (2015). ANOVA tests show that the means of the variable *patience* are different ($p = 0.02$) in the replication treatments of Fatas, Lacomba, and Lagos (2007).

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.[18]

Finally, we evaluated the effects of individual subject characteristics on their economic efficiency of decisions across the experiments, with results shown in Table 2.4.[19] Across four experiments,[20] *female* subjects made less efficient decisions than males, and such gender effect is only statistically significant in the replication of Koehler, Langstaff, and Liu (2015). In three experiments, higher *risk-taking* subjects performed significantly worse in most studies except the replication of Koehler, Langstaff, and Liu (2015).[21] The subjects who have participated in *financial training* performed better than those who have not, and the effect of financial training is statistically significant only in the replication of Koehler, Langstaff, and Liu (2015). These differences do not seem to arise from

[16] In addition to a variable incentive payoff, subjects who finished the experiment earned €4.76 for participating in the study.

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

[18] Even though the 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.

[19] This analysis was not preregistered.

[20] 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.

[21] 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.

different effect levels between the treatment assignments, as the specifications incorporate round × treatment fixed-effects for the relevant replicated experiments.

Table 2.4 – Effects of Individual Characteristics on the Efficiency of Decisions

	Anderhub *et al.* (2000)		Koehler *et al.* (2015)		Meissner (2016)		Blaufus & Milde (2021)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.000	-0.000	-0.006***	-0.006***	-0.000	-0.000	-0.300*	-0.291*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.123)	(0.121)
Female	0.091	0.092	-0.052*	-0.052*	-0.036	-0.024	1.572	1.510
	(0.049)	(0.049)	(0.023)	(0.023)	(0.019)	(0.019)	(4.037)	(4.019)
Education	0.041	0.042	0.033**	0.032**	0.009	0.010	-4.323	-3.863
	(0.029)	(0.029)	(0.012)	(0.012)	(0.010)	(0.010)	(2.211)	(2.184)
Financial training	-0.072	-0.072	0.058*	0.058*	0.031	0.018	-5.410	-4.945
	(0.056)	(0.056)	(0.028)	(0.028)	(0.024)	(0.023)	(4.710)	(4.698)
Income	-0.016	-0.016	-0.003	-0.003	-0.006*	-0.007*	-0.011	-0.028
	(0.008)	(0.008)	(0.004)	(0.004)	(0.003)	(0.003)	(0.696)	(0.684)
Risk-taking	0.003**	0.003*	-0.000	-0.000	-0.002***	-0.002***	0.416***	0.417***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.102)	(0.100)
Patience	-0.032	-0.033	0.040***	0.040***	-0.011	-0.011	1.583	1.556
	(0.019)	(0.019)	(0.008)	(0.008)	(0.008)	(0.007)	(1.394)	(1.392)
Constant	1.181***	1.175***					66.596***	54.145***
	(0.111)	(0.116)					(9.768)	(10.397)
Round/treatment FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	309	309	1252	1252	968	968	954	954

Notes: The dependent variables are the measurements of the efficiency of decisions: for Anderhub et al. (2000), it is the mean of deviations of the observed decisions from the optimal 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 mean of absolute deviations from the optimal saving of the periods in a round. The optimal saving is the saving that maximizes the expected payoff. Given that the periods have an equal chance to determine the final payoff, the optimal saving is same for each period and it is 74 points for treatment Immediate and Matching and 124 points for treatment Deferred. The results of the first and last columns are OLS estimations, and the results of the second and third columns are marginal effects of logistic regressions. *Age* is in years old. *Female* is an indicator for female. *Education* equals 1 if the subjects have no qualification, 2 if vocational education, 3 if Bachelor degree, 4 if Master degree and 5 if Doctoral degree. *Financial training* is a dummy indicating that subjects state that they had participated in courses on financial decision making. *Income level* equals 1 if the monthly household disposable income is below €400, 2 if the income is between €400 and €800, and the value increases with the interval of €400 to 11 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 and equal to 1, 2, 3 and 4 for the choice of no delay, 1 month, 2 months and 3 months, respectively. The round/treatment control covariates include the round number and the treatment dummies. The observations who chose to give no answer to the questions of gender, education, income or financial training are not included. The number of observations equals the number of subjects for the first study and the number of the decisions made by all subjects in all the rounds for the last three studies. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.3.3 Replication Results of Anderhub, Güth, Müller and Strobel (2000)

The experiment analysed how ambiguity (and its resolution) in the probability of survival affects consumption decisions over time. In the main task, subjects started a round – comprising four to six periods – facing three possible chances of being terminated in one period (1/6, 1/3, or 1/2).[22] In the first and second periods, subjects did not face termination while one of the probabilities was removed, reducing ambiguity until only one of the probabilities was left. Then, in the third, fourth, and fifth periods, subjects faced the termination probability that remained. Subjects made consumption decisions out of an initial endowment until they were terminated. Upon termination, any unspent amount from the initial endowment was lost. The round ended automatically after six periods if subjects were not terminated earlier. A subject went through six rounds, and each round had a different sequence of the three termination probabilities.

To see how the behaviour changes with different risk structures, the treatment conditions implement two different forms of the induced lifetime utility based on period consumption c for subject i at periods t from the first until termination period T. In the *Summation* condition, the payoff is given by the sum of the square root of period consumption $\left(U_i = \sum_{t=1}^{T} \sqrt{c_{i,t}}\right)$. In the *Product* condition, the payoff is given by the product of period consumption $\left(U_i = \prod_{t=1}^{T} c_{i,t}\right)$. The smoothing incentives are larger in the *Product* condition, since the expected payoff in that condition is substantially reduced if subjects spend all their endowment before termination (as one of the periods would have zero consumption and, thus, the lifetime utility for that round would be zero).

Analysis of the behaviour with respect to a risk-neutral optimal benchmark suggests two distinct behavioural patterns. First, subjects need to dynamically adjust their spending based on the sequential resolution of the ambiguity of the termination probabilities. When a larger or smaller termination probability is removed, the expected remaining length of the round of the subject increases or decreases, respectively; this should lead subjects to increase or decrease spending in the following period accordingly. Second, the consumption should monotonically decrease from the third period on.

[22] In our study, the different levels of termination risk were implemented using different distributions of colours in card decks (the original study used numerical ranges of a dice).

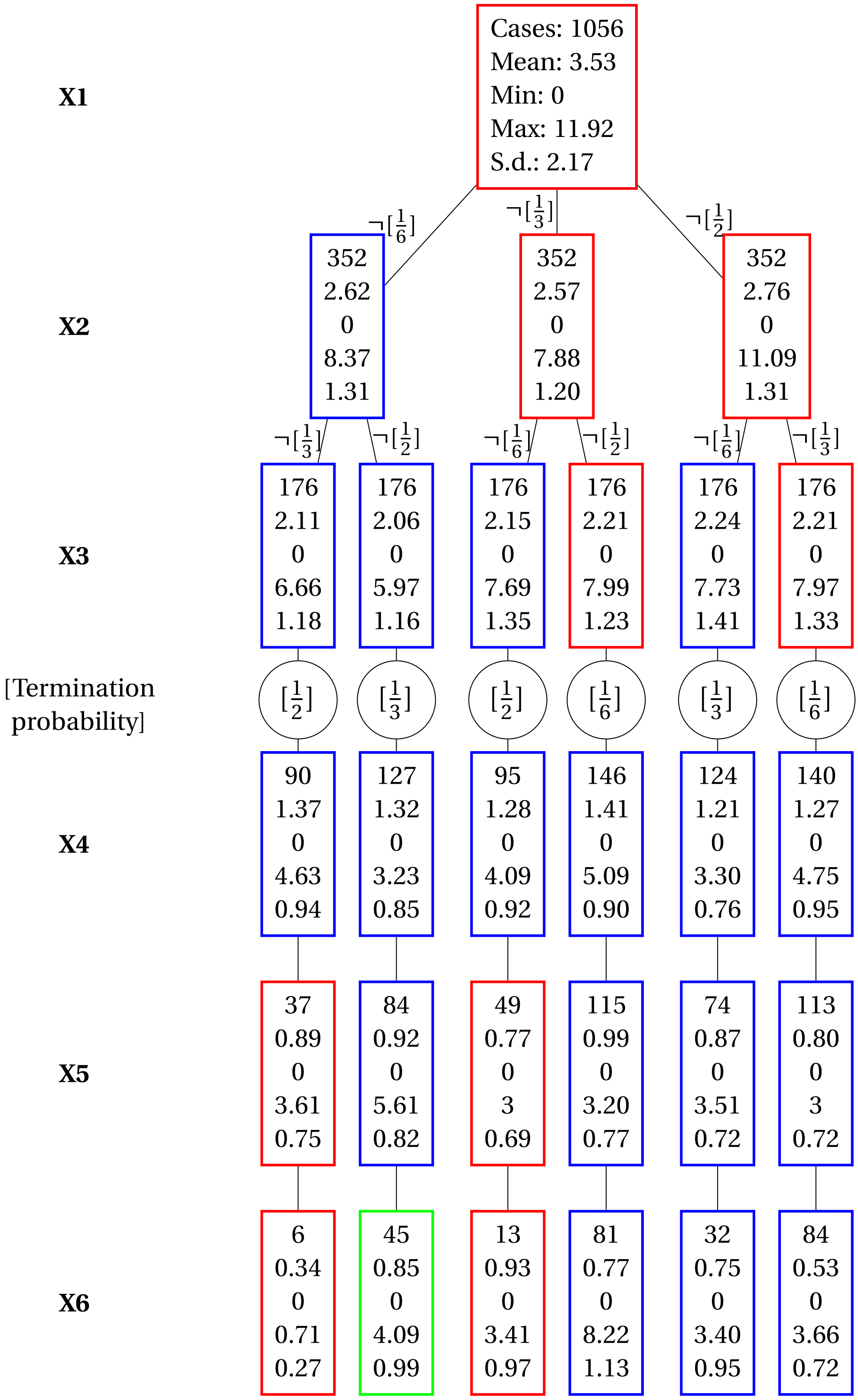


Figure 2.2 – Anderhub et al. (2000) Replication: Average Behaviour on Product Treatment.

Notes: The labels on the left X1 to X6 indicate the period of the spending decision for *Product* treatment. In each node box, the five values (from the top to the bottom) indicate the number of cases, mean, maximal, minimal, and standard deviation. $\neg[\frac{1}{2}]$ ($\neg[\frac{1}{3}]$, $\neg[\frac{1}{6}]$) indicates the card deck indicating termination probability $\frac{1}{2}$ ($\frac{1}{3}$, $\frac{1}{6}$) is removed. After Period 3 (X3), the finally stayed card deck determines the termination probability in the next periods and it can be $\frac{1}{2}$, $\frac{1}{3}$, or $\frac{1}{6}$. The color of the node box border indicates how the observed mean of spending decision is compared to the optimal spending paths: red means over-spending, blue means under-spending, and green means same as the optimal spending.

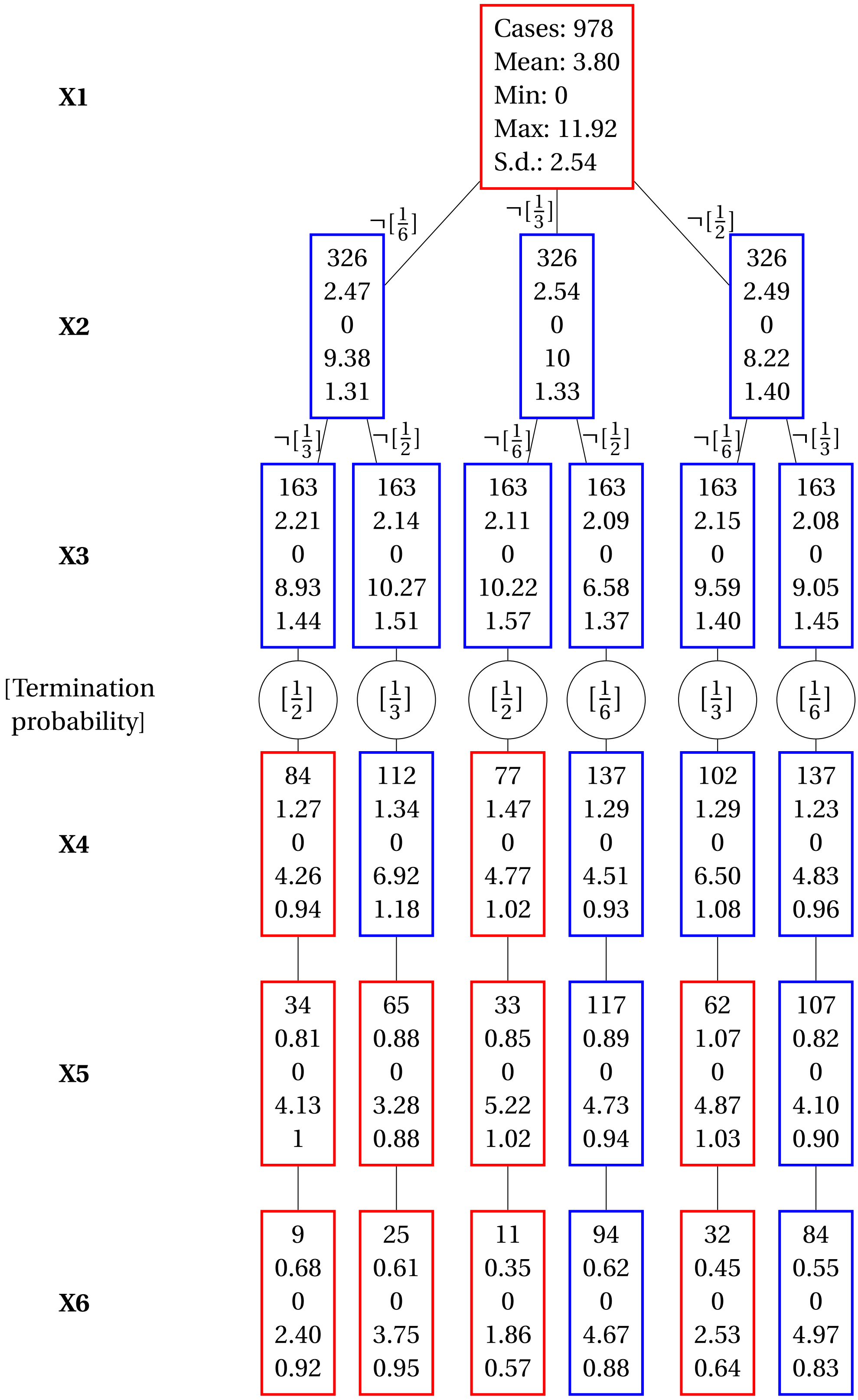


Figure 2.3 – Anderhub et al. (2000) Replication: Average Behaviour in Summation Treatment.

Notes: The labels on the left X1 to X6 indicate the period of the spending decision for *Summation* treatment. In each node box, the five values (from the top to the bottom) indicate the number of cases, mean, maximal, minimal, and standard deviation. $\neg[\frac{1}{2}]$ ($\neg[\frac{1}{3}]$, $\neg[\frac{1}{6}]$) indicates the card deck indicating termination probability $\frac{1}{2}$ ($\frac{1}{3}$, $\frac{1}{6}$) is removed. After Period 3 (X3), the finally stayed card deck determines the termination probability in the next periods and it can be $\frac{1}{2}$, $\frac{1}{3}$, or $\frac{1}{6}$. The color of the node box border indicates how the observed mean of spending decision is compared to the optimal spending paths: red means over-spending and blue means under-spending.

Table 2.5 – Anderhub et al. (2000) Replication: Consumption and Resolution of the Survival Ambiguity

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

Notes: 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 in the first and second period, which eventually eliminates ambiguity of the actual survival probabilities subjects will face (the remaining card deck being then used to determine survival after periods 3, 4 and 5). Sequences are ranked in descending order of optimal consumption fraction (in parentheses). *Product* and *Summation* are the treatments. ¬ means the removal of a card deck and the following fraction indicates the termination probability of the removed card deck. E.g., ¬[1/2] means the card deck with termination probability 1/2 is removed.

Subjects participated in six rounds (twelve in the original study), comprising all permutations (twice each in the original study) of the sequence of resolution of the uncertainty of termination risk (the order in which the card decks – with different termination probabilities – are removed). We used the same initial endowment for both treatments (11.92 ECU) as in the original study, and adjusted the exchange rates of lifetime utility-induced points into currency to account for the different conditional expectation of payoffs given the induced utility function of both treatments.

In our replication, 339 subjects completed the experiment: 176 in the treatment *Product* and 163 in *Summation* (in the original study, 50 subjects participated in each treatment). In the *Product* treatment, there were 1056 cases (176 subjects × 6 rounds) and the average reward was €3.33; in the *Summation* treatment, there were 978 cases (163 subjects × 6 rounds) and the average reward was €4.16.[23]

[23]The average reward is among all the cases in a treatment (all subjects in all the rounds). For a subject, one out of the six rounds is randomly chosen to determine the incentive payoff and the average incentive payoff among the subjects is €3.48 for the *Product* treatment and €4.74 for the *Summation* treatment.

Table 2.6 – Anderhub et al. (2000) Replication: Reactions to the first removed card deck and finally stayed card deck

Panel A: **Mean consumption share in period 2 and 3**

Treatment	Mean consumption share in period 2 ($\frac{x_2}{S_2}$)			Mean consumption share in period 3 ($\frac{x_3}{S_3}$)			Obs.	Obs. fulfilling Condition 1	Obs. fulfilling Condition 2
	First period removal			Finally stayed deck					
	¬[1/6]	¬[1/3]	¬[1/2]	[1/6]	[1/3]	[1/2]			
Product	0.33	0.34	0.35	0.44	0.42	0.40	176	19	29
Summation	0.35	0.35	0.34	0.43	0.41	0.42	163	39	24

Panel B: ***P*-value of *t*-test of the inequality between the consumption shares**

Treatment	*P*-value of *t*-test					
	Null hypothesis for Condition 1			Null hypothesis for Condition 2		
	$(\frac{x_2}{S_2}\|\neg[1/6]) \leq (\frac{x_2}{S_2}\|\neg[1/2])$	$(\frac{x_2}{S_2}\|\neg[1/3]) \leq (\frac{x_2}{S_2}\|\neg[1/2])$	$(\frac{x_2}{S_2}\|\neg[1/6]) \leq (\frac{x_2}{S_2}\|\neg[1/3])$	$(\frac{x_3}{S_3}\|[1/6]) \geq (\frac{x_3}{S_3}\|[1/2])$	$(\frac{x_3}{S_3}\|[1/3]) \geq (\frac{x_3}{S_3}\|[1/2])$	$(\frac{x_3}{S_3}\|[1/6]) \geq (\frac{x_3}{S_3}\|[1/3])$
Product	0.886	0.863	0.545	0.978	0.807	0.874
Summation	0.387	0.336	0.554	0.664	0.260	0.857

Notes: The mean consumption share is computed from all the subjects in period 2 ($\frac{x_2}{S_2}$) and 3 ($\frac{x_2}{S_2}$) correspondingly. For each subject, there are two rounds out of the six rounds where the first removed card deck (finally stayed card deck) is of a same termination probability. The consumption share in period 2 (period 3) for each subject is the mean of the shares of the two rounds with the same termination probability of first removed card deck (finally stayed card deck). Condition 1 refers to $(\frac{x_2}{S_2}|\neg[1/6]) > (\frac{x_2}{S_2}|\neg[1/3]) > (\frac{x_2}{S_2}|\neg[1/2])$, implying that the spending in period 2 is larger when a card deck with a low termination probability is removed than a card deck with a high termination probability is removed. Condition 2 refers to $(\frac{x_3}{S_3}|[1/2]) > (\frac{x_3}{S_3}|[1/3]) > (\frac{x_3}{S_3}|[1/6])$, implying that the spending in period 3 is larger when a card deck with a high termination probability finally stays than a deck with a low termination probability stays. The *p*-value is from *t*-tests of the null hypotheses.

