INDUSTRY SERIES

A comparison of risk preference elicitation methods and presentation formats

Jorgo Goossens, Marike Knoef, Bart Kuijpers, Rogier Potter van Loon, Eduard Ponds, Arno Riedl, Siert Vos

NETSPAR

DESIGN PAPERS are part of the **refereed Industry Paper Series**, which are refereed by the Netspar Editorial Board. Design Papers discuss the design of a component of a pension system or product. A Netspar Design Paper analyzes the objective of a component and the possibilities for improving its efficacy. These papers are easily accessible for industry specialists who are responsible for designing the component being discussed. Authors are allowed to give their personal opinion in a separate section. Design Papers are presented for discussion at Netspar events. Representatives of academic and private sector partners, are invited to these events. Design Papers are published at the Netspar website.

CONTENTS

Affiliations

Jorgo Goossens − Radboud University Nijmegen, Tilburg University, APG

Marike Knoef − Tilburg University

Bart Kuijpers − APG

Rogier Potter van Loon − Erasmus University Rotterdam, TKP

Eduard Ponds − Tilburg University, APG

Arno Riedl − CESifo, IZA, Maastricht University

Siert Vos − PGB Pensioendiensten, Universiteit van Amsterdam

Summary

We study the consistency of risk preferences between two presentation formats and three well-known risk preference elicitation methods. The presentation formats are a traditional questionnaire and a novel serious game. We conduct our survey with participants of a pension fund in the construction sector. Our results show that the distributions of risk preferences are similar in both presentation formats in the aggregate and for each elicitation method. However, we observe sizable differences in risk aversion levels and behavioral consistency between the three risk elicitation methods. For white-collar workers the consistency between elicitation methods is higher than for blue-collar workers. Additionally, older individuals and blue-collar workers show higher risk aversion than younger and white-collar workers. Overall, this study shows that the risk elicitation method and individual characteristics have a greater influence on the measured risk preferences than the presentation format.

Samenvatting

We bestuderen de consistentie van risicovoorkeuren met betrekking tot twee verschillende presentatievormen en drie bekende methodieken om risicovoorkeuren uit te vragen. De presentatievormen zijn een traditionele vragenlijst en een vernieuwende animatieweergave ("serious game"). De vragenlijst is uitgezet onder deelnemers van een pensioenfonds in de bouwsectoren. Onze uitkomsten laten zien dat de geaggregeerde verdeling van risicovoorkeuren vrijwel identiek is voor beide presentatievormen. Dit geldt ook voor elk van de drie afzonderlijke methodieken. We zien echter wel grote verschillen in risicovoorkeuren tussen de drie methodieken, zowel qua niveau als qua onderlinge consistentie. Voor werknemers die niet op de bouwplaats werken (zoals werknemers in de bouwsectoren met een kantoorbaan) vinden we een grotere consistentie tussen de drie methodieken dan voor werknemers op de bouwplaats. Ook hebben oudere werknemers en werknemers op de bouwplaats gemiddeld een hogere risicoaversie dan jongere werknemers en werknemers die niet op de bouwplaats werken. Samenvattend vinden we dat de methodiek om risicovoorkeuren uit te vragen en de individuele kenmerken van een deelnemer, zoals leeftijd en aard van het werk, een grotere invloed hebben op de gemeten risicovoorkeur dan de presentatievorm.

1. Introduction

Pension funds and insurance companies must elicit risk preferences at least every five years according to tentative new legislation. Therefore, it is important to know whether the presentation format of the risk preference survey impacts the elicited risk aversion levels. Based on expected utility theory (EUT), the presentation format of an elicitation method should not matter for eliciting risk preferences. If we find that different formats matter, then pension funds could take the layout of the survey into consideration. If we find that the formats do not matter for risk preferences, or to a lesser extent, then pension funds need to care less about the graphical layout. The goal of this paper is to better understand how the presentation format of risk elicitation methods impacts risk preferences. We pay specific attention to lower literacy groups. In this paper, we study whether framing an individual risky choice in a game results in different revealed risk preference estimates than when the decision is described in a neutral frame without graphical background effects.