Following the original study, we show the univariate statistics for the average consumption decision per period according to the ambiguity resolution path in Figure 2.2 and Figure 2.3. The red outline of the nodes indicates that on average, the subjects of that treatment, period, and uncertainty resolution path spent above the optimal consumption levels for that node; the blue indicates spending below these optimal levels. 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.

In our sample, the efficiency rate in the condition *Summation* is higher than in the condition *Product*, as in the original study, but the average efficiency of consumption decisions is smaller in both treatments as compared to the original study. Additionally, we found that there is substantially less differential adjustment to the resolution of uncertainty of termination probabilities than in the original study. For example, in the second period (X2), for the *Summation* condition the average spending ranges from 2.47 to 2.54 points, while in the original study the averages ranged between 2.56 and 3.23. We also found that, contrary to the original study, when ambiguity of termination risk is eliminated, the fraction of endowment consumed did not vary substantially according to the optimal levels under either treatment condition, as seen in Table 2.5. 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 observed levels in the original study of 0.70 to 0.83). Overall, these observations suggest that the subjects in our sample reacted much less to changes in their termination risk than subjects in the original study, of which the main result was that "subjects' reactions to information about termination probabilities are qualitatively correct."

To evaluate whether subjects respond qualitatively correctly to the resolution of uncertainty, we first checked the reactions to the removal of the first termination probability (card deck). The removal of a card deck for a low termination probability after the first period decreases the expected length of the round for the subject, and thus he/she should consume more in the second period; conversely, removal of a card deck with a high termination probability increases the expected length of the round, and incentivizes a reduction in consumption. This implies the condition $\left(\frac{x_2}{S_2}|\neg[1/6]\right) > \left(\frac{x_2}{S_2}|\neg[1/3]\right) > \left(\frac{x_2}{S_2}|\neg[1/2]\right)$, where $\neg$ is the removal of a card deck (set for one termination probability), x_2 is the spending decided in the second period, and S_2 is the disposable amount in that period. E.g., $\left(\frac{x_2}{S_2}|\neg[1/2]\right)$ means the proportion of the decided spending to the disposable

Table 2.7 – Anderhub et al. (2000) Replication: Facing an uncertain future

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						

Notes: 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.

endowment in Period 2 when the card deck with termination probability 1/2 is removed. Likewise, in the third period, when all ambiguity has been resolved and one probability remains, subjects should consume more when the final termination probability is high and less when that probability is low. This implies the condition $\left(\frac{x_3}{S_3}|[1/2]\right) > \left(\frac{x_3}{S_3}|[1/3]\right) > \left(\frac{x_3}{S_3}|[1/6]\right)$, where $\left(\frac{x_3}{S_3}|[1/2]\right)$, for example, means the proportion of the decided spending to the disposable endowment in Period 3 when the card deck with termination probability 1/2 remains. The mean consumption shares in Panel A of Table 2.6 exhibit no obvious difference when different card decks are removed. The statistical tests in Panel B of Table 2.6 also show that, on average, the subjects in our sample do not fulfil either of the two conditions. For each condition, the original study rejects the hypothesis that the subjects do not fulfil the condition with a binomial test for both treatments. A summary of these analyses for socio-demographic subsamples is reported in Table A2 and Table A3 in the Appendix.

Another check of the quality of decisions relies on the observation that consumption should always be larger in the earlier periods when the period of termination (the length of the round) is still undetermined. In Table 2.7, we tabulate the percentage of cases where this condition is met, according to the termination periods of each subject in each round. Our results confirm the observation in the original study that a large fraction of cases does not adhere to relatively relaxed conditions. For example, in the right column of panel A for the *Product* condition, only 35.6% of the cases who reached period 6 had 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).[24]

2.3.4 Replication Results of Fatas, Lacomba and Lagos (2007)

This study investigated the impact of the structure of retirement payouts on the choice of when to retire when the subjects face longevity risks. The three considered structures are *Annuity*, *Lump sum*, and *Combined*. At the start of a round, subjects chose the period in which they wanted to start collecting retirement benefits. In every period, there was risk of being terminated, and subjects only earned payoffs in a round while they are still active.

In the *Annuity* treatment, subjects received a fixed payout per period, starting at their chosen retirement period. In the *Lump sum* treatment, subjects earned a single payout at their chosen retirement period and nothing in any other active period. In the *Combined* treatment, they earned both a lump sum and an annuity, as in the previous treatments. In all the treatments, the payout was higher if the subjects chose a later period to retire (start collecting the payout). However, the subjects received payouts only if they were active when the retirement period arrived. The expected value of the payoff per round was equal for all the treatments and chosen periods of retirement. In our study, subjects underwent three rounds (the original study comprised a single round) to allow evaluating learning effects, of which one was randomly chosen for compensation based on the total

[24] In this analysis, most of the discrepancies between our results and the original study 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 ($x_3 \geqslant x_4 \geqslant x_5 \geqslant x_6$). By contrast, in the original study many subjects violated the strict condition, but kept consumption numerically constant between two rounds. This specific difference between violations of strictly and weak conditions is, arguably, due in part to our use of a slider precise to increments of 0.01, rather than requiring a numerical input.

payoff accrued in the chosen round.

Termination in a round was determined by a random draw of cards without replacement at each period – starting with 14 green cards and one red card (in the original study, coloured balls were used instead). Each period, a card was selected from the stack, and the subject was terminated in a round when a red card is drawn. This procedure generates a survival function with interesting properties: known maximum length of experimental life (as a red card will be eventually drawn with certainty), decreasing one-period survival probabilities, and increasing rate-of-change of survival probabilities. These are properties also found on stylized human survival curves for individuals approaching the typical age of retirement.

The three treatments have equal expected lifetime payout for any period of retirement chosen by subjects (when adjusted for the implicit survival probabilities) such that in theory there should be no difference between the treatments in the choice of the timing of retirement if subjects were neutral to the structure of the payoffs.

In our replication, 530 subjects completed the experiment (177 in the *Annuity*, 170 in the *Combined*, and 183 in the *Lump sum* condition – in the original study, these numbers were 28, 26, and 22, respectively). Similarly to the results of the original study, we found 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 2.4. On average, subjects in the *Annuity* condition chose to retire after 5.49 periods, those in the *Lump sum* condition retired latest (after 6.32 periods), and those in *Combined* after 6.13 periods (in the original experiment, they chose to retire after 5.0, 9.0, and 7.0 periods, respectively).

As in the original study, we found significant treatment effects between the treatments of *Lump sum* and *Annuity*. Following the original study, using *Lump sum* as a baseline, we regressed the chosen retirement period on the treatment indicator variables while using our own measures of *risk taking* and *patience* as controls. The results are shown in Table 2.8. In the full specification (4), subjects in the *Annuity* condition chose to retire 0.916 periods earlier than those in *Lump-sum*. The difference was smaller (0.863 periods) but still significant before controlling for patience in (5). The difference in the estimated coefficients of *Annuity* and *Combined*, shown in the bottom

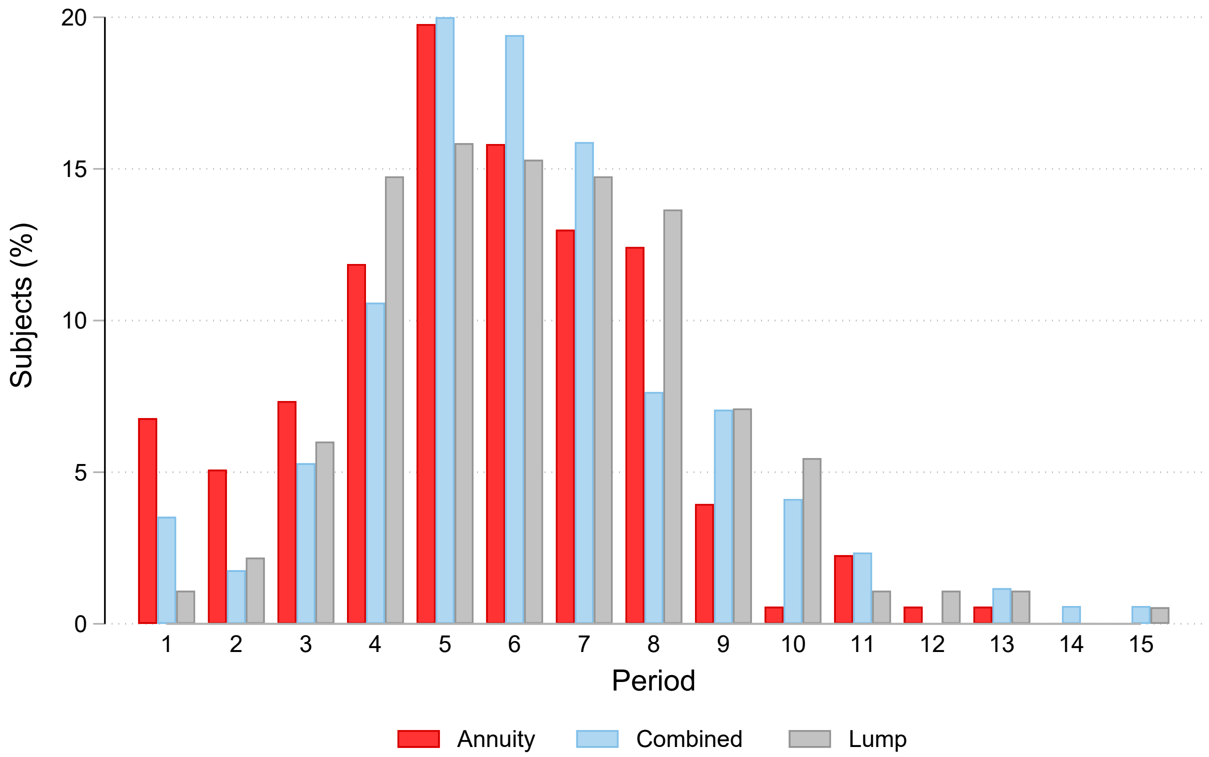


Figure 2.4 – Fatas, Lacomba, and Lagos (2007) Replication: Timing of Retirement.
Notes: Period chosen by subjects to (start) collecting payoffs, conditional on not having been terminated.

panel (0.635 and 0.645 periods in (4) and (5)), is significant, as it was in the original study. However, contrary to the original study, the difference in the coefficients between *Lump sum* and *Combined* was not significant in either specification. Table A4, in the Appendix, shows results for the same analysis repeated for socio-demographic subsamples.

Similarly to the original study, we found 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 was associated with a delayed retirement timing of 0.026 to 0.028 periods (the original study used a different risk-preference elicitation method). Patience was also positively associated with a delay in retirement. Each month that subjects chose to wait for their payoff in exchange for 5% interest (per month) is associated with a delay in the choice of retirement period between 0.184 and 0.255 periods.

In an additional preregistered analysis that was not part of the original study, we analysed how termination at round 1 and/or 2 affected the choice of timing of retirement in later rounds. Termination

Table 2.8 – Fatas, Lacomba, and Lagos (2007) Replication: Timing of Retirement Treatment Effects

	(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

Notes: The results are from OLS estimations. The dependent variable is the mean retirement period chosen in the three rounds. *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 and equal to 1, 2, 3 and 4 for the choice of no delay, 1 month, 2 months and 3 months, respectively. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

is the most salient event in a round, and being terminated before one's chosen retirement period means that no payoff is accrued in that round. The experience of termination in earlier rounds might influence the subsequent decisions of subjects in later rounds, as they learn throughout the rounds. The results of this additional analysis are presented in Table 2.9.

We found that generally, a later termination in earlier round(s) was associated with a significantly delayed choice of retirement in subsequent round(s). In specification (3), controlling for the treatment, a first round that lasted one period longer delayed the retirement timing chosen in the second round by 0.06 periods. A much more salient event is that the subjects survived at least until the period they had chosen to earn (or start earning) their payoffs. In specification (4), we regressed the choice of timing of retirement in round 2 on whether the subject survived until their chosen timing of retirement during round 1. In round 1, surviving at least until the chosen period delayed the subsequent choice of timing of retirement chosen in round 2 by 2.782 periods. The direct effect of one later period for termination was then a further delay of 0.277 periods for round

Table 2.9 – Fatas, Lacomba, and Lagos (2007) Further Analysis: Effects of Experienced Termination Period on Later Decisions

	(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)
Non-zero pay round 1				2.782***				2.290***
				(0.332)				(0.335)
Non-zero pay 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

Notes: The results are from OLS estimations. 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. *Non-zero pay round 1* (*2*) is a dummy indicating the payoff in round 1 (2) is non-zero (one of the three rounds is randomly chosen at the end of the study to determine the final payoff). *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 and equal to 1, 2, 3 and 4 for the choice of no delay, 1 month, 2 months and 3 months, respectively. Standard errors are in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2. In specifications (5-8), we tested retirement timing in round 3 given outcomes of the first two rounds: the effect of termination period and survival until the chosen period remained significant in all the specifications.

2.3.5 Replication Results of Meissner (2016)

This study evaluated the consumption smoothing behaviour when debt is treated differently than savings. To study this question, the study allowed interest-free borrowing. Over a set number of periods in a life (round), subjects decided on savings and consumption while facing different broad income paths, increasing or decreasing throughout a round, with local stochastic perturbations (such that the income paths were not strictly monotonically increasing or decreasing). The study tested the hypothesis that with an induced CARA utility reward function of the consumption in a period, subjects should smooth their consumption throughout all periods in a round. On a downward income path, smoothing lifetime consumption requires *saving* from earlier periods for later consumption (*Saving* condition). On an upward income path, lifetime consumption smoothing requires *borrowing* from later periods when income will be higher (*Borrowing* condition).

Treatment groups differed in the sequence of conditions that the subjects faced. In the treatment *Savings First*, subjects played two rounds in the *Saving* condition, then switched to *Borrowing* for another two rounds, while in the *Borrowing First* treatment this order was reversed. In the last period of either condition, no decision was made, and its consumption (spending) was set such that lifetime consumption would equal lifetime income. Our replication focused on the treatment effect of symmetric financial decisions (saving or borrowing) on lifetime consumption smoothing.[25]

The null hypothesis of the study is that regardless of the income path, the consumption paths should be equally smooth if debt is not evaluated differently than saving. The behaviour under the first of the conditions to which subjects were randomly assigned should not differ from the behaviour under the second condition in the last half of the session. The expected payoff is maximal when subjects smooth their consumption regardless of the income path.

To simplify the task and make it viable to implement with our sample, we first reduced the length of the experimental life (from 20 to 16 periods) and the repetitions (from three to two rounds per condition). We also modified the variables in the experimental environment of the original study. In our replication, subjects earned income in points and variable incentives, per period, in the form of induced CARA utility over their consumption, which was then converted into euros ('Eurocent Rewards'). In the original experiment, subjects earned 'Talers' instead, which they converted into utility-induced 'points,' summed across each round and then converted into monetary units. We bypassed this intermediate utility computational variable and presented the CARA-induced utility conversion as both a static graph and as dynamic text information per period, as subjects manipulated a slider prior to confirming their decisions. We also simplified the variable incentive

[25]The original study further investigated the roles of myopia and learning on these consumption decisions.

to be the lifetime sum of ‘Eurocent Rewards’ in one randomly chosen round (the original study used the average of total payoffs of one round of the first treatment and one round of the second treatment, per subject). To reduce the task complexity, we also did not allow for negative spending (to be distinguished from negative savings, i.e., borrowing) in periods other than the last.[26]

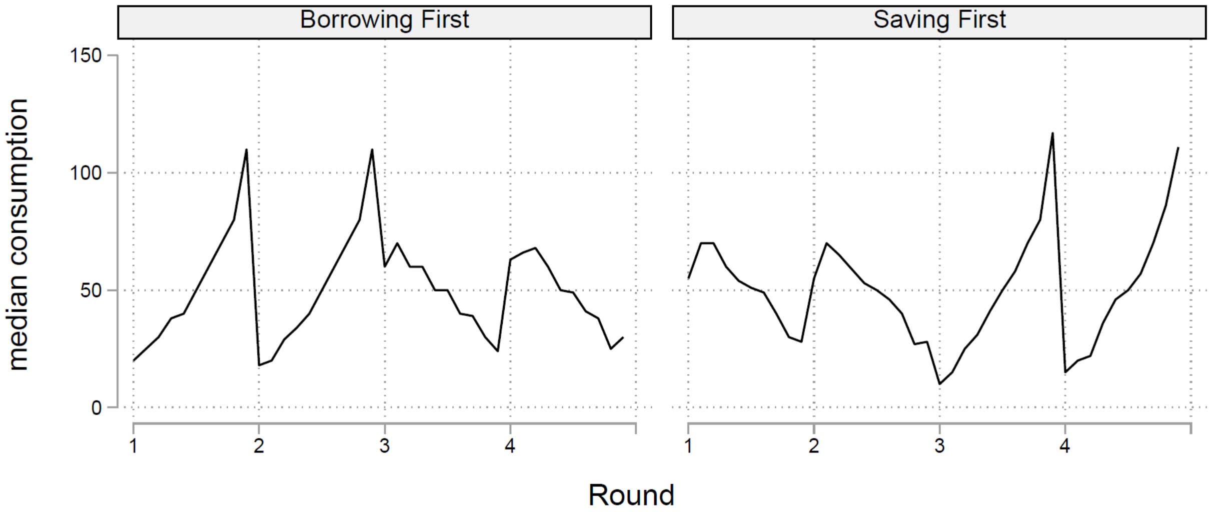


Figure 2.5 – Meissner (2016) Replication: Median Consumption per Period over Sequential Rounds by Treatments.
Notes: Borrowing (saving) first subjects play rounds 1 and 2 in the borrow (saving) condition; and rounds 3 and 4 in the saving (borrow) condition.

In total, 278 subjects completed the experiment, of whom 147 in the *Borrowing First* treatment and 131 in *Savings First* (the original experiment recruited 38 subjects for each treatment).[27]

In Figure 2.5, 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 from future income to smooth consumption in earlier periods. Lifetime consumption is smoother under the *Saving* condition, as its subjects have to save a part of the income they have already earned at present instead of borrowing from future expected higher income. However, compared to the original study, the median consumption among subjects in the *Saving* condition is not as smooth in our study.

[26]Voluntary negative spending, as allowed in the original study, is a very hard feature to conceptualize for subjects, and it would have required a significant expansion of the instruction set. 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 × period × round) spending decisions were negative.

[27]In the original preregistered plan, we had proposed excluding 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 the experimental data collection had been live for less than one day, and only ten subjects had 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 2.3.1.

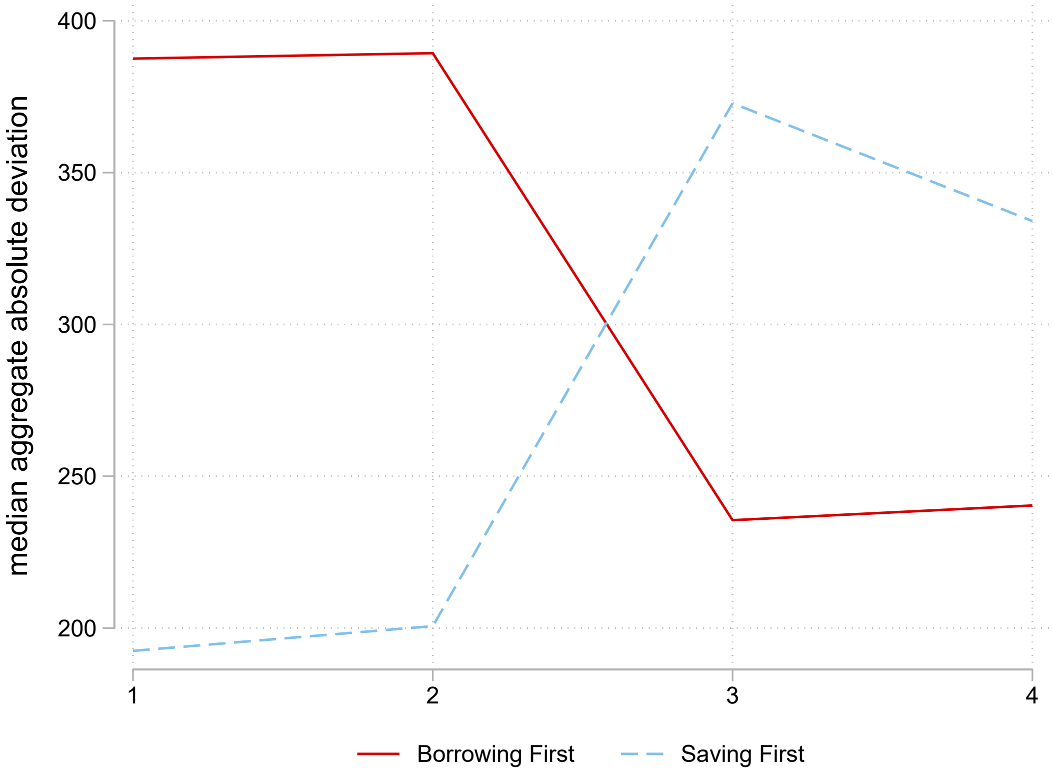


Figure 2.6 – Meissner (2016) Replication: Sub-Optimal Consumption.
Notes: Medians of *Measure 2* (mean absolute deviation of consumption from optimal path at each round, per subject × round) by treatment condition. Borrowing (saving) first subjects play rounds 1 and 2 in the borrow (saving) condition; and rounds 3 and 4 in the saving (borrow) condition.

Order effects of the income paths from treatments *Borrowing First* or *Savings First* did not appear to significantly affect the results in each treatment of our experiment, as in the original study.

Following the original study, we use three measures to evaluate deviations from optimal consumption. *Measure 1* is the lifetime sum (within a round) of the period deviations between observed consumption and the optimal consumption at each period, conditioned on the wealth (unspent endowment) of the subject at the start of each period. *Measure 2* is the lifetime sum of the absolute value of those same period deviations.[28] As in the original study, subjects deviated more from conditionally optimal consumption paths in the *Borrowing* condition (i.e., rounds 1-2 for *Borrowing First* and 3-4 for *Savings First*) than in the *Saving* condition. In turn, *Measure 3* is the lifetime sum of the period utility losses between observed consumption and optimal consumption at ex-ante

[28]This implies that Measure 1 and Measure 2 recalculate the optimal consumption path for the remaining periods of each round, for each subject, considering both the past income path and the previous decisions the subject already made in previous periods of that round.

(start of a round) optimal wealth levels.

In an additional preregistered analysis, we controlled for the impact of *risk taking* and *patience*, and found that the results remained qualitatively unchanged: treatment has significant effects on *Measure 1* and *Measure 2*, but not on *Measure 3*.[29]

Table 2.10 – Meissner (2016) Replication: Sub-Optimal Consumption Paths

		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

Notes: 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. $N = 278$.

In bivariate analyses with Mann-Whitney U tests, reported in Table 2.10, we found that *Measure 1* and *Measure 2* differ statistically significantly between treatments in all rounds (effect size – in the first round – 0.470 and 0.412; statistical power (5% level) 0.973 and 0.916 for *Measure 1* and *2* respectively). In the original study, *Measure 1* was statistically significant in all rounds, and *Measure 2* was significant in three of the six rounds (5% level). Thus, deviations from conditionally optimal consumption paths are higher for the *Borrowing First* condition than for the *Savings First* condition, regardless of the within-subject order of both conditions. This lends supports to the debt-aversion hypothesis, as subjects are less willing to borrow from the future to consume now than to save from

[29]ANOVA analysis was used in the additional analysis. The independent variables in ANOVA include treatment dummy (if *Borrowing First*), the condition (if *Borrowing*), the risk-taking choice, the delay choice (patience), and the interaction between the treatment and risk-taking choice, treatment and delay choice, condition and risk-taking choice, and condition and delay choice.

the present to consume in the future in order to smooth consumption. The utility loss from the deviation from the unconditionally optimal consumption path (*Measure 3*) is significant only for the first two rounds before the switch of the conditions, making it resemble the results from the original study, in which it was significant only for the first three rounds before the switch. In the Appendix, Table A5 and Table A6 show a similar analysis over different socio-demographic subsamples of subjects for *Measure 1* and *2*, respectively.

2.3.6 Replication Results of Blaufus and Milde (2021)

For this replication, we were interested in the main treatment effects of different but economically equivalent taxation regimes on retirement savings decisions. The experiment consisted of a "working" phase and a "rest" phase. During the working phase, subjects decided between saving and spending. Each round had ten working periods (with fixed wages) and five resting periods. Each subject completed two rounds in a treatment condition that did not change for these first two rounds. The treatment conditions varied the taxation regime for savings. In *Immediate* taxation, subjects paid income taxes immediately, but their savings were tax-free upon withdrawal during retirement. In *Deferred* taxation, subjects did not pay income taxes on their savings (they got a tax rebate from income taxes) but were taxed later when they withdrew savings during retirement. Finally, in the *Matching* condition, subjects received matching contributions to their savings and paid taxes later, upon withdrawal, during retirement. The balance in all savings accounts earned an interest of 5% per period, with interest earned being taxable or tax-exempt according to the tax rule applied to the principal amount of savings. Withdrawals after retirement were automatically calculated and made equal for all periods of the rest phase.[30]

Subjects completed two rounds, and were compensated based on their consumption decision in one randomly chosen period of one round, to incentivize them to smooth their consumption. As the three treatment conditions yield economically-equivalent returns on savings, they should command equal after-tax effective savings rates.

To simplify the experimental design, we removed an attention check and reassurance screen of tax

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

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 the Gill and Prowse (2012) sliders task. In terms of control variables, we retained age and gender, but used our own risk-taking measure for identification of *High risk-taking* subjects taking the 75th percentile cut-off here from the original study. Furthermore, we used our measure of *financial ability* as a replacement for the original study measure of financial knowledge. Due to session time constraints, we did not collect information on tax aversion or procrastination.

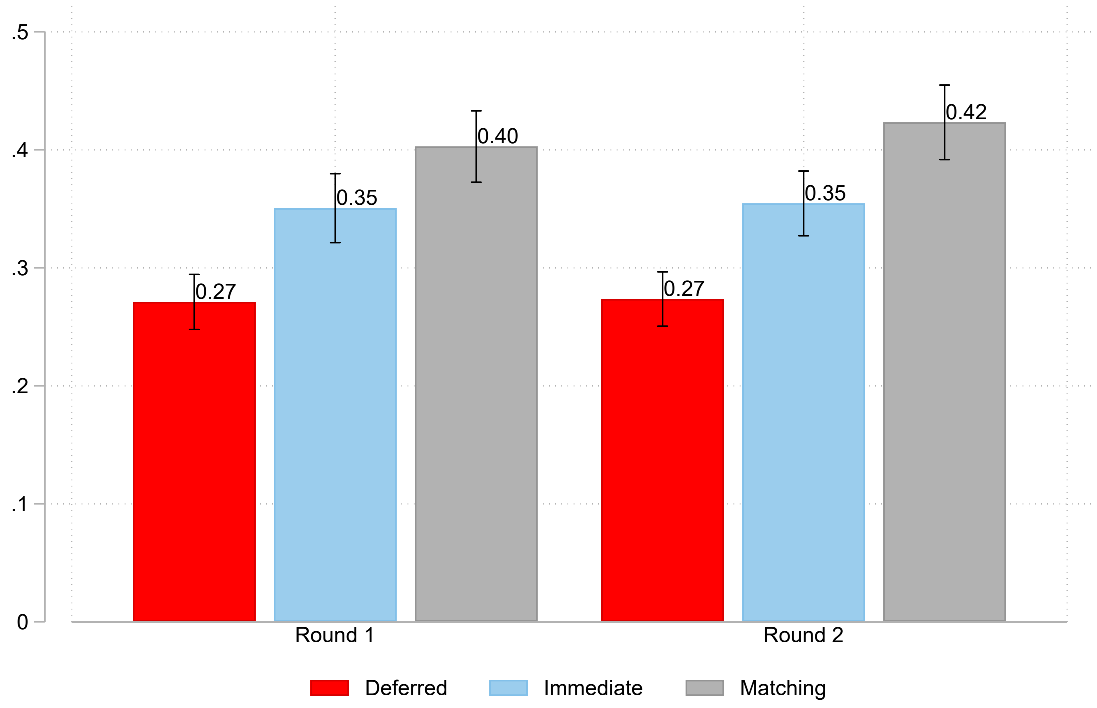


Figure 2.7 – Blaufus and Milde (2021) Replication: Average Savings Rates (95% confidence interval).
Notes: Direct (total) saving rates used for *Immediate* condition, and effective savings rates for *Deferred* and *Matching*, per round.

As in the original study, our main dependent variables were *savings rate* (naïve 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 replication, we collected 522 valid responses (306 in the original study), of which 182 in the *Immediate* treatment condition, 162 in *Deferred*, and 178 in *Matching* (in the original study, 104, 105 and 97, respectively).

We first calculated the unconditional means of the compatible savings rates across treatments, with 95% confidence intervals (see Figure 2.7). As in the original study, we observed that the savings rates did not change significantly between the first and second round, and *Immediate* savings rates were higher than *Deferred* effective savings rate.

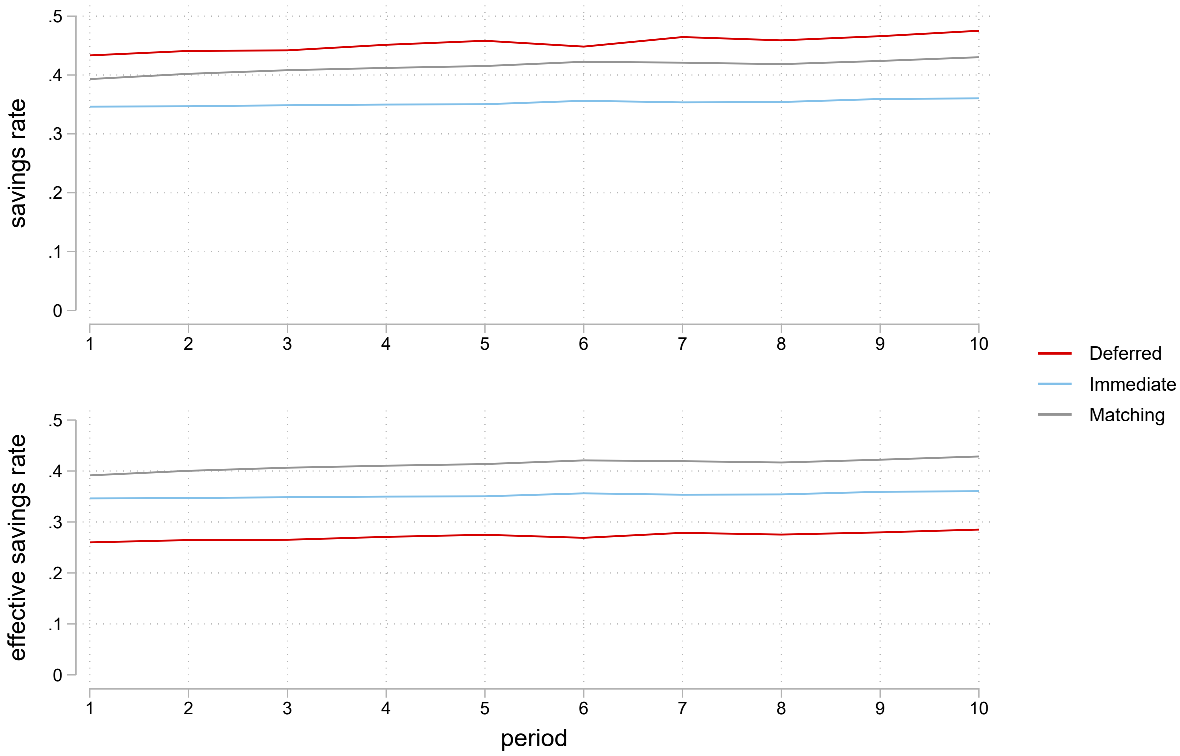


Figure 2.8 – Blaufus and Milde (2021) Replication: Savings Persistence.
Notes: Average (effective) saving rates per period across rounds.

Both savings measures were reasonably stable over periods, as their aggregate levels per period and round show in Figure 2.8.

Following the analysis of the original study, we regressed savings rates and effective savings rates, observed at the subject × period × round level,[31] on the binary indicators of treatment and the aforementioned covariates. The results of the estimation are presented in Table 2.11. All models

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

include subjects of the *Immediate* treatment. For treatment contrasts, models (1-5) include *Deferred* subjects only, while models (6-10) add *Matching* subjects only.

Both treatment coefficients are statistically significant in all estimation models, and all treatments and the magnitudes of our estimated coefficients are similar to those of the original study. In our replication, both the *Deferred* and the *Matching* tax-protected savings schemes increased the base savings rate from the *Immediate* condition (models 1-3 and 6-8 in Table 2.11). In model (2), the base savings rate of the *Deferred* subjects was on average 9.2 percentage points higher than that of the *Immediate* subjects in the first round. In model (8), the base savings rate of the *Matching* subjects was on average 6.9 percentage points higher than that of the *Immediate* subjects. Tax rebates and matching contributions appeared to attract savings in nominal terms, as in the original study.

However, this comparison of base savings rates ignores the fact that, in both *Deferred* and *Matching* conditions, withdrawals will be taxed, whereas *Immediate* withdrawals are tax-exempt. Like the original study, our analysis of effective savings rates shows that the economically equivalent savings rate of the *Deferred* subjects is on average 8.6 percentage points lower than that of the *Immediate* subjects (see model (4) in Table 2.11). However, the effective savings rate of the *Matching* subjects is on average 5.0 percentage points higher than that of the *Immediate* subjects (see model (9) in Table 2.11). In other words, the *Matching* contribution tax regime generates higher average post-tax net pension savings than the baseline *Immediate* taxation scheme. We repeated these analyses for socio-demographic subsamples of subjects and report the results in the Appendix, in Table A7 and Table A8 for savings rate and effective savings rate, respectively.

In contrast to the original study, we found that *male* was a significant negative predictor of savings rates in the *Immediate* and *Deferred* treatment group. Furthermore, in our replication, *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 found that *Period* has a positive and significant coefficient in our sample, while in the original study it had a significant negative coefficient. However, the effect magnitude of *Period* is small. In period 10, subjects in our sample would save 2% to 3% more from their income than in period 1. In the original study, savings and effective savings rates decreased over periods.

Table 2.11 – Blaufus and Milde (2021) Replication: Drivers of Saving Behaviour

	(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	0.101***	0.092***	0.096***	-0.086***	-0.085***					
	(0.025)	(0.024)	(0.024)	(0.018)	(0.018)					
Matching						0.054*	0.052*	0.069**	0.050*	0.067**
						(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Period	0.003**	0.003**	0.003**	0.002**	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		0.103***	0.034	0.078***	0.030		0.043	0.022	0.043	0.022
		(0.027)	(0.027)	(0.021)	(0.021)		(0.023)	(0.024)	(0.023)	(0.023)
Male		-0.050*	-0.048	-0.040*	-0.036		-0.019	-0.019	-0.019	-0.019
		(0.025)	(0.025)	(0.020)	(0.020)		(0.023)	(0.023)	(0.023)	(0.023)
Financial training		-0.008	0.019	0.002	0.020		0.005	0.019	0.005	0.019
		(0.027)	(0.028)	(0.022)	(0.022)		(0.024)	(0.026)	(0.024)	(0.026)
High risk-taking		0.116***	0.085**	0.089***	0.066**		0.088***	0.085***	0.088***	0.085***
		(0.028)	(0.028)	(0.022)	(0.021)		(0.026)	(0.025)	(0.026)	(0.025)
Constant	0.335***	0.316***	0.337***	0.323***	0.340***	0.334***	0.314***	0.326***	0.314***	0.326***
	(0.016)	(0.024)	(0.024)	(0.021)	(0.021)	(0.016)	(0.022)	(0.022)	(0.022)	(0.022)
Observations	3,440	3,440	3,440	3,440	3,440	3,600	3,600	3,600	3,600	3,600
Subjects	344	344	344	344	344	360	360	360	360	360
R2	0.0429	0.1298	0.0813	0.1240	0.0835	0.0164	0.0560	0.0578	0.0553	0.0568

Notes: The table presents regression results of random-effects models explaining subject's (effective) savings rates. The savings rate (SR) is defined as the saving amount in a given period divided by the income in this period. The effective savings rate (ESR) is the savings rate 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 the observations. *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. *R2* is the R-squared for overall model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.3.7 Replication Results of Koehler, Langstaff and Liu (2015)

This study analyses whether subjects adjust their consumption behaviour to changes in the availability of income over a life-cycle. To evaluate this behaviour, subjects are asked to made decisions over several rounds of multiple periods. Each round had a working phase and a retirement phase. During the working phase, subjects earned a predetermined income, which increased over periods. They decided how much to spend and how much to save in a simple, interest-free cash account. During the retirement phase, income was zero. We focused on the main treatment effect of the relative length of the retirement phase (*Short* or *Long* retirement) to the total life length (in periods). In our replication, out of 16 periods per round, subjects were 'retired' for four periods in the *Short* retirement condition and for eight periods in the *Long* condition. In the original study, each round lasted 24 periods, with *Short* retirement consisting of 6 periods and *Long* retirement of 12 periods. In our study, subjects played two rounds under one condition, then changed to the other for another two rounds, with a random assignment of the starting condition (in the original study, subjects played four rounds, switched conditions and then played another four rounds). The compensation in our replication depended on the spending in one randomly selected period. The original study did not use variable incentives.

In every period, subjects had to pay expenses, which were automatically deducted from their income. At the start of a round, a card deck with the value of all possible expenses for every period was shown. Then, at each period one card was randomly chosen and removed (without replacement), determining the actual expenses of that period. During the working phase, income was always larger than mandatory expenses, such that even subjects who always consumed all their income in all periods would still be able to meet their expenses. In the retirement phase, subjects who did not save enough in the working phase would be unable to meet expenses, or go 'bankrupt.'[32] In the original, non-incentivized study, bankruptcy did not have any further repercussions for the subject; in our replication, however, a bankrupt subject would earn zero variable payoff if a round in which he/she was bankrupt was selected for compensation.[33] Compared to the original design,

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

[33] This is implemented to prevent strategic but unwanted behaviour on consumption decisions. For instance, consider a subject who, as periods advance, sees that the random realization of expenses will backload the high expense periods during the retirement phase. This subject could decide to spend more during the lower expenses period, while his

this bankruptcy penalty strengthens the incentive for subjects to smooth consumption and, 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 (149 in the original study), of whom 166 started the session under the retirement condition *Long* and 178 started under the condition *Short* before switching.

Following the original study, we analysed (1) whether the subjects saved enough for retirement (i.e., whether they made sufficient adjustments in saving in response to the manipulation of retirement length), and (2) whether the subjects smoothed their consumption over periods. With respect to the first question, we found that participants saved more when faced with a *Long* retirement period than when faced with a *Short* retirement period, as in the original study. In ANOVA analyses, the retirement length treatment has a significant effect on retirement savings, with $F(1375) = 1495$, $adjusted\ R^2 = 0.752$, $p < 0.001$ (effect size 0.52, statistical power > 0.999), whereas the original study found $F(1147) = 379$, $adj.\ R^2 = 0.72$, $p < 0.001$. With respect to the second question, we found that consumption smoothing as measured by the variability of spending did not differ between conditions, with $F(1375) = 0.52$, $adj.R^2 = 0.648$, $p = 0.471$ (effect size 0.01, statistical power 0.071). This observation is in contrast to the original study, which found a significantly greater mean spending variation (lower consumption smoothing, on average) in the *Long* condition than in the *Short* condition. These results of our replication do not change qualitatively after controlling for *risk taking* and *patience*.[34] The observation that consumption smoothing activities do not differ between treatments could be related to the incentives for consumption smoothing that we introduced.

budget slack to spend is higher, even while knowing that he would eventually go bankrupt, in order to maximize lifetime spending before bankruptcy.

[34]The results regarding the retirement savings and the spending variability remain the same when the observations with savings left unspent at the end of a round are excluded from the ANOVA analyses.

Table 2.12 – Koehler, Langstaff, and Liu (2015) Replication: Decision Constraints and Outcomes

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

Notes: 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 the standard deviation of spending. *Difference savings to previous period* is average change in retirement savings from the previous round.

In the Appendix, Table A9, we repeated these analyses for subsamples split according to socio-demographic characteristic of the subjects. We also used this replication to evaluate whether having a sample drawn from the general population, that is on average older and has a lower level of education than the sample of the original study,[35] matters for the significance of the treatment effects. The results suggest that neither age nor education level affects the significance of the main treatment effects.

As in the original study, we did not force subjects to automatically spend all points they had in the last period and allowed them to end a round with the remaining points in the savings account.[36] As part of our additional analysis, we investigated the implication of this feature on the subjects' decision behaviour.

Table 2.12 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 left 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 the other subjects are effective planners.

[35]The average age in our sample is 42 years old, the average age in the original study is 29 years old. The average qualification in our sample is 3.4 (3 is vocational qualification, 4 is Bachelor level), the average qualification in the original study is at a Bachelor level.

[36]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.

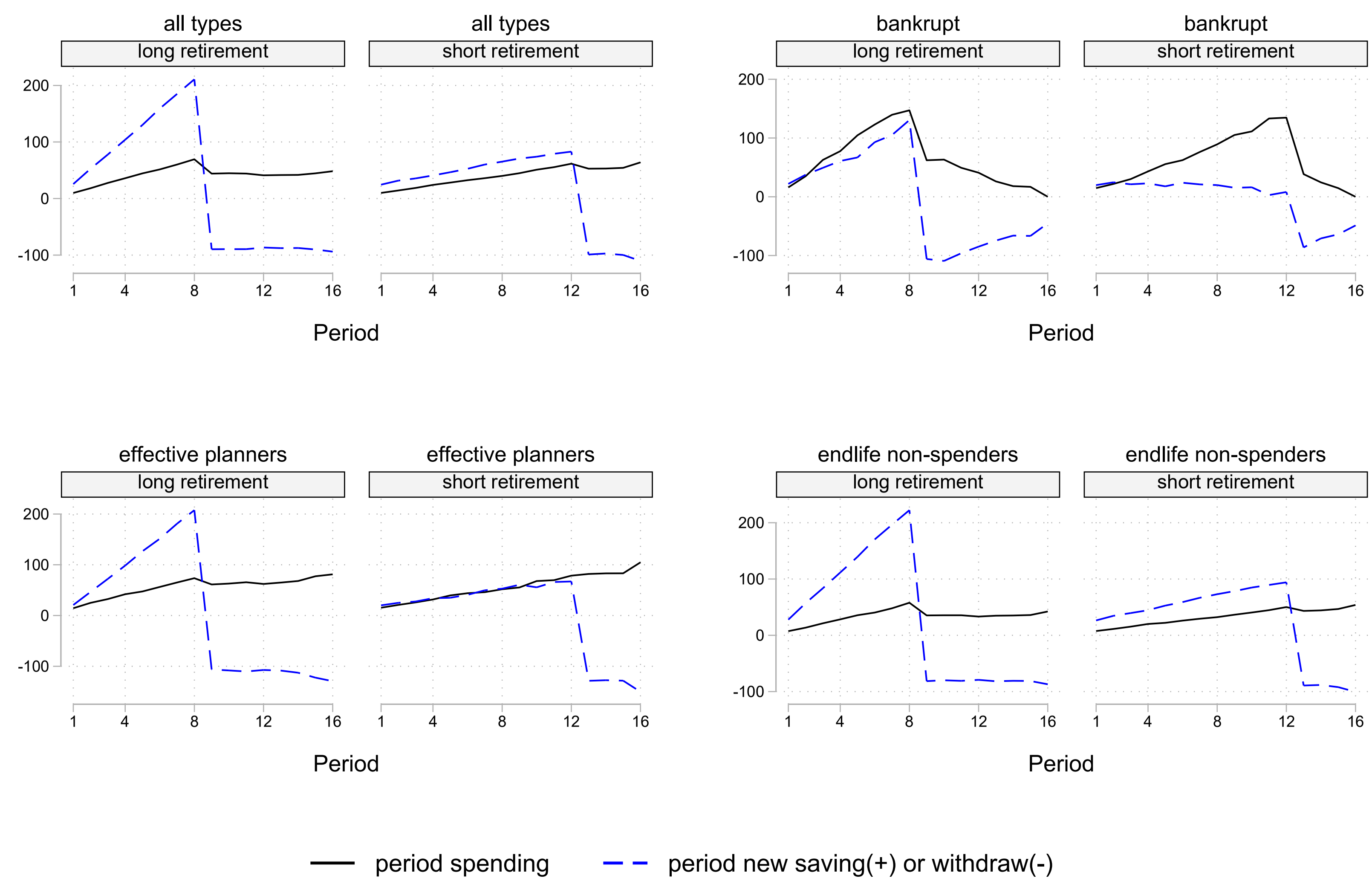


Figure 2.9 – Koehler, Langstaff, and Liu (2015) Replication: Spending and Saving per Period.

Notes: 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.

The subjects 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*. Likewise, 9% of the subjects in the first round *Short* did the same, 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 did not change noticeably between rounds. We do not have original study results to compare the prevalence of this outcome for each type of subject there.

The average savings and consumption paths for each type and for the entire sample are shown in Figure 2.9.

Since income increases along the periods during the working phase, while expenses do 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 many unspent savings behind.

2.4 Discussion and Implications for Future Research

In this section, combining insights from our replications 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 the limitations of our replication study and offer a modest suggestion for an agenda for future experimental research on retirement decision-making, considering also the current state of the literature and its limitations. Finally, we briefly discuss some policy implications of our findings.

2.4.1 Replication of Modified Tasks, Task Design Features, and Implementation Challenges

We replicated most of the main effects of the five studies we reviewed. We 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 allowed the use of a heterogeneous unassisted online sample without yielding excessive noise in the observed results. This highlights the potential for adopting general features of simplified life-cycle experimental tasks, like those we used, in future experimental work, echoing the proposition of Koehler, Langstaff, and Liu (2015). However, some important considerations and precautions, which we discuss in the following, may be the concern of future experiments.

We observed that in general, subjects' consumption smoothing still is fairly suboptimal, regardless of whether incentives for smoothing are presented in the form of lifetime induced utility or selection of one period per round. With respect to the latter, we did not observe consistent high-stakes gambling behaviour, i.e., subjects did not concentrate consumption or spending in just one period, creating a low chance of a high-value payoff.

One task design feature of concern is to impose a lifetime budget constraint, such that lifetime income matches lifetime consumption (with interest if applicable). In experiments that do not impose the constraints, subjects might spend too little throughout the periods, and leave unspent experimental currency units that are of no value after the end of a round. In particular, underconsumption (or oversaving) in later life periods has been identified in other studies using intertemporal allocation tasks, outside the context of retirement-like decision-making (e.g., Yamamori, Iwata, & Ogawa, 2018). Future experiments that impose lifetime budget constraints and then study lifetime outcomes (such as induced utility from spending or consumption in all periods) should look at the impacts of such constraints that self-resolve in the last period. Simultaneous aggregation of lifetime utility from subjects who on the one hand, in violation of a lifetime budget constraint, leave money unspent at the end of a round, and on the other who consume everything before the last period(s) does not allow distinguishing between these different decision-making phenomena. If both groups of subjects are present in a 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 identify such inefficient decisions. Additionally, the estimates 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 2.3.2). 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 imposed a minimum engagement that resembles the requirement in an in-person lab session of answering all questions of a quiz correctly before being allowed to proceed. At the same time, we allowed the subjects to revisit the instructions throughout the quiz and all subsequent tasks.

We reduced the number of discrete periods and/or rounds. Such changes did not materially affect the panel structure of the data collected on relevant points. More severe reductions to fewer periods should be implemented with caution to avoid degenerating the natural computational and sequencing complexity present in life cycle optimization decisions (through dynamic programming) in the field or in the laboratory.[37]

Further experiments might help learn the particular impacts of other features on life cycle experimental tasks. These often sidestep any implementation of time-discounting factors across periods, other than interest on savings. Relatively complex utility forms can be imposed through incentive-reward functions. However, we still have limited knowledge of how subjects would react if decisions were measured non-parametrically (as in Abdellaoui, Attema, & Bleichrodt, 2010), when, for example, longevity uncertainty and changes in the institutional environment are simultaneously introduced into the same task.

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

2.4.2 Limitations

First, we replicate studies on different topics related to the retirement decision-making problems, but we cannot jointly evaluate the success of our replications. This is the case because there are not enough studies on any single topic in the experimental retirement decision-making literature. This field of experimental research is still relatively young and encompasses several topics covered by only few studies each, and even then the outcome measures are not clearly defined as to allow for such joint analyses through the possible replication of several studies on the same topic and research question.

Second, our replication study covers only a subset of the topics addressed in experimental studies on retirement decision-making as summarized in Table 2.1. There were practical and operational restrictions imposed by the use of an online sample from a research panel of the general population, with limited attention, no possibility for interaction between subjects, and for real-time experimenter assistance. We thus were not able to cover other relevant topics, such as social learning and social interactions, that would require experimental tasks unfeasible for deployment in our sample. These restrictions also limited the scope of topics from which we could select studies to replicate, as certain topics had no feasible experiments for replication with our general population online sample.

Third, the characteristics of our sample required adaptation of certain features of the experimental designs of the original studies. Although these adaptations did not prevent the evaluation of the main treatment effects, we can only speculate about the reasons for which we were not able to replicate some of the original results. It is not particularly reasonable, although possible, that certain simplifications of the original experimental designs led to the non-replications that we observed.

Finally, in an effort to use the limited time and attention of our respondents efficiently, we assessed only a set of individual characteristics that is common in all original studies and that we additionally consider as important for the underlying decision problems. It is possible that some other personal characteristics – unrelated to the characteristics that we considered in our replications – also have an impact on the main outcome measures.

2.4.3 Future Research

Our general results suggest that most main treatment effects on individual decisions in the life-cycle can be studied with much simpler task designs, apt for deployment in online samples from the general population. This should open up opportunities for future experimental research that broadens our understanding of possible heterogeneous treatment effects in a more systematic framework, once the simpler designs reduce the hurdles for recruitment of broader and more heterogeneous samples.

Apart from questions related to the generalizability of experimental findings, future research should consider more systematic studies on specific topics. The overall complexity of life cycle 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 features of those decisions. 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 of optimal behaviour (as in Benartzi & Thaler, 2007). Winter, Schlafmann, and Rodepeter (2012) showed 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. There is also 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 in interaction with different characteristics of retirement decisions should be investigated in more depth.

Furthermore, experimental work should contribute to assessing how individuals break down the complex inputs of decisions (such as the annuitization choice) and the interaction between the inputs and other factors that determine decision behaviour in controlled settings. This is necessary since the theoretical or simulation-based literature does not sufficiently agree on what the necessary assumptions are for the unsettled and unsolved annuity puzzle. With simulations, Peijnenburg, Nijman, and Werker (2016) questioned some previous assumptions about the attractiveness (or lack thereof) of pension wealth annuitization for many subjects, which implies that normative prescriptions for rational annuitization decisions are less likely to break down than in the earlier work of Davidoff, Brown, and Diamond (2005) or J. R. Brown et al. (2008). More experiments are needed to simultaneously implement several key features of the annuitization decision. This could allow descriptive models to emerge and better explain whether, why, and to what extent subjects should (or should not) annuitize their pension wealth.

2.4.4 Policy Implications

Our results, taken together, suggest that individuals have limitations in their capacity to solve dynamic programming problems even in stylized and simplified form as in our replications. In the field, these decision problems are much more complex and, for the most part, do not allow subjects to learn from their own mistakes.

In particular, pension reforms over the last two decades have often focused on increasing individual control over certain financial choices in retirement, relaxing compulsory elements, creating opt-outs, and introducing flexible financial arrangements. A large empirical literature evaluates their impact (see Gough & Niza, 2011, for an overview). Our results show the evidence that the systematic patterns in retirement decisions lead to suboptimal outcomes in the behaviour of the general population. These patterns and the observation that the participants of the general population are not sensitive enough to changes of the decision environment should be considered when designing pension reforms.

2.5 Conclusion

Individual retirement financial decisions are complex, which makes them prone to magnification of biases and cognitive mistakes with adverse effects on the decision outcomes. The suboptimal outcomes are likely persistent, since retirement saving decisions also offer limited learning opportunities due to long lags between the moment of a decision and its outcome. Experimental research on retirement decisions and on how heterogeneous individuals engage in these decisions is therefore acutely needed 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 four experimental studies, each addressing different topics and incorporating different features of the retirement decision problem, and attempted to replicate 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 replicated most of the main effects of the original studies we selected for this exercise, which might raise the external validity of the findings.

Finally, we note that limitations remain in the extant simulation-based and field empirical literature on several topics concerning retirement decision-making. These present opportunities for a promising future agenda for experimental research.

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. https://doi.org/10.1111/j.1468-0297.2009.02308.x

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. https://doi.org/10.1257/aer.98.2.418

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. https://doi.org/10.1080/15427560.2015.1034857

Anderhub, V., Güth, W., Müller, W., & Strobel, M. (2000). An experimental analysis of intertemporal allocation behavior. *Experimental Economics*, *3*(2), 137–152. https://doi.org/10.1007/BF01669305

Ballinger, T. P., Hudson, E., Karkoviata, L., & Wilcox, N. T. (2011). Saving behavior and cognitive abilities. *Experimental Economics*, *14*(3), 349–374. https://doi.org/10.1007/s10683-010-9271-3

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. https://doi.org/10.1111/1468-0297.t01-1-00158

Benartzi, S., & Thaler, R. H. (2007). Heuristics and Biases in Retirement Savings Behavior. *Journal of Economic Perspectives*, *21*(3), 81–104. https://doi.org/10.1257/jep.21.3.81

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. https://doi.org/10.1016/j.jpubeco.2020.104144

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. https://doi.org/10.1016/j.jebo.2011.08.010

Blaufus, K., & Milde, M. (2021). Tax Misperceptions and the Effect of Informational Tax Nudges on Retirement Savings. *Management Science*, *67*(8), 5011–5031. https://doi.org/10.1287/mnsc.2020.3761

Bohr, C. E., Holt, C. A., & Schubert, A. V. (2019). Assisted savings for retirement: An experimental analysis. *European Economic Review*, *119*, 42–54. https://doi.org/10.1016/j.euroecorev.2019.05.020

Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and Visceral Temptation in Dynamic Saving Experiments. *Quarterly Journal of Economics*, *124*(1), 197–231. https://doi.org/10.1162/qjec.2009.124.1.197

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. https://doi.org/10.1257/aer.98.2.304

Browning, M., & Crossley, T. F. (2001). The Life-Cycle Model of Consumption and Saving. *Journal of Economic Perspectives*, *15*(3), 3–22. https://doi.org/10.1257/jep.15.3.3

Carbone, E. (2005). Demographics and Behaviour. *Experimental Economics*, *8*(3), 217–232. https://doi.org/10.1007/s10683-005-1464-9

Carbone, E. (2006). Understanding intertemporal choices. *Applied Economics*, *38*(8), 889–898. https://doi.org/10.1080/00036840500399313

Carbone, E., & Duffy, J. (2014). Lifecycle consumption plans, social learning and external habits: Experimental evidence. *Journal of Economic Behavior and Organization*, *106*, 413–427. https://doi.org/10.1016/j.jebo.2014.07.010

Carbone, E., & Hey, J. D. (2004). The Effect of Unemployment on Consumption: An Experimental Analysis. *Economic Journal*, *114*(497), 660–683. https://doi.org/10.1111/j.1468-0297.2004.00236.x

Carbone, E., & Infante, G. (2014). Comparing behavior under risk and under ambiguity in a lifecycle experiment. *Theory and Decision*, *77*(3), 313–322. https://doi.org/10.1007/s11238-014-9443-2

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. https://doi.org/10.1016/J.JBEF.2015.12.001

Davidoff, T., Brown, J. R., & Diamond, P. A. (2005). Annuities and Individual Welfare. *American Economic Review*, *95*(5), 1573–1590. https://doi.org/10.1257/000282805775014281

Druckman, J. N., & Kam, C. D. (2009). *Students as Experimental Participants: A Defense of the 'Narrow Data Base'*. https://doi.org/10.2139/ssrn.1498843

Duffy, J., & Li, Y. (2019). Lifecycle consumption under different income profiles: Evidence and theory. *Journal of Economic Dynamics and Control*, *104*, 74–94. https://doi.org/10.1016/j.jedc.2019.05.006

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. https://doi.org/10.3758/BRM.41.4.1149

Fatas, E., Lacomba, J. A., & Lagos, F. (2007). An experimental test on retirement decisions. *Economic Inquiry*, *45*(3), 602–614. https://doi.org/10.1111/j.1465-7295.2007.00027.x

Feigenbaum, J., Gahramanov, E., & Tang, X. (2013). Is it really good to annuitize? *Journal of Economic Behavior and Organization*, *93*, 116–140. https://doi.org/10.1016/j.jebo.2013.07.005

Feltovich, N., & Ejebu, O.-Z. Z. (2014). Do positional goods inhibit saving? Evidence from a lifecycle experiment. *Journal of Economic Behavior & Organization*, *107*(PB), 440–454. https://doi.org/10.1016/j.jebo.2014.01.015

Gechert, S., & Siebert, J. (2020). Preferences over wealth: Experimental evidence. *Journal of Economic Behavior & Organization*. https://doi.org/10.1016/j.jebo.2020.06.006

Gill, D., & Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, *102*(1), 469–503. https://doi.org/10.1257/aer.102.1.469

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645. https://doi.org/10.1162/003355397555217

Gough, O., & Niza, C. (2011). Retirement Saving Choices: Review of the Literature and Policy Implications. *Journal of Population Ageing*, *4*(1-2), 97–117. https://doi.org/10.1007/s12062-011-9046-4

Hanel, P. H., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? (M. Voracek, Ed.). *PLoS ONE*, *11*(12), e0168354. https://doi.org/10.1371/journal.pone.0168354

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. https://doi.org/10.1111/jofi.12828

Hey, J. D., & Dardanoni, V. (1988). Optimal Consumption Under Uncertainty: An Experimental Investigation. *The Economic Journal*, *98*(390), 105. https://doi.org/10.2307/2233308

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425. https://doi.org/10.1007/s10683-011-9273-9

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. https://doi.org/10.1016/j.jebo.2019.11.008

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. https://doi.org/10.1016/j.joep.2014.12.004

Krupnikov, Y., & Levine, A. S. (2014). Cross-Sample Comparisons and External Validity. *Journal of Experimental Political Science*, *1*(1), 59–80. https://doi.org/10.1017/xps.2014.7

Levy, M. R., & Tasoff, J. (2020). Exponential-growth Bias in Experimental Consumption Decisions. *Economica*, *87*(345), 52–80. https://doi.org/10.1111/ecca.12306

Lugilde, A., Bande, R., & Riveiro, D. (2019). Precautionary Saving: a Review of the Empirical Literature. *Journal of Economic Surveys*, *33*(2), 481–515. https://doi.org/10.1111/joes.12284

Meissner, T. (2016). Intertemporal consumption and debt aversion: an experimental study. *Experimental Economics*, *19*(2), 281–298. https://doi.org/10.1007/s10683-015-9437-0

Meissner, T., & Rostam-Afschar, D. (2017). Learning Ricardian Equivalence. *Journal of Economic Dynamics and Control*, *82*, 273–288. https://doi.org/10.1016/j.jedc.2017.07.004

Peijnenburg, K., Nijman, T., & Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, *70*, 18–35. https://doi.org/10.1016/j.jedc.2016.05.023

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, *28*(3), 450–461. https://doi.org/10.1086/323732

Pu, B., Peng, H., & Xia, S. (2017). Role of Emotion and Cognition on Age Differences in the Framing Effect. *The International Journal of Aging and Human Development*, *85*(3), 305–325. https://doi.org/10.1177/0091415017691284

Winter, J. K., Schlafmann, K., & Rodepeter, R. (2012). Rules of Thumb in Life-Cycle Saving Decisions. *Economic Journal*, *122*(560), 479–501. https://doi.org/10.1111/j.1468-0297.2012.02502.x

Yamamori, T., Iwata, K., & Ogawa, A. (2018). Does money illusion matter in intertemporal decision making? *Journal of Economic Behavior and Organization*, *145*, 465–473. https://doi.org/10.1016/j.jebo.2017.11.019

Appendix – Main Effects and Socio-Demographic Characteristics

Table A1 – Summary of main effects in socio-demographic subsamples

		Anderhub *et al.* (2000)				Fatas *et al.* (2007)	
		Fulfil Condition 1		Fulfil Condition 2		Choice of retirement timing	
		Product	Summation	Product	Summation	Lump>Annuity	Combined>Annuity
Full sample		✗	✗	✗	✗	✓	✓
Age	≤=35	✗	✗	✗	✗	✗	✗
	>35 & ≤=50	✗	✗	✗	✗	✓	✗
	>50	✗	✗	✗	✗	✓	✓
Gender	Non-female	✗	✗	✗	✗	✓	✓
	Female	✗	✗	✗	✗	✓	✗
High education	No	✗	✗	✗	✗	✓	✓
	Yes	✗	✗	✗	✗	✗	✗
Income (Euro)	<2000	✗	✗	✗	✗	✓	✓
	2000 to 3200	✗	✗	✗	✗	✗	✗
	>3200	✗	✗	✗	✗	✓	✓
Financial training	No	✗	✗	✗	✗	✓	✓
	Yes	✗	✗	✗	✗	✗	✗

(Continued)

		Koehler *et al.* (2015)		Meissner (2016)		Blaufus & Milde (2021)	
		Retirement savings	Spending variablility	*m1*	*m2*	Savings rate (Effective savings rate)	
		Long>Short	Long>Short	BF>SF	BF>SF	Def.>(<)Imme.	Mat.>(>)Imme.
Full sample		✓	✗	✓	✓	✓(✓)	✓(✓)
Age	<=35	✓	✗	✓	✓	✓(✓)	✗(✗)
	>35 & <=50	✓	✗	✓	✓	✗(✓)	✗(✗)
	>50	✓	✗	✓	✓	✓(✓)	✓(✓)
Gender	Non-female	✓	✗	✓	✓	✓(✓)	✓(✓)
	Female	✓	✗	✓	✓	✓(✓)	✗(✗)
High education	No	✓	✗	✓	✓	✓(✓)	✓(✓)
	Yes	✓	✗	✓	✓	✓(✓)	✗(✗)
Income (Euro)	<2000	✓	✗	✓	✓	✓(✗)	✗(✗)
	2000 to 3200	✓	✗	✓	✓	✓(✓)	✗(✗)
	>3200	✓	✗	✓	✓	✗(✓)	✓(✓)
Financial training	No	✓	✗	✓	✓	✓(✓)	✓(✓)
	Yes	✓	✓	✓	✓	✗(✓)	✗(✗)

Notes: For the replication of Anderhub *et al.* (2000), Condition 1 $((\frac{x_2}{S_2}|\neg[1/6]) > (\frac{x_2}{S_2}|\neg[1/3]) > (\frac{x_2}{S_2}|\neg[1/2]))$ indicates that the spending is larger when a card deck with a low termination probability is removed than a card deck with a high termination probability is removed. Condition 2 $((\frac{x_3}{S_3}|[1/2]) > (\frac{x_3}{S_3}|[1/3]) > (\frac{x_3}{S_3}|[1/6]))$ indicates that the spending is larger when a card deck with a high termination probability finally stays than a deck with a low termination probability stays. For the replication of Fatas *et al.* (2007), the larger the choice of retirement timing is, the later a subject chooses to retire. For the replication of Koehler *et al.* (2015), the retirement savings is the savings balance after the last period of the working phase. The spending variability is the standard deviation of spending. For the replication of Meissner (2016), *m1* and *m2* are the deviations and absolute deviations from conditional optimal consumption, respectively, and the results are from Round 1. For the replication of Blaufus & Milde (2021), *Def.* is treatment Deferred, *Imme.* is treatment Immediate, and *Mat.* is treatment Matching. The savings rate is defined as the saving amount divided by the income in a period and the effective savings rate is defined as the savings rate multiplied by $(1 - tax\ rate)$. ✓ indicates that an effect is confirmed in the full sample or a subsample, and ✗ indicates that an effect is not found.

Table A2 – Anderhub et al. (2000) Replication: Reactions to first removed card deck and finally stayed card deck in soc.-demogr. subsamples

		Treatment	Mean consumption share in period 2 ($\frac{x_2}{S_2}$)			Mean consumption share in period 3 ($\frac{x_3}{S_3}$)			Obs.	Obs. fulfilling Condition 1	Obs. fulfilling Condition 2
			($\frac{x_2}{S_2}$|¬[1/6])	($\frac{x_2}{S_2}$|¬[1/3])	($\frac{x_2}{S_2}$|¬[1/2])	($\frac{x_3}{S_3}$|[1/6])	($\frac{x_3}{S_3}$|[1/3])	($\frac{x_3}{S_3}$|[1/2])			
Age (years old)	≤35	Product	0.34	0.31	0.32	0.43	0.45	0.41	67	13	12
		Summation	0.32	0.32	0.32	0.47	0.42	0.44	60	13	9
	>35 & ≤50	Product	0.34	0.35	0.39	0.49	0.41	0.43	46	4	7
		Summation	0.35	0.36	0.34	0.46	0.44	0.45	17	4	4
	>50	Product	0.32	0.35	0.36	0.42	0.38	0.37	63	2	10
		Summation	0.37	0.37	0.36	0.40	0.39	0.41	86	22	11
Gender	Non-female	Product	0.32	0.31	0.32	0.40	0.40	0.38	102	14	18
		Summation	0.36	0.35	0.34	0.44	0.42	0.46	79	20	16
	Female	Product	0.35	0.37	0.40	0.49	0.44	0.43	74	5	11
		Summation	0.34	0.35	0.34	0.43	0.40	0.39	84	19	8
High education	No	Product	0.33	0.33	0.35	0.44	0.42	0.40	115	14	17
		Summation	0.34	0.36	0.34	0.43	0.41	0.43	101	21	14
	Yes	Product	0.34	0.36	0.36	0.43	0.43	0.41	54	4	11
		Summation	0.36	0.35	0.35	0.44	0.41	0.42	59	16	10
Income (Euro)	<2000	Product	0.36	0.34	0.33	0.42	0.42	0.40	63	9	10
		Summation	0.35	0.36	0.33	0.44	0.40	0.41	57	11	7
	2000 to 3200	Product	0.33	0.33	0.38	0.45	0.44	0.42	50	5	11
		Summation	0.34	0.37	0.34	0.42	0.38	0.42	44	11	7
	>3200	Product	0.33	0.34	0.36	0.46	0.41	0.39	53	5	5
		Summation	0.37	0.35	0.37	0.46	0.45	0.47	51	14	9
Financial training	No	Product	0.34	0.34	0.36	0.45	0.43	0.40	130	15	22
		Summation	0.35	0.35	0.33	0.44	0.43	0.45	118	30	19
	Yes	Product	0.32	0.32	0.33	0.40	0.40	0.39	41	2	7
		Summation	0.36	0.36	0.38	0.43	0.37	0.37	41	9	4

Notes: The mean consumption share is computed from all the subjects in period 2 ($\frac{x_2}{S_2}$) and 3 ($\frac{x_2}{S_2}$) correspondingly. For each subject, there are two out of the six rounds where the first removed card decks (the finally stayed card deck) have the same termination probability. The consumption share in period 2 (period 3) for each subject is the mean of the shares of the two rounds with the same termination probability of first removed card deck (finally stayed card deck). Condition 1 refers to ($\frac{x_2}{S_2}|\neg[1/6]$) > ($\frac{x_2}{S_2}|\neg[1/3]$) > ($\frac{x_2}{S_2}|\neg[1/2]$), implying that the spending is larger when a card deck with a low termination probability is removed than a card deck with a high termination probability is removed. Condition 2 refers to ($\frac{x_3}{S_3}|[1/2]$) > ($\frac{x_3}{S_3}|[1/3]$) > ($\frac{x_3}{S_3}|[1/6]$), implying that the spending is larger when a card deck with a high termination probability finally stays than a deck with a low termination probability stays.

Table A3 – Anderhub et al. (2000) Replication: Statistical tests on the reactions to the first removed card deck and finally stayed card deck in socio-demographic subsamples

		Treatment	*P*-value of *t*-test						Obs.
			Null hypothesis for Condition 1			Null hypothesis for Condition 2			
			$(\frac{x_2}{S_2}\|\neg[1/6]) \leq (\frac{x_2}{S_2}\|\neg[1/2])$	$(\frac{x_2}{S_2}\|\neg[1/3]) \leq (\frac{x_2}{S_2}\|\neg[1/2])$	$(\frac{x_2}{S_2}\|\neg[1/6]) \leq (\frac{x_2}{S_2}\|\neg[1/3])$	$(\frac{x_3}{S_3}\|[1/6]) \geq (\frac{x_3}{S_3}\|[1/2])$	$(\frac{x_3}{S_3}\|[1/3]) \geq (\frac{x_3}{S_3}\|[1/2])$	$(\frac{x_3}{S_3}\|[1/6]) \geq (\frac{x_3}{S_3}\|[1/3])$	
Age (years old)	≤35	Product	0.239	0.610	0.162	0.288	0.109	0.749	67
		Summation	0.385	0.424	0.460	0.232	0.695	0.108	60
	>35 & ≤50	Product	0.956	0.913	0.639	0.064	0.612	**0.036**	46
		Summation	0.466	0.331	0.637	0.457	0.558	0.400	17
	>50	Product	0.900	0.654	0.813	0.072	0.318	0.161	63
		Summation	0.448	0.394	0.554	0.526	0.655	0.369	86
Gender	Non-female	Product	0.453	0.688	0.271	0.152	0.173	0.465	102
		Summation	0.329	0.383	0.442	0.719	0.874	0.285	79
	Female	Product	0.965	0.856	0.773	**0.030**	0.398	0.051	74
		Summation	0.498	0.381	0.617	0.131	0.424	0.176	84
High education	No	Product	0.856	0.912	0.385	**0.036**	0.233	0.141	115
		Summation	0.480	0.289	0.694	0.455	0.749	0.217	101
	Yes	Product	0.730	0.517	0.715	0.241	0.287	0.443	54
		Summation	0.357	0.504	0.353	0.351	0.631	0.237	59
Income (Euro)	<2000	Product	0.218	0.354	0.343	0.260	0.302	0.451	63
		Summation	0.300	0.177	0.657	0.181	0.522	0.167	57
	2000 to 3200	Product	0.950	0.936	0.551	0.175	0.328	0.311	50
		Summation	0.461	0.211	0.760	0.541	0.787	0.244	44
	>3200	Product	0.831	0.752	0.609	**0.024**	0.257	0.094	53
		Summation	0.440	0.716	0.235	0.557	0.626	0.430	51
Financial training	No	Product	0.892	0.843	0.593	**0.030**	0.175	0.173	130
		Summation	0.216	0.154	0.593	0.672	0.813	0.329	118
	Yes	Product	0.656	0.685	0.468	0.382	0.405	0.476	41
		Summation	0.654	0.641	0.514	0.087	0.495	0.089	41

Notes: The mean consumption share is computed from all the subjects in period 2 ($\frac{x_2}{S_2}$) and 3 ($\frac{x_2}{S_2}$) correspondingly. For each subject, there are two out of the six rounds where the first removed card decks (the finally stayed card deck) have the same termination probability. The consumption share in period 2 (period 3) for each subject is the mean of the shares of the two rounds with the same color of first removed card deck (finally stayed card deck). Condition 1 refers to $(\frac{x_2}{S_2}|\neg[1/6]) > (\frac{x_2}{S_2}|\neg[1/3]) > (\frac{x_2}{S_2}|\neg[1/2])$, implying that the spending is larger when a card deck with a low termination probability is removed than a card deck with a high termination probability is removed. Condition 2 refers to $(\frac{x_3}{S_3}|[1/2]) > (\frac{x_3}{S_3}|[1/3]) > (\frac{x_3}{S_3}|[1/6])$, implying that the spending is larger when a card deck with a high termination probability finally stays than a deck with a low termination probability stays. The *p*-values in bold and blue font are those that are smaller than 0.05.

Table A4 – Fatas, Lacomba, and Lagos (2007) Replication: Timing of Retirement Treatment Effects in Socio-demographic Subsamples

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
Risk-taking	0.023**	0.021*	0.029***	0.027***	0.024***	0.024***	0.028***	0.009	0.010	0.036***	0.024***	0.034***
	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)	(0.011)	(0.007)	(0.006)	(0.009)
Patience	0.235	0.436**	0.058	0.171	0.236*	0.186	0.194	0.155	0.352	0.131	0.242*	0.076
	(0.140)	(0.160)	(0.117)	(0.115)	(0.110)	(0.110)	(0.115)	(0.158)	(0.182)	(0.117)	(0.096)	(0.145)
Annuity	-0.682	-1.044*	-0.923*	-0.853*	-0.998**	-1.330***	-0.562	-1.499**	-0.845	-0.805*	-0.976**	-0.753
	(0.436)	(0.463)	(0.378)	(0.363)	(0.337)	(0.349)	(0.350)	(0.510)	(0.520)	(0.376)	(0.297)	(0.456)
Combined	-0.083	-0.660	-0.118	-0.004	-0.602	-0.224	-0.357	-0.483	-1.506**	0.030	-0.325	-0.196
	(0.433)	(0.495)	(0.373)	(0.370)	(0.331)	(0.334)	(0.376)	(0.493)	(0.566)	(0.368)	(0.306)	(0.437)
Constant	4.567***	4.753***	5.426***	5.095***	5.034***	5.132***	5.073***	5.645***	5.203***	5.039***	5.080***	4.952***
	(0.495)	(0.557)	(0.400)	(0.437)	(0.344)	(0.380)	(0.403)	(0.529)	(0.565)	(0.422)	(0.323)	(0.553)
(Annuity–Combined)	-0.598	-0.384	-0.805*	-0.849*	-0.396	-1.106**	-0.205	-1.016*	0.661	-0.835*	-0.651*	-0.557
	(0.443)	(0.494)	(0.372)	(0.359)	(0.345)	(0.350)	(0.362)	(0.498)	(0.560)	(0.385)	(0.300)	(0.461)
R2	0.133	0.130	0.076	0.086	0.111	0.104	0.093	0.076	0.111	0.147	0.088	0.128
Prob. >F	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.020	0.000	0.000	0.001
Observations	114	154	262	286	243	284	246	131	103	234	387	139

Notes: The results are from OLS estimations. The dependent variable is the mean retirement period chosen in the three rounds. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income. *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 and equal to 1, 2, 3 and 4 for the choice of no delay, 1 month, 2 months and 3 months, respectively. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5 – Meissner (2016) Replication: Sub-Optimal Consumption Paths (Measure 1) in Socio-demographic subsamples

				round 1	round 2	round 3	round 4	Obs.
Age (years old)	≤35	mean (m1)	BF	201.65	188.98	-61.75	13.38	57
			SF	-147.82	-95.37	320.52	303.33	60
		p-value		<0.001	<0.001	<0.001	<0.001	
	>35 & ≤50	mean (m1)	BF	61.98	71.42	-344.87	-220.76	40
			SF	-7.73	97.46	404.02	378.25	23
		p-value		0.003	0.013	<0.001	<0.001	
	>50	mean (m1)	BF	168.98	288.83	66.28	-25.31	50
			SF	-182.09	-213.50	245.41	302.34	48
		p-value		0.003	0.013	<0.001	<0.001	
Gender	Non-female	mean (m1)	BF	132.30	171.13	-83.13	-81.56	73
			SF	-127.78	-98.66	312.44	300.34	79
		p-value		<0.001	<0.001	<0.001	<0.001	
	Female	mean (m1)	BF	170.52	240.25	-114.40	-57.66	71
			SF	-147.29	-116.68	301.54	342.06	51
		p-value		<0.001	<0.001	<0.001	<0.001	
High education	No	mean (m1)	BF	181.63	229.63	-112.34	-61.36	75
			SF	-128.41	-116.47	305.74	328.43	73
		p-value		<0.001	<0.001	<0.001	<0.001	
	Yes	mean (m1)	BF	112.50	143.92	-76.28	-66.71	69
			SF	-142.27	-87.98	310.04	300.40	57
		p-value		<0.001	<0.001	<0.001	<0.001	
Income (Euro)	<2000	mean (m1)	BF	247.40	210.75	-39.78	-115.63	46
			SF	-173.80	-36.02	407.10	398.98	36
		p-value		<0.001	<0.001	<0.001	<0.001	
	2000 to 3200	mean (m1)	BF	79.49	272.58	-88.32	-11.40	41
			SF	-84.55	-89.88	350.92	327.57	31
		p-value		<0.001	<0.001	<0.001	<0.001	
	>3200	mean (m1)	BF	94.38	89.70	-169.12	-69.72	46
			SF	-149.48	-174.37	242.94	268.68	50
		p-value		<0.001	<0.001	<0.001	<0.001	
Financial training	No	mean (m1)	BF	125.46	190.89	-118.58	-72.56	115
			SF	-123.68	-102.43	308.58	338.58	95
		p-value		<0.001	<0.001	<0.001	<0.001	
	Yes	mean (m1)	BF	240.80	263.57	36.99	-7.83	28
			SF	-177.07	-115.21	297.82	251.92	34
		p-value		<0.001	<0.001	<0.001	<0.001	

Notes: Deviations from conditional optimal consumption, following the original study's *m1*. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income. *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.

Table A6 – Meissner (2016) Replication: Sub-Optimal Consumption Paths (Measure 2) in Socio-demographic subsamples

				round 1	round 2	round 3	round 4	Obs.
Age (years old)	≤35	mean (m2)	BF	510.52	531.84	326.63	364.56	57
			SF	314.38	256.40	387.06	370.01	60
		p-value		<0.001	<0.001	<0.001	0.271	
	>35 & ≤50	mean (m2)	BF	518.07	571.53	481.64	385.62	40
			SF	189.36	287.66	409.68	380.70	23
		p-value		<0.001	<0.001	0.003	0.037	
	>50	mean (m2)	BF	516.74	467.93	306.78	300.63	50
			SF	399.43	436.67	533.38	462.17	48
		p-value		<0.001	<0.001	<0.001	<0.001	
Gender	Non-female	mean (m2)	BF	549.51	526.71	367.45	376.13	73
			SF	301.69	289.76	397.29	387.16	79
		p-value		<0.001	<0.001	<0.001	<0.001	
	Female	mean (m2)	BF	473.24	482.51	334.88	305.95	71
			SF	360.06	390.15	521.98	437.55	51
		p-value		<0.001	<0.001	<0.001	<0.001	
High education	No	mean (m2)	BF	512.03	496.54	407.47	357.75	75
			SF	354.20	388.88	485.55	440.98	73
		p-value		<0.001	<0.001	<0.001	<0.001	
	Yes	mean (m2)	BF	524.92	556.51	320.51	339.47	69
			SF	284.72	251.94	394.58	362.01	57
		p-value		<0.001	<0.001	<0.001	0.028	
Income (Euro)	<2000	mean (m2)	BF	449.93	500.55	288.54	332.57	46
			SF	426.27	368.14	489.96	482.55	36
		p-value		<0.001	<0.001	<0.001	<0.001	
	2000 to 3200	mean (m2)	BF	576.14	526.44	470.50	359.82	41
			SF	223.21	237.29	356.32	336.12	31
		p-value		<0.001	<0.001	0.019	0.049	
	>3200	mean (m2)	BF	505.56	517.98	350.38	363.76	46
			SF	313.77	364.47	446.72	396.15	50
		p-value		<0.001	<0.001	<0.001	0.148	
Financial training	No	mean (m2)	BF	541.19	512.36	371.00	351.32	115
			SF	332.22	329.23	452.80	400.17	95
		p-value		<0.001	<0.001	<0.001	<0.001	
	Yes	mean (m2)	BF	429.68	485.62	303.25	318.44	28
			SF	310.37	336.83	422.65	424.82	34
		p-value		<0.001	<0.001	0.002	0.012	

Notes: Absolute deviations from conditional optimal consumption, following the original study's *m2*. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income. *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.

Table A7 – Blaufus and Milde (2021) Replication: Drivers of Saving Behaviour (Savings Rate) in Socio-demographic Subsamples

Panel A: **Treatment Deferred vs. Immediate**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
Deferred	0.163**	0.055	0.090**	0.092**	0.112**	0.115***	0.082*	0.111*	0.129**	0.068	0.119***	0.034
	(0.051)	(0.047)	(0.033)	(0.031)	(0.040)	(0.032)	(0.038)	(0.049)	(0.043)	(0.041)	(0.029)	(0.052)
Observations	670	950	1,820	2,080	1,360	2,280	1,080	940	1,040	1,240	2,630	730
Subjects	67	95	182	208	136	228	108	94	104	124	263	73
R2	0.1031	0.0177	0.0376	0.0376	0.0500	0.0510	0.0363	0.0471	0.0720	0.0216	0.0568	0.0103

Panel B: **Treatment Matching vs. Immediate**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
Matching	-0.014	0.071	0.072*	0.070*	0.027	0.066*	0.038	0.041	0.034	0.094**	0.072**	0.011
	(0.042)	(0.047)	(0.029)	(0.028)	(0.034)	(0.028)	(0.034)	(0.043)	(0.041)	(0.033)	(0.025)	(0.039)
Observations	720	810	2,070	2,180	1,420	2,390	1,170	980	1,100	1,370	2,740	800
Subjects	72	81	207	218	142	239	117	98	110	137	274	80
R2	0.0033	0.0304	0.0300	0.0273	0.0049	0.0226	0.0090	0.0092	0.0062	0.0511	0.0260	0.0057

Notes: The table presents regression results of random-effects models explaining subject's saving rates. The savings rate (SR) is defined as the saving amount in a given period divided by the income in this period. *Deferred* is a dummy variable equal to one if the observation belongs to the deferred-tax treatment. The other covariates include *Period* and the constant term. *Matching* is a dummy variable equal to one if the observation belongs to the matching treatment. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income. Standard errors clustered at subject level are reported in parentheses. *R2* is the R-squared for overall model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A8 – Blaufus and Milde (2021) Replication: Drivers of Saving Behaviour (Effective Savings Rate) in Socio-demographic Subsamples

Panel A: **Treatment Deferred vs. Immediate**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
Deferred	-0.075*	-0.117**	-0.071**	-0.078**	-0.085**	-0.067**	-0.097***	-0.071	-0.067*	-0.093**	-0.065**	-0.129***
	(0.038)	(0.037)	(0.025)	(0.024)	(0.031)	(0.025)	(0.029)	(0.040)	(0.033)	(0.031)	(0.022)	(0.037)
Observations	670	950	1,820	2,080	1,360	2,280	1,080	940	1,040	1,240	2,630	730
Subjects	67	95	182	208	136	228	108	94	104	124	263	73
R2	0.0405	0.0952	0.0354	0.0423	0.0460	0.0291	0.0748	0.0298	0.0320	0.0612	0.0284	0.1183

Panel B: **Treatment Matching vs. Immediate**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
Matching	-0.016	0.069	0.070*	0.069*	0.026	0.064*	0.036	0.040	0.033	0.092**	0.070**	0.010
	(0.041)	(0.047)	(0.029)	(0.028)	(0.034)	(0.028)	(0.034)	(0.043)	(0.041)	(0.033)	(0.025)	(0.039)
Observations	720	810	2,070	2,180	1,420	2,390	1,170	980	1,100	1,370	2,740	800
Subjects	72	81	207	218	142	239	117	98	110	137	274	80
R2	0.0035	0.0293	0.0290	0.0263	0.0044	0.0217	0.0083	0.0087	0.0057	0.0497	0.0250	0.0055

Notes: The table presents regression results of random-effects models explaining subject's effective saving rates. The effective savings rate (ESR) is defined as the saving amount in a given period divided by the income in this period and multiplied by $(1 - tax\ rate)$. *Deferred* is a dummy variable equal to one if the observation belongs to the deferred-tax treatment. The other covariates include *Period* and the constant term. *Matching* is a dummy variable equal to one if the observation belongs to the matching treatment. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income. Standard errors clustered at subject level are reported in parentheses. *R2* is the R-squared for overall model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A9 – Koehler, Langstaff, and Liu (2015) Replication: Effects of Retirement Length in Socio-demographic Subsamples

Panel A: **Retirement Savings**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
F-statistic	$F(559)=788$	$F(335)=366$	$F(479)=392$	$F(647)=934$	$F(723)=615$	$F(859)=793$	$F(479)=692$	$F(447)=390$	$F(339)=475$	$F(435)=648$	$F(1051)=1110$	$F(303)=380$
$adjusted\ R^2$	0.772	0.760	0.736	0.781	0.729	0.719	0.799	0.675	0.780	0.806	0.749	0.760
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	560	336	480	648	724	860	480	448	400	436	1052	304
Subjects	140	84	120	162	181	215	120	112	100	218	263	76

Panel B: **Variability of Spending**

	Age (years old)			Gender		High education		Income (Euro)			Financial training	
	≤35	>35 & ≤50	>50	Non-female	Female	No	Yes	<2000	2000 to 3200	>3200	No	Yes
F-statistic	$F(559)=2.96$	$F(335)=0.49$	$F(479)=0.09$	$F(647)=3.19$	$F(723)=0.35$	$F(859)=0.42$	$F(479)=0.28$	$F(447)=0.17$	$F(399)=0.00$	$F(435)=0.45$	$F(1051)=0.01$	$F(303)=4.82$
$adjusted\ R^2$	0.646	0.649	0.641	0.653	0.645	0.661	0.626	0.572	0.541	0.763	0.658	0.677
p-value	0.086	0.486	0.769	0.075	0.552	0.518	0.599	0.678	1.000	0.504	0.904	0.029
Observations	560	336	480	648	724	860	480	448	400	436	1052	304
Subjects	140	84	120	162	181	215	120	112	100	109	263	76

Notes: The table shows the effects of retirement length treatment on retirement savings and spending variability from ANOVA analyses. *High education* means Bachelor, Master or Doctoral degree. *Income* means the monthly household disposable income.

Chapter 3

Opening and Holding Tax-Incentivized Retirement Savings Accounts

ANDRE LOT KREMENA BACHMANN[1] FRANCISCO SANTOS[2]

Tax-incentivized retirement savings accounts (TIRSA) are one of the main policy tools for promotion of private voluntary retirement savings. We study a longitudinal panel of Swiss households to evaluate the determinants of TIRSA ownership, and whether these factors are also significant drivers of the decision of households who open a TIRSA. We show that several financial and socio-demographic factors that determine the likelihood TIRSA ownership (financial satisfaction, wealth, education, civil status, adult household composition) are not relevant for the households' decision to open a TIRSA.

JEL Classification: D14, G51

Keywords: household finance, retirement savings, tax incentives

[1]Zürich University of Applied Sciences, University Zürich
[2]NHH – Norwegian School of Economics

3.1 Introduction

The fundamental question of 'why people save' has eluded researcher for a long time (Browning & Lusardi, 1996; Devaney, Anong, & Whirl, 2007; Wärneryd, 1989). Standard life-cycle models (Browning & Crossley, 2001) assume that individuals have an incentive to smooth their lifetime consumption, while they only earn income in adulthood, before retirement. To smooth consumption, individuals should thus build up substantial savings while they are active in the labor force, in order to afford a commensurable lifestyle after they retire. Individual portfolio risk exposure should also adjust to the life-cycle (Cocco, Gomes, & Maenhout, 2005).

Normative model predictions of consumption smoothing require only few assumptions, such as concave marginal utility of consumption and positive intertemporal discount rate. Consumption smoothing incentives are stronger under habit formation models, which assumes a marginal disutility from a reduction on consumption that has become a reference baseline for the individual.

In most OECD countries, consumption smoothing during retirement is, in great part, provided by government-sponsored or privately managed retirement schemes with mandatory participation. Regardless of their design, these schemes automatically force a certain baseline level of consumption smoothing, even if that is achieved through intergenerational transfers through the general taxation system. Such is the case in Switzerland, which adopts a tax-incentivized retirement savings account (TIRSA) as part of its third pillar of its retirement and pension framework.

We first analyze the determinants of a household having a TIRSA. We find that several variables have a positive impact on the likelihood of having TIRSA. Namely, income, years of education, being married, divorced, or widower, all drive positively TIRSA ownership. For a sense of the economical magnitude, being married increases the odds of having TIRSA by 4.49. Additionally, self-assessed financial satisfaction and savings ability of households also have a positive impact on the ownership of TIRSA. On the other hand, we find that the amount of direct taxes paid, the average age of the household, the average number of adults in the household, and not being in the labor all contribute negatively to the household's choice to not have TIRSA. Our results in explaining the choice of having TIRSA in the cross-section are consistent with the broader results in the retirement savings literature in different countries under different institutional arrangements (as in Stinglhamber et al., 2007).

Next, we analyze whether these same determinants that explain the cross-sectional variation of TIRSA ownership can explain the initial decision of opening TIRSA – the decision to save for the first time under this scheme. We find, surprisingly, that many of the previous determinants do not explain the decision open a TIRSA. Income is the biggest driver for opening TIRSA. The higher the income, the higher the odds ratio of saving for the first time in TIRSA. Taxes are also a factor, but with a negative impact on the odds ratio of opening a TIRSA. The next determinants of the decision to open TIRSA are financial slack and, even more, savings ability. We also find that the levels of these variables drive the decision to open TIRSA, not the time-series changes of these variables. Our results in this regard are particularly interesting because the literature on what drives the initial decision to save is quite scarce. We then contribute to the literature by showing that several variables that explain the differences in savings behavior in the cross-section do not explain the initial choice of behavior. We also show that households are not opening their first TIRSA account as a result of an increase in annual income that could otherwise create easier opportunities to change the household's overall consumption rate over its income. Finally, we show that opening a TIRSA account for the first time is not predictive of future changes in income, financial satisfaction, or financial slack.

In general, existing empirical evidence suggests that individuals, on a voluntary basis, save too little for retirement (Banks, Blundell, & Tanner, 1998; Skinner & Hubbard, 1994). This *undersaving* behavior has important implications for the individual and his or her household, and also for policymakers. If a country has some form of a welfare safety net, individuals might undersave in expectation that welfare programs will eventually provide financial support at old age. Then, it could be an ex ante optimal choice to undersave (and consume more) during the working period and then draw maximum support from state-provided welfare at old age. In tandem, previous research has also identified a puzzle in the form of *retirement dip*, which is a one-time drop in the level of consumption after retirement that is difficult to reconcile with standard life-cycle models (Battistin et al., 2009).

Nonetheless, the *old-age underconsumption* puzzle suggests that, conditional on their previous savings accumulated at retirmenet, individuals drawdown their long-term savings too slowly afterwards (Love, Palumbo, & Smith, 2009), given their unconditional survival expectations. This stylized behavior directly contradicts the previous argument for strategic undersaving with the goal of maximizing the extraction of state welfare, which could then be explained by – among other drivers – a certain level of aversion to public care (Ameriks et al., 2011). Furthermore, individuals may attach some level of utility in expectation that any unspent wealth at the end of their life will be enjoyed by their heirs, that is, motivated to reduce spending in retirement due to bequest motives (the utility derived in anticipation from consumption of one's own inheritance by their next-of-kin) (Yaari, 1965). There is considerable heterogeneity on voluntary retirement savings (Bernheim, Skinner, & Weinberg, 2001), in part driven by generational effects (Brounen, Koedijk, & Pownall, 2016) and also by financial literacy (Lusardi & Mitchell, 2011) pertaining to the complex task of optimizing consumption throughout one's lifetime.

Within this context, TIRSA are the main policy tool used by governments to incentivize more long-term voluntary savings, mostly framed as provisions for retirement self-financing. The structure of TIRSA schemes vary across countries, as they usually provide some type of tax relief by immediate reduction of the taxable base, exemption of taxation for accrued capital gains, deferred taxation at lower rates, exemption from wealth and asset taxes, or a combination thereof. In exchange, a TIRSA imposes some form of temporal liquidity restriction that limits withdraws for a period of time and/or charges punitive taxes for withdraws beyond the scope of preapproved conditions (reaching retirement age, minimum TIRSA holding period, etc.). TIRSA schemes offer economic benefits that are sometimes non-intuitive (Brown, Cederburg, & O'Doherty, 2017) and difficult to assess.

Participation in these TIRSA schemes is highly heterogeneous (Engen, Gale, & Scholz, 1996), and plausibly low, given the economic incentives embedded in their tax advantages. While policymakers promote TIRSA schemes as tools to incentivize middle- and lower-middle-income households to save more (i.e. households that have some financial capacity to save, but are also likely to demand more welfare support if they do not save privately for old age), there is evidence that those who actively take advantage of TIRSA schemes tend to be wealthier, more educated, and already be in a better financial situation before they enroll in a TIRSA and derive its economic advantages from its tax relief features.

The extant literature offers conflicting explanations on why subjects would or would not participate in a TIRSA scheme and the effectiveness of these schemes (Hubbard & Skinner, 1996). One main concern is short-term precautionary savings driven by prudence in the face of uncertainty (Ventura & Eisenhauer, 2006; Van Schie, Donkers, & Dellaert, 2012; Levenko, 2020; Lugilde, Bande, & Riveiro, 2019). In anticipation of an increase in the risk of losing their jobs (for instance), individuals would increase short-term savings and decrease short-term consumption. Such a precautionary savings motive could deter individuals commiting to an illiquid TIRSA that makes them unable to access their savings should they suffer an income shock before withdrawals are allowed. As such, the economic benefits of a TIRSA might not be enough to overcome the impact of prudence on savings behavior.

Although the existing literature has studied the cross-sectional determinands of participation in TIRSA schemes, and the effects of randomized control trials and other interventions aimed at promoting new TIRSA schemes (Dolls et al., 2018) or changing the decision architecture surrounding retirement investment planning in TIRSA accounts (Duflo et al., 2006), less is known about the characteristics and drivers that motivate individuals to open a TIRSA. Furthermore, many long-term decisions about the financing of retirement consumption are taken at the household level, while most TIRSA schemes are offered on an individual basis with respect to their implicit tax benefits. In this paper, we study what drives individuals in households to make their first deposit on a TIRSA.

Engaging in a TIRSA scheme for the first time is a decision with some particular features that might get obfuscated in pooled cross-sectional analyses. It could require clearance of any informational and financial literacy hurdles (Lusardi & Mitchell, 2011) that individuals might have with respect to the TIRSA scheme, which are much less relevant when an individual already uses a TIRSA and increases its investment in it at a later time. The decision to open a first TIRSA also does not rely on a habit of saving or, where applicable, reliance on commitment devices embedded in certain schemes.

Moreover, as most TIRSA schemes restrict the ability to withdraw investments from the accounts, a cross-section of TIRSA account ownserhip conceals the fact that individual and household conditions that led to the opening and deposits in the TIRSA might have changed substantially over the possible many years lapsed since the original investment was made. Withdraw restrictions conceal possible preference for non-participation in a TIRSA. Studying the first opening of a TIRSA allows us, to a certain extent, to bypass these concerns and analyze the decisions of potential savers. We investigate the decision to open (or not) a TIRSA among individuals in households that have the ability to save – given the TIRSA institutional setup and their time-variant financial situation – while having not owned a TIRSA before.

The remainder of this paper is organized as follows: Section 3.2 presents the specific institutional framework of a TIRSA scheme in Switzerland, Section 3.3 discusses the survey data, key variables, and their descriptive statistics. Section 3.4 outlays the empirical analysis, and Section 3.5 concludes.

3.2 Tax-Incentivized Voluntary Savings Accounts in Switzerland

In this section, we provide a brief institutional overview of the tax-incentivized voluntary savings account scheme in Switzerland, where it is known as 'pillar 3a'.

The Swiss pension system consists of three basic pillars. The first pillar is a federal old age pension system (OASI), which provides a variable amount based on years of contribution and salary at the time of contribution. Participation is mandatory, and benefits are approportioned on basis of years of contribution. The first-pillar pension is not means tested.

The second pillar relies on private occupational pension funds with mandatory participation through individual pension rights. Second-pillar accounts accrue individually through contribution from employee and employer externally managed at the sector or employer level. There are statutory employee and employer contribution levels. In some circumstances, employees might elect to increase their contribution above the statutory percentage of their gross income. Upon retirement, employees can withdraw their balance as a lump sum or keep the benefits to be collected as annuities, with or without survival benefits. Generally, the discount rates for redeeming the pension as a lump sum from second-pillar pension funds are substantially higher than the neutral actuarial valuation of the accrued pension rights.

The third pillar consists of voluntary contributions. These can be restricted and carry tax advantages (pillar 3a) or unrestricted savings accounts (pillar 3b) without withdrawal restrictions or tax advantages. In this paper, we focus on the former '3a' Tax-Incentivized Retirement Savings Accounts

(TIRSA).

A TIRSA can be opened by wage earners, salaried workers, self-employed individuals, cross-border commuters, persons temporarily out of the workforce (for instance, due to military service), and in a few other special situations.

The amounts contributed to a TIRSA are directly deductible from the investor's annual income tax base. This implies that the tax benefit of a TIRSA depends on the marginal tax rate of each potential holder. TIRSA contributions must happen within the calendar year for which the holder is claiming the associated tax benefit. In Switzerland, personal income taxes have a federal standardized bracket structure component and cantonal add-ons that vary considerably in terms of the applicable bracket step thresholds and marginal rates.

There are annual caps for contributions to TIRSA. For salaried or wage-earner employees, who have a second-pillar occupational pension, the annual contributional cap to a TIRSA is CHF 7 056 as of 2023. For self-employed persons (without a second-pillar pension), the annual cap is CHF 35 280. If a self-employed person becomes employed, there are mechanisms through which the TIRSA balance can be transferred as a buy-in to the second-pillar occupational pension fund.

TIRSA funds can be invested in managed accounts or vetted mutual funds. Banks, brokerage firms, and other financial service providers can register vetted investment vehicles and make them available to their TIRSA clients. These eligible funds and managed accounts available to TIRSA investors have restrictions in terms of risk, fees, and costs that are more stringent than those normally applicable to retail investment products in Switzerland.

Funds invested on a TIRSA can be withdrawn after the minimum retirement age (as of 2023, 64 and 65 years for women and men, respectively), for purchase of a home with intent or being the primary residency of the owner, for opening of a business, transition to self-employment, or transfer to a second pillar fund in case of missing contributions. In practice, retirement or acquisition of real estate accounts for an overwhelming majority of all cases of the use of the said funds.

Upon withdrawal of funds from a TIRSA, a flat tax is charged. This flat tax varies by canton, and it is lower than the standard income tax for the lowest bracket. In most cantons, the withdraw flat tax is around half the first bracket tax. This structure of the embedded tax incentive means that, ex ante, individuals who are closer to retirement and are taxes at the highest marginal rate have the most direct financial benefits to accrue from participating in the TIRSA scheme: they will have few years to capitalize the difference between standard and reduced taxation, and already have a larger starting gap between both rates when investing in a TIRSA – as it is standard in similar schemes in other jurisdictions.

3.3 Data

We use data from the Swiss Household Longitudinal Panel (FORS - Swiss Centre of Expertise in the Social Sciences, 2022, hereby SHP),[3] which has assembled an overlapping panel representative of the Swiss population since 1999. More specifically, we use data from personal, household, and proxy survey responses collected in waves 1 through 22 (1999-2020),[4] augmented with matched imputed household income (available from 1999-2018) and wealth (available only 2012 and 2016).

Households are recruited to the SHP periodically, through large recruitment drives every few years (in 1999, 2004, 2013 and 2020), in addition to small intake of new subjects in other years. When recruited, subjects respond to an intake questionnaire. Then, they are elicited on hundreds of structured questions that are asked every year, part of questions allocated through recurring blocks and themes that are rotated every few waves. New questions and themes are also added and dropped across waves. The SHP has natural attrition (households that drop off the panel over time), but does not impose a strict limit on how long a household can remain in the panel.

The SHP has questionnaires regarding the household and its individual members. Each household on the SHP also has a 'reference person': an individual subject who answers the main questionnaire for the household. Minors, persons with certain disabilities, and similar conditions might have questions about themselves asked through proxy questionnaires to other members of their household.

The output variables of each SHP wave consist of both original answers to the survey questionnaires

[3] Managed by FORS.
[4] Each wave is a yearly repeated data gathering process over a few months, not necessarily over a single calendar year

and also constructed variables following standard practice on longitudinal survey methodology. In this paper, we refer to the original variable codes as `SHP-[questionnaire][*][question]`, whereas '*' is a standout for the specific wave number, and otherwise match the SHP codebook.

The reference person in each household is not necessarily the person making financial decisions on behalf of the household (if the household has more than one person). Using information on who manages the finances in the household[5], we keep in sample the single-person households, the reference person her/himself makes decisions alone, or the decisions are made jointly between the reference person and another member of the household. Thus, when matching variables elicited at the household and personal level, we remove households where the reference person does not manage the finances of the household.

Our main original variable of interest is a binary indicator on whether a member of the household saves in a TIRSA.[6] We then identify, on each wave, households that did not have a TIRSA in the previous wave, and own one in the current wave, attributing an indicator of value one to such a household as a *new saver*. Trivially, we can only identify new savers when they previously reported not having such an account.

In Table 3.1, we show the frequency of the subjects in our sample according to their TIRSA status and changes from the previous household × wave. *N* shows the number of households in each SHP wave with non-missing TIRSA ownership information. Recruitment of new households to participate in the panel happens mostly through big drives every few years, which reflects the spikes in *Enter* – the number of households that appear for the first time in the sample – in 1999, 2004, 2013 and 2020.

Due to the rules that govern the withdrawal from TIRSA (see Section 3.2), we exclude households where *age*[7] of the household's reference person is greater than 66. Although specific regulations vary, the TIRSA scheme is geared toward people who obtain labor or income from work in general. Overall, the sample gets a larger proportion of older persons who remain in the panel for many years. These observations are excluded from all further analyses. Across all waves, we observe

[5] `SHP_H*F01`
[6] `SHP_H*I22`
[7] `SHP_AGE*`

Table 3.1 – Panel subject enrollment and TIRSA ownership. Frequency of households for each wave (year) of the Swiss Household Survey panel. *N* is the total number of subjects in the sample with information on TIRSA ownership on each yearly wave. *Enter* are subjects who appeared for the first time in the panel. *Exit* are subjects who appeared for the last time in the panel. *Old age* are those 66 years or older. *New* subjects have opened a TIRSA in that year and did not have a TIRSA in the previous year. *Converted* are subjects who did not have a TIRSA when they entered the panel, opened a TIRSA since then and still have a TIRSA opened in the current year. *Continued* are subjects who already had a TIRSA when they entered the panel and still own it in the current year. Subjects who did not have and did not open a TIRSA in the current year stated they *Cannot afford* to, or had *Other* reasons not to open a TIRSA. *Withdraw* are subjects who closed their TIRSA (for statutory reasons such as reaching the minimum age, buying a residence with intent of establishing residence, opening an own business, or definitive immigration).

wave	N	enter	exit	old age	new	converted	continued	cannot afford	other	withdraw
1999	4,081	4,081	448	612			2,301	540	1,195	
2000	3,548	321	288	576	305		1,707	504	972	297
2001	3,316	144	322	554	297	160	1,333	378	915	254
2002	2,799	107	164	511	212	245	1,023	314	777	219
2003	2,484	47	255	466	183	305	808	302	639	194
2004	4,323	2,215	561	841	154	325	1,706	503	1,280	147
2005	3,359	146	261	718	263	329	1,180	354	920	209
2006	3,309	103	297	730	189	438	954	299	972	208
2007	3,314	87	216	788	150	512	774	304	1,023	188
2008	3,270	62	176	796	154	554	654	286	1,039	180
2009	3,366	59	157	885	131	605	564	332	1,099	195
2010	3,385	61	175	957	194	641	524	315	1,054	140
2011	3,323	40	185	982	145	719	469	312	1,046	178
2012	3,247	30	201	1,029	156	765	419	278	1,042	153
2013	6,454	3,379	837	1,900	142	789	2,303	670	1,941	160
2014	5,643	292	596	1,760	252	802	1,743	498	1,807	377
2015	5,060	102	524	1,690	245	889	1,325	419	1,662	268
2016	4,628	59	537	1,674	211	968	1,051	385	1,521	246
2017	4,314	69	460	1,623	174	1,033	859	373	1,418	218
2018	4,192	76	554	1,579	149	1,121	754	347	1,428	219
2019	3,970	53	678	1,533	167	1,155	662	327	1,387	198
2020	7,445	3,804		2,289	136	1,188	2,931	606	2,318	167
Total	88,830	15,337	7,892	24,493	4,009	13,543	26,044	8,646	27,455	4,415

4009 instances of households that open a TIRSA account, becoming a new saver. On every wave, those who report that they did not have a TIRSA are asked for an explanation for not saving in such account:[8] "because you cannot affort it", or "for another reason". They are reported in the respective columns of Table 3.1. A few reference persons, on each wave, do not have a TIRSA while being elibigle to open one, but did not answer this question on the motivation of not having done it. The SHP does not have data on the individual balances invested in a TIRSA.

Each household can only become a new saver once, and is only identified as such in the first

[8] `SHP_H*I23`

instance of opening an account having not held a TIRSA in the previous wave. Therefore, between the years in which there are large recruitment drives, the pool of latent savers who can open their first TIRSA having not held a TIRSA before dwindles. On average, across all waves, there were 36 101 instances of households that could have opened a TIRSA and become new savers but did not (8646 because of being unable to afford it, 27 455 for other reasons). Therefore, on average across all waves, the conversion rate of non-savers (households without a TIRSA in previous waves) into new savers is 11.1% per year.

The count of households converted from non-savers to savers that keep their TIRSA opened grows over time, mechanically, as households remain in the panel long after being observed opening a TIRSA. Meanwhile, a larger number of households already reported ownership of a TIRSA when they entered the panel. In this case, we cannot observe when they opened an account or assess what might have triggered this decision.

Some households enter the panel, but have some missing data for intermediate waves, returning more answers in later waves. In such a case, we only consider the household to have entered the panel at the earliest wave and to have left the panel at the latest wave for which there are data for them. For this reason, the difference of total households observed at each wave is not merely the difference between new households that entered the panel minus the ones that exit the panel, compared to the previous wave.

In Table 3.2, we divide the sample according to TIRSA ownership and assess their distribution according to each category of four demographic characteristics. Reference persons older than 66 years are excluded, as stated, to the extent that age mechanically drives the institutional setup of who can open and withdraw a TIRSA.

The contribution to household income[9] is different between households that own or do not own a TIRSA: 51% of households that do not own a TIRSA are single-income, which is the status of 36% of those who own it. Similarly, of the reference persons whose households do not own a TIRSA, 43% are married, while a higher share of 61% of households that have a TIRSA are married. There are two features plausibly contributing to these differences: the so-called ‘marriage tax penalty’

[9] SHP_H*I57

Table 3.2 – Individual profile according to TIRSA ownership. Distribution (% of observations) of household × wave observations, according to TIRSA ownership, over household (income contribution) and individual reference persons (civil status, birhtplace and working status) demographic characteristics. Reference person is the person who answers the main questionnaire on behalf of each household, each year (wave).

	TIRSA ownership	
	No	Yes
Contribution to household income		
one earner only	50.70	35.96
one main earner with supplement from others	27.27	41.38
equally from two or more earners	17.57	20.24
other	4.47	2.41
Civil status		
single, never married	29.11	21.45
married	42.76	60.50
separated	3.89	2.15
divorced	19.90	13.01
widower/widow	4.23	2.60
registered partnership	0.11	0.26
dissolved partnership	0.01	0.02
Birthplace		
in Switzerland	74.32	86.43
outside Switzerland	25.68	13.57
Working status		
active occupied	72.29	87.19
unemployed	3.37	1.33
not in labor force	24.34	11.48
Sex		
Man	35.23	40.27
Woman	64.77	59.73

(through which married couples have a higher joint income tax liability than they would if they filled as single persons) and the income differences brought upon the presence of two earners in the household.

Households with TIRSA have fewer reference persons born outside Switzerland[10] (14%) than those without a TIRSA (26%). In addition to its effects on income and ability to save, being an immigrant could discourage some households from opening a TIRSA if they have plans to move out of Switzerland to their native countries once they retire. Finally, the distribution according to work status[11] is trivial as expected: 11% of those who own TIRSA are in the labor force, contrasting with 24% of those who do not. The distribution of owners and non-owners of TIRSA does not seem to be

[10] SHP_P*D160
[11] SHP_WSTAT*

somehow similar with respect to *sex*[12] of the reference person.

Financial variables are imputed in the SHP through the processing of a larger set of structured questions, aimed at eliciting information that is possibly difficult for individuals to answer immediately upon being asked. Through this process, the SHP checks the internal consistency of the reported measures and then attritutes values that consider, among other factors, the systematic mistakes that the SHP participants make. Through this process, the SHP imputes Swiss Franc values for yearly net household income,[13] disposable household income,[14] direct taxes,[15], real estate equity[16] and household wealth other than real estate.[17]

Considering the non-normal distribution of the household financial variables. We transform the original values of net household income, thus defining *Income* as the inverse sine hyperbolic log transformation of original values (Pence, 2006; MacKinnon & Magee, 1990), as:

$$Income = \ln\left(net\ disposable\ income + \left(net\ disposable\ income^2 + 1\right)^{\frac{1}{2}}\right) \tag{3.1}$$

Likewise, *Taxes* and *Wealth* are defined as similar transformations of the values of direct taxes and wealth other than real estate.

We use three measures of self-assessment of the finances of the household. The first measure is the *financial satisfaction* of the reference person with respect to the financial situation of the household,[18], measured on a scale of 0-10, with extremes explicitly defined at 0 – “not at all satisfied” and 10 – “completely satisfied”. The second measure is *financial slack* on the household budget,[19] also measured on a 0-10 scale as subjects assess “how do you manage on your household current income”, with 0 defined as “with great difficulty” and 10 defined as “very easily”. The final measure is *savings ability*, which we construct as a binary indicator that takes the value of one for the first of

[12] SHP_SEX*
[13] SHP_I*HTYNI
[14] SHP_I*DISPYI
[15] SHP_I*HTAXI
[16] SHP_WEALTH*HI
[17] SHP_WEALTH*OTI
[18] SHP_H*I30
[19] SHP_H*I51

the original response categories of assessment of income and expenses*sex*[20] (household that "can save money") and zero otherwise (household that "spends what it earns", "eats into its assets and savings" or "gets into debt").

As a proxy for prudence, we use *unemployment risk* as the self-assessed risk, by the reference person, of losing his or her own job within the next 12 months.[21] Unemployment risk is assessed on a 0-10 scale, whereas 0 is "no risk at all" and 10 is "a real risk". Finally, in the main analysis, *Age* and *Education* are the average (in years) of the adult members of each household.

3.4 Empirical Analysis

We start our analysis by looking at how the characteristics of the households differ between the households with and without TIRSA. Table 3.3 shows that variables related to income and wealth are significantly different. We see that households with a TIRSA have higher net income, disposable income, real estate equity, and wealth beyond real estate. Naturally, these households also pay more direct taxes due to their higher income and wealth. Note that information on wealth is scarce, as many of the households surveyed do not report wealth. The differences are statistically (except for wealth other than household) and economically significant. For example, the net income of households with TIRSA is approximately 50% higher than that of households without TIRSA.

Table 3.3 also shows that households with TIRSA have a higher level of education and a higher number of adults and minors in the household. No differences are observed in the average age in adults of households with and without TIRSA. Of the subsample of households that are employed in a given year, we see that households without a TIRSA self-assess a higher risk of being unemployed within a year. Lastly, Table 3.3 shows that there are significant differences in savings ability, financial slack, and financial satisfaction between TIRSA and non-TIRSA households. TIRSA households are more likely to be able to save from their income, are more likely to be able to manage their household budget, and in general, are more satisfied with the financial situation of the household.

The univariate analysis in Table 3.3 indicates that households with TIRSA have higher income

[20] SHP_H*I50
[21] SHP_W*101

Table 3.3 – Household characteristics and TIRSA ownership. Descriptive statistics of variables for household characteristics according to TIRSA ownership. Yearly household income, taxes, and wealth measures, in Swiss Francs, are imputed for each household from their components elicited in the survey. *Age* and *Education* are average (years) across the adults in the household. *Adults* and *Children in household* are the number of adults (18 or older) and minors (17 or younger) living in the household. *Unemployment risk* is the self-assessed risk of losing your job within the next 12 months (0-10 scale, higher means higher risk). *Savings ability* is an indicator equal to one if the household regularly spends less than its combined broad income. *Financial slack* is the self-assessed manageability of the household budget (0-10 scale, higher means more slack). *Financial satisfaction* is the self-assessed satisfaction with the financial situation of the household (0-10 scale, higher means more satisfied). All household × year observations with non-missing obs. for each variable are pooled. The last column shows t-statistics for means tests of each variable over TIRSA ownership status.

	TIRSA ownership (mean)					
	No	Yes	All	s.d.	N	t-stat (diff.)
Net household income	79,388	120,623	108,759	95,431	55,811	-47.122***
Disposable household income	64,536	95,88	87,529	67,779	46,644	-45.109***
Direct taxes	9,718	17,530	15,449	31,955	46,644	-23.480***
Real estate equity	181,637	387,535	340,664	2,468,212	5,122	-2.505*
Wealth other than real estate	237,566	332,252	310,697	2,376,053	5,122	-1.196
Age	48.32	48.18	48.22	10.34	63,420	1.635
Education	13.01	14.09	13.79	2.89	62,947	-42.619***
Adults in household	1.79	2.06	1.98	0.85	63,430	-35.804***
Children in household	0.50	0.70	0.64	1.00	63,430	-21.890***
Unemployment risk	2.32	1.91	2.01	2.48	48,147	15.493***
Savings ability	0.36	0.65	0.57	0.49	62,355	-66.653***
Financial slack	6.32	7.68	7.30	2.21	63,319	-72.808***
Financial satisfaction	6.45	7.63	7.30	2.03	63,366	-67.912***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

and wealth, are more educated, are closer to a traditional family structure, and are generally more satisfied with their financial situation.

Next, we perform a multivariate analysis of the drivers of having a TIRSA. First, we perform a logit regression with random effects where the dependent variable is an indicator equal to one if the household has a TIRSA in that year. Table 3.4 presents the results.

In column 1, we see that – as expected – income increases the likelihood of having a TIRSA. However, paying higher taxes reduces the odds of a TIRSA. It could be that the fiscal benefit of a TIRSA is small when the total tax paid is very large. Some sociodemographic variables are also important to explain why some households decide to have a TIRSA. Age has a negative impact, which can be explained in part by TIRSA withdrawals for alternative purposes, such as opening a business or purchasing residential real estate. Unemployment or not being in the labor force negatively

Table 3.4 – TIRSA ownership determinants. Coefficients from panel logit regressions with random effects. The dependent variable is an indicator equal to one if a household owns a TIRSA in a given wave (year) of the Swiss Household Panel. *Income*, *Taxes* and *Wealth* are inverse hyperbolic sine log-transformations of imputed yearly household income, wealth and direct tax liabilities. *Financial satisfaction* is the self-assessed satisfaction with the financial situation of the household (0-10 scale, higher means more satisfied). All household × year observations with non-missing obs. for each variable are pooled. *Financial slack* is the self-assessed manageability of the household budget (0-10 scale, higher means more slack).*Savings ability* is an indicator equal to one if the household regularly spends less than its combined broad income. *Unemployment risk* is the self-assessed risk of losing your job within the next 12 months (0-10 scale, higher means higher risk). *Age* and *Education* are averages (in years) of adult members of the household. See Table 3.3 for a description of the remaining variables. Robust standard errors in brackets.

dep. var: TIRSA ownwerhip	(1)	(2)	(3)	(4)
Income	1.095***	0.967***	1.284***	1.241***
	[0.066]	[0.067]	[0.081]	[0.170]
Taxes	-0.159***	-0.157***	-0.245***	-0.062
	[0.016]	[0.016]	[0.024]	[0.038]
Wealth				0.050***
				[0.009]
Financial satisfaction		0.077***	0.107***	0.130**
		[0.017]	[0.021]	[0.063]
Financial slack		0.077***	0.093***	0.085
		[0.017]	[0.020]	[0.059]
Savings ability		0.534***	0.483***	0.886***
		[0.057]	[0.066]	[0.189]
Unemployment risk			-0.008	
			[0.012]	
Age	-0.013**	-0.015***	0.000	-0.028**
	[0.005]	[0.005]	[0.006]	[0.013]
Education	0.205***	0.185***	0.188***	0.115***
	[0.019]	[0.019]	[0.021]	[0.036]
Married	1.502***	1.480***	1.675***	0.843***
	[0.141]	[0.140]	[0.159]	[0.283]
Separated	0.329	0.550**	0.547**	-0.289
	[0.225]	[0.224]	[0.265]	[0.513]
Divorced	0.711***	0.796***	0.812***	0.465
	[0.176]	[0.174]	[0.203]	[0.290]
Widower	0.674***	0.635**	0.977***	-0.586
	[0.260]	[0.254]	[0.330]	[0.463]
Registered partnership	0.889	1.230	1.204	1.101
	[0.925]	[0.922]	[1.031]	[1.391]
Adults in household	-0.163***	-0.111*	-0.281***	-0.161
	[0.058]	[0.058]	[0.065]	[0.127]
Children in household	-0.054	-0.003	-0.132**	0.135
	[0.054]	[0.055]	[0.061]	[0.122]
Unemployed	-0.421***	-0.209		-0.704
	[0.159]	[0.163]		[0.594]
Not in labor force	-1.026***	-0.978***		-1.849***
	[0.084]	[0.085]		[0.257]
Man	-0.019	0.012	0.154	0.246
	[0.092]	[0.092]	[0.107]	[0.179]
constant	-21.908***	-20.354***	-26.364***	-26.861***
	[1.230]	[1.224]	[1.460]	[3.208]
$\ln(\sigma_v^2)$	2.303***	2.218***	2.424***	2.294***
Observations	43195	42577	35048	4891

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

affects the likelihood of having a TIRSA. This is also expected, as these households are more likely not to be in a situation where they can lock in some of their wealth for a long period as required when investing in a TIRSA. Column 1 also shows that the number of adults in the household also negatively impacts the choice of having a TIRSA. Regarding variables that increase the likelihood that a household has a TIRSA, we see that years of education, being married, divorced, or widower increase the odds of having a TIRSA. For a sense of the economical magnitude of these effects, an additional year of average education of the household increases the odds of having an account by 1.22. Being married increases the odds by 4.49.

In column 2 of Table 3.4, we add three variables: financial satisfaction, financial slack, and savings ability of the household. All three variables are significant and positively impact the likelihood of having a TIRSA. An extra point on the 0-10 scale for financial satisfaction and slack increases the TIRSA odds by 1.08. The impact is significantly higher for savings ability – being able to save increases the TIRSA ownership odds by 1.70. This is a large impact, yet still significantly smaller than being married. Relative to column 1, the impact of the other remaining variables remains roughly the same, except for adults in the households that become marginally significant.

In column 3 of Table 3.4, we investigate whether unemployment risk plays a role in TIRSA. We only have a self-assessed unemployment risk for currently employed individuals, so the number of observations drops in this specification. We see that the unemployment risk does not drive the choice of having a TIRSA. Compared to previous specifications, the only new result that appears is the negative impact of the number of children in the household. Finally, in column 4 we add wealth. Due to the lack of data on wealth, the number of observations drops dramatically. Nonetheless, we see that wealth drives positively TIRSA. As in previous specifications, income, financial satisfaction, savings ability, and being married all have a positive impact on the odds of having a TIRSA. Only age and not being in the labor force reduce the odds of having a TIRSA.

The results in Table 3.4 are consistent with the literature (Stinglhamber et al., 2007). We contribute to the savings literature by analyzing what drives the choice of opening a first TIRSA among those who are able, by institutional design, to do so. For that purpose, we run again a logit regression with random effects, but here our dependent variable is an indicator variable equal to one if a household did not own a TIRSA in the previous year but owns it in the following year. Our sample

Table 3.5 – First-time savers: opening a TIRSA. Coefficients from panel logit regressions with random effects. The dependent variable is an indicator equal to one if a household did not own a TIRSA in the previous wave (year) and opened a TIRSA in the current wave of the Swiss Household Panel. *Income*, *Taxes* and *Wealth* are inverse hyperbolic sine log-transformations of imputed yearly household income, wealth and direct tax liabilities. *Financial satisfaction* is the self-assessed satisfaction with the financial situation of the household (0-10 scale, higher means more satisfied). All household × year observations with non-missing obs. for each variable are pooled. *Financial slack* is the self-assessed manageability of the household budget (0-10 scale, higher means more slack).*Savings ability* is an indicator equal to one if the household regularly spends less than its combined broad income. *Unemployment risk* is the self-assessed risk of losing your job within the next 12 months (0-10 scale, higher means higher risk). *Age* and *Education* are averages (in years) of adult members of the household. See Table 3.3 for a description of the remaining variables. Robust standard errors in brackets.

dep. var: opening of first TIRSA	(1)	(2)	(3)	(4)
Income	0.331***	0.233***	0.332***	0.230
	[0.059]	[0.058]	[0.065]	[0.179]
$\Delta_{(t,t-1)}$Income	0.012	0.033	0.066	0.053
	[0.049]	[0.044]	[0.059]	[0.160]
Taxes	-0.055***	-0.056***	-0.092***	-0.076*
	[0.015]	[0.014]	[0.017]	[0.042]
Wealth				0.001
				[0.013]
Financial satisfaction		0.023	0.032	0.017
		[0.026]	[0.030]	[0.082]
Financial slack		0.066***	0.053*	0.163**
		[0.024]	[0.028]	[0.074]
Savings ability		0.293***	0.259***	0.070
		[0.075]	[0.083]	[0.246]
Unemployment risk			0.005	
			[0.015]	
Age	-0.023***	-0.024***	-0.025***	-0.010
	[0.004]	[0.004]	[0.005]	[0.015]
Education	-0.012	-0.016	-0.016	0.008
	[0.015]	[0.015]	[0.016]	[0.045]
Married	-0.049	-0.073	-0.122	0.473
	[0.115]	[0.118]	[0.124]	[0.397]
Separated	0.287	0.401**	0.311	1.235**
	[0.184]	[0.191]	[0.206]	[0.596]
Divorced	0.044	0.123	0.130	0.669*
	[0.124]	[0.128]	[0.134]	[0.401]
Widower/widow	0.302	0.278	0.500**	0.074
	[0.187]	[0.188]	[0.203]	[0.773]
Registered partnership	-0.556	-0.317	-0.366	
	[0.439]	[0.477]	[0.392]	
Adults in household	0.043	0.100*	0.073	0.278*
	[0.053]	[0.053]	[0.059]	[0.156]
Children in household	-0.102**	-0.063	-0.066	-0.352*
	[0.043]	[0.042]	[0.046]	[0.182]
Unemployed	0.155	0.282		0.542
	[0.196]	[0.205]		[0.583]
Not in labor force	-0.082	-0.061		-0.362
	[0.100]	[0.101]		[0.302]
Man	0.033	0.067	0.160*	0.226
	[0.081]	[0.081]	[0.086]	[0.233]
constant	-6.830***	-5.453***	-6.898***	-7.304**
	[1.058]	[1.027]	[1.151]	[3.199]
$\ln(\sigma_v^2)$	-0.914***	-0.904***	-1.184***	-0.490
Observations	9064	9018	6785	902

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

now comprises households that never had a TIRSA and households that opened a TIRSA. However, the latter households are excluded from the sample after the year when they initiate a TIRSA. The goal is to study what motivates people to start saving in these accounts. Results are presented in Table 3.5, where the regression specifications follow the same structure as in Table 3.4. The only difference is the addition of the change in income from the previous year, which indicates whether a one-period increase in household income contributes to the decision to use part of this change in income level to increase savings through a TIRSA.

Columns 1 to 3 in Table 3.5 provide consistent results. Income is the biggest driver for opening TIRSA. The higher the income, the higher the odds ratio of saving for the first time in TIRSA. Taxes are also a driver, but with a negative impact on savings. The higher the direct taxes paid, the less likely to open TIRSA. This is a puzzling result. One could argue that for households that pay very high taxes, the marginal benefit of TIRSA is lower, and thus inertia or availability of other tax-saving products could make these households shy away from TIRSA. However, this reasoning would also apply to high-income households, for whom we do not see the same effect there. The next determinants of the decision to open TIRSA are financial slack and, even more, savings ability. As before, the impact of savings ability is particularly high – being able to save increases the TIRSA odds by 1.29. Of the remaining variables included, the only consistently significant variable is age, which negatively affects the choice to open TIRSA.

In column 4 of Table 3.5, we add wealth as an explanatory variable. Unfortunately, this reduces our sample size to only 902 observations. This probably explains the lack of statistical significance for most of the variables in column 4. In particular, the coefficient of income is not significant.

Table 3.5 shows that several variables that appear to explain the cross-sectional variation in having TIRSA do not explain the decision to open TIRSA. For example, we see that being married or not being in the labor force are economically and statistically significant in Table 3.4, but not in Table 3.5. In unreported results, we tested different specifications. Instead of using the values of these variables in the year that the household opened TIRSA, we used the change from the previous year. The results remained unchanged from the ones in Table 3.5. We also used changes in income from the last 3 and 5 years. For 3-year changes, this did not change the results. For 5-year changes, this dramatically reduces the sample size, which results in problems in statistical power.

Table 3.6 – Effects of opening a TIRSA. Coefficients of of panel OLS (1-3) and panel logit (4) regressions of the depedent variables in the upper row. On (1) the dependent variable Δ *Income* is the one-year change in the inverse hyperbolic sine log-transformation of yearly net household income. On (2), the dependent variable *financial satisfaction* is the self-assessed satisfaction with the financial situation of the household (0-10 scale, higher means more satisfied). On (3), the dependent variable *financial slack* is the self-assessed manageability of the household budget (0-10 scale, higher means more slack). On (4) the depedent variable *Savings ability* is an indicator equal to one if the household regularly spends less than its combined broad income. $Newsaver_{t-1}$ is an indicator equal to one if a household did not own a TIRSA in the previous wave (year) and opened a TIRSA in the current wave of the Swiss Household Panel. *Age* and *Education* are averages (in years) of adult members of the household. See Table 3.3 for definition of the other variables. Robust standard errors in brackets.

	(1) Δ Income	(2) Fin. satisfaction	(3) Fin. slack	(4) Savings ability
New saver$_{t-1}$	0.016	0.002	-0.029	-0.024
	[0.021]	[0.038]	[0.038]	[0.069]
Age	-0.001***	0.022***	0.017***	-0.014***
	[0.000]	[0.002]	[0.002]	[0.004]
Education	0.001	0.115***	0.142***	0.161***
	[0.001]	[0.007]	[0.007]	[0.012]
Married	-0.115***	0.159***	0.059	0.232**
	[0.009]	[0.053]	[0.055]	[0.095]
Separated	-0.171***	-0.740***	-0.952***	-1.088***
	[0.025]	[0.105]	[0.102]	[0.160]
Divorced	-0.037***	-0.292***	-0.435***	-0.514***
	[0.009]	[0.067]	[0.071]	[0.113]
Widower/widow	-0.039**	0.185*	0.052	0.080
	[0.017]	[0.105]	[0.106]	[0.188]
Registered partnership	-0.090***	-0.037	-0.282	-0.811
	[0.034]	[0.248]	[0.247]	[0.593]
Dissolved partnership	0.001	0.411***	0.455***	
	[0.007]	[0.127]	[0.128]	
Adults in household	0.084***	0.031*	-0.012	-0.037
	[0.004]	[0.017]	[0.018]	[0.035]
Children in household	0.035***	-0.113***	-0.139***	-0.294***
	[0.003]	[0.018]	[0.019]	[0.033]
Unemployed	-0.158***	-1.091***	-1.178***	-1.111***
	[0.049]	[0.088]	[0.084]	[0.129]
Not in labor force	-0.099***	-0.180***	-0.361***	-0.926***
	[0.011]	[0.032]	[0.035]	[0.066]
Man	0.000	-0.120***	-0.105***	0.005
	[0.005]	[0.037]	[0.040]	[0.066]
constant	-0.027	4.761***	4.809***	-0.656***
	[0.020]	[0.137]	[0.151]	[0.253]
Wald χ^2	908.17***	46607.12***	108371.90***	665.61***
Observations	44116	48511	48482	47610

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Knowing which households open TIRSA allows us to investigate whether that decision has an impact on the future financial outlook of the household. It could be that by tying funds into an iliquid TIRSA, households lose financial flexibility. Alternatively, it could be the case that the decision to open TIRSA is a consequence of the anticipation of higher income or an improved financial situation in the future. Therefore, we regress 1-year changes in income, financial satisfaction, financial slack,

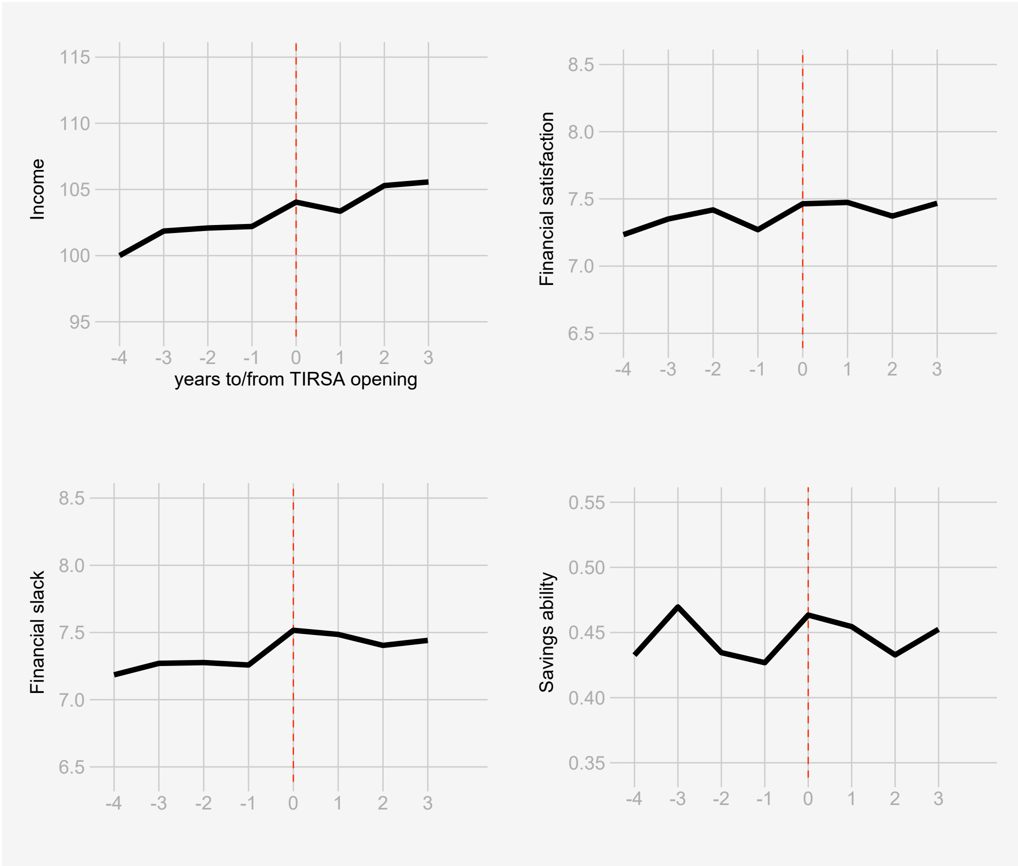


Figure 3.1 – Changes around opening of a TIRSA. Graphs show changes of individual variables around the opening of the first TIRSA for each relevant subject. Year zero is the year when a TIRSA was opened by each subject, and the other years are offset years before (-) and after (+) that event. *Income* and *Taxes* are net household yearly income and direct taxes, normalized to 100 for each subject 4 years (-4) before he/she opened a TIRSA.

and savings ability of the household on a set of socioeconomic variables plus a dummy variable if the household opened TIRSA in the previous year. The results presented in Table 3.6 clearly show that the opening of TIRSA does not affect income or self-assessed financial variables for the next year. In unreported results, we also test the impact of having opened a TIRSA in 2 to 5 previous years, and also find no effect.

Lastly, we complement the previous analysis with Figure 3.1 which shows changes in income, financial satisfaction, financial slack, and savings ability of the household 4 years prior to the year TIRSA is opened to 3 years after. We do not see any clear trends, again supporting the result that it is not changes in these variables that drive the decision to open an account. The regressions point towards a level effect but not a change effect.

3.5 Discussion and Conclusion

TIRSA have constrained withdrawal rules that limit the ability of households that own them to redeem their investments. One consequence of this fact is that, over time, households that own a TIRSA are not necessarily similar to the set households that, conditional on already owning a TIRSA, would hold their investment if they could also choose to withdraw. Analyzing the cross-section of determinants of TIRSA ownership might obfuscate what actually motivates households to actively make an investment in a TIRSA.

Thus, taking into consideration only the subset of households that do not have a TIRSA but could open one (within the institutional framework of the third-pillar tax-incentivized accounts in Switzerland), we show that some cross-sectional explanatory factors of TIRSA ownership are not significant in determining who opens a TIRSA at a given period. Furthermore, our results suggest that a one-year increase in household income does not affect the odds that a household opens a TIRSA, even though an increase in income plausibly releases households of certain financial restrictions that could have prevented them from opening a TIRSA in previous periods.

Then, from a policy perspective, we confirm that the households most likely to take advantage of the TIRSA scheme are wealthier, more educated, dual-income, and with a better financial outlook than households without a TIRSA.

References

Ameriks, J., Caplin, A., Laufer, S., & Van Nieuwerburgh, S. (2011). The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives. *Journal of Finance*, *66*(2), 519–561. https://doi.org/10.1111/j.1540-6261.2010.01641.x

Banks, J., Blundell, R., & Tanner, S. (1998). Is there a retirement savings puzzle? *American Economic Review*, *88*(4), 769–788. https://www.jstor.org/stable/117005

Battistin, E., Brugiavini, A., Rettore, E., & Weber, G. (2009). The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach. *American Economic Review*, *99*(5), 2209–2226. https://doi.org/10.1257/aer.99.5.2209

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. https://doi.org/10.1257/aer.91.4.832

Brounen, D., Koedijk, K. G., & Pownall, R. A. (2016). Household financial planning and savings behavior. *Journal of International Money and Finance*, *69*, 95–107. https://doi.org/10.1016/j.jimonfin.2016.06.011

Brown, D. C., Cederburg, S., & O'Doherty, M. S. (2017). Tax uncertainty and retirement savings diversification. *Journal of Financial Economics*, *126*(3), 689–712. https://doi.org/10.1016/j.jfineco.2017.10.001

Browning, M., & Crossley, T. F. (2001). The Life-Cycle Model of Consumption and Saving. *Journal of Economic Perspectives*, *15*(3), 3–22. https://doi.org/10.1257/jep.15.3.3

Browning, M., & Lusardi, A. (1996). Household Saving: Micro Theories and Micro Facts. *Journal of Economic Literature*, *34*(4), 1797–1855. https://doi.org/https://www.jstor.org/stable/2729595

Cocco, J. F., Gomes, F. J., & Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *Review of Financial Studies*, *18*(2), 491–533. https://doi.org/10.1093/rfs/hhi017

Devaney, S. A., Anong, S. T., & Whirl, S. E. (2007). Household savings motives. *Journal of Consumer Affairs*, *41*(1), 174–186. https://doi.org/10.1111/j.1745-6606.2006.00073.x

Dolls, M., Doerrenberg, P., Peichl, A., & Stichnoth, H. (2018). Do retirement savings increase in response to information about retirement and expected pensions? *Journal of Public Economics*, *158*, 168–179. https://doi.org/10.1016/j.jpubeco.2017.12.014

Duflo, E., Gale, W., Liebman, J., Orszag, P., & Saez, E. (2006). Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block. *Quarterly Journal of Economics*, *121*(4), 1311–1346. https://doi.org/10.1162/qjec.121.4.1311

Engen, E. M., Gale, W. G., & Scholz, J. K. (1996). The Illusory Effects of Saving Incentives on Saving. *Journal of Economic Perspectives*, *10*(4), 113–138. https://doi.org/10.1257/jep.10.4.113

FORS - Swiss Centre of Expertise in the Social Sciences. (2022). SHP Group, Living in Switzerland waves 1-22 + Beta version wave 23 + Covid 19 data [Dataset]. https://doi.org/doi.org/10.48573/vaxa-3w42

Hubbard, R. G., & Skinner, J. S. (1996). Assessing the Effectiveness of Saving Incentives. *Journal of Economic Perspectives*, *10*(4), 73–90. https://doi.org/10.1257/jep.10.4.73

Levenko, N. (2020). Perceived uncertainty as a key driver of household saving. *International Review of Economics and Finance*, *65*(March 2019), 126–145. https://doi.org/10.1016/j.iref.2019.10.005

Love, D. A., Palumbo, M. G., & Smith, P. A. (2009). The trajectory of wealth in retirement. *Journal of Public Economics*, *93*(1-2), 191–208. https://doi.org/10.1016/j.jpubeco.2008.09.003

Lugilde, A., Bande, R., & Riveiro, D. (2019). Precautionary Saving: a Review of the Empirical Literature. *Journal of Economic Surveys*, *33*(2), 481–515. https://doi.org/10.1111/joes.12284

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and retirement planning in the United States. *Journal of Pension Economics and Finance*, *10*(4), 509–525. https://doi.org/10.1017/S147474721100045X

MacKinnon, J. G., & Magee, L. (1990). Transforming the Dependent Variable in Regression Models. *International Economic Review*, *31*(2), 315. https://doi.org/10.2307/2526842

Pence, K. M. (2006). The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving. *Contributions in Economic Analysis & Policy*, *5*(1). https://doi.org/10.2202/1538-0645.1430

Skinner, J., & Hubbard, R. (1994). Expanding the life-cycle model: Precautionary saving and public policy. *American Economic Review*, *84*(2), 174–179. https://doi.org/10.2307/2117824

Stinglhamber, P., Zachary, M.-D., Wuyts, G., & Valenduc, C. (2007). The determinants of savings in the third pension pillar. *Economic Review*, (3), 97–113. https://ideas.repec.org/a/nbb/ecrart/y2007mdecemberiiiip97-113.html

Van Schie, R. J., Donkers, B., & Dellaert, B. G. (2012). Savings adequacy uncertainty: Driver or obstacle to increased pension contributions? *Journal of Economic Psychology*, *33*(4), 882–896. https://doi.org/10.1016/j.joep.2012.04.004

Ventura, L., & Eisenhauer, J. G. (2006). Prudence and precautionary saving. *Journal of Economics and Finance*, *30*(2), 155–168. https://doi.org/10.1007/BF02761482

Wärneryd, K. E. (1989). On the psychology of saving: An essay on economic behavior. *Journal of Economic Psychology*, *10*(4), 515–541. https://doi.org/10.1016/0167-4870(89)90041-X

Yaari, M. E. . (1965). Uncertain Lifetime , Life Insurance, and the Theory of the Consumer. *Review of Economic Studies, 32*(2), 137–150. http://www.jstor.org/stable/2296058