We design two presentation formats to study how the graphical layout influences elicited risk preferences. The first consists of a traditional questionnaire with three well-known risk elicitation methodologies: (i) a choice sequence (CS) list (Barsky et al., 1997), (ii) a single choice (EG) list (Eckel and Grossman, 2002 and 2008), and (iii) convex time budgets (CTB, Andreoni and Sprenger, 2012). We also develop the same questionnaire in the form of a novel game. The game is identical to the traditional questionnaire in terms of the three risk elicitation methods (i.e., pension benefits, payout probabilities, and wording), except for differences in background graphics. In both formats we ask how certain participants are about their choices (cognitive certainty) after each elicitation method. We conduct our survey with actual pension fund participants in the construction sector. To proxy for low and high literacy participants, the sample includes both blue-collar and white-collar workers.

Our main finding is that the distributions of elicited risk preferences are similar in the two presentation formats. Median risk preferences are almost identical at the aggregate level, but also for each elicitation method. At the median, we find constant relative risk aversion (CRRA) parameter values of 5.6 for CS, 3.0 for EG, and 0.3 for CTB. However, the risk elicitation methods show sizable differences in median CRRA parameter values. Additionally, behavioral consistency – in terms of correlations – between risk-elicitation methods at the aggregate level is about 0.5 for CS and EG, while it is 0.07 for CS and CTB and 0.06 for EG and CTB. These results suggest that participants' preferences towards risk may vary when measured using different elicitation methods – a finding recently introduced as the risk elicitation puzzle (Holzmeister

and Stefan, 2021; Pedroni et al., 2017). White-collar workers typically display higher consistency than blue-collar workers. There are no differences between the questionnaire and the game. Finally, we find that age and being a blue-collar worker correlate positively with risk aversion.

Regarding the quality of the answers in the questionnaire and the game, we make two observations. First, the effect of the presentation format on completion rates is ambiguous: completion rates in the game can be up to 5% lower than in the questionnaire. This is mainly driven by white-collar workers completing the game less often than the questionnaire, not by more blue-collar workers completing the game. In general, the type of worker has a statistically and economically significant influence on the completion rate, as blue-collar workers have a 6% to 10% lower rate than white-collar workers in both presentation formats. Second, we find no effect of the presentation format on self-reported cognitive (un)certainty. Participants in the questionnaire are as cognitively certain as participants in the game.
However, blue-collar workers are typically less cognitively certain than white-collar workers. Also, blue-collar workers are more likely to respond with answers in a similar fixed order (i.e., answers in the form 'AAAA...', 'BBBB...', etc.).

Our paper contributes in three ways to the literature. First, we contribute on how the presentation format of risk preference elicitation tasks relates to final risk preference estimates. Alekseev et al. (2017) provide an overview of how textual context – rather than graphical context – affects behavior in several experimental tasks, such as public good and trust games, but they do not address risk-taking tasks. A paper close to ours is that of Friedman et al. (2022). They study whether variation in spatial representation, or variation in prices or probabilities, matters for risk preference elicitation procedures. They find that observed variation across elicitation methods, within the same individual, can be explained by design attributes.

Second, we contribute to the literature on the observed variation in risk preferences between different elicitation methods. Several papers have studied the consistency (or lack thereof) between risk preference elicitation methods and, thereby, the risk elicitation puzzle (Holzmeister and Stefan, 2021; Frey et al., 2017; Pedroni et al., 2017; Meeren et al., 2016; Crosetto and Filippin, 2016; Dave et al., 2010). However, to the best of our knowledge, the convex time budgets have not yet been compared with well-known existing risk elicitation methods in the pension domain, which typically concerns high stakes and long time periods.

Third, we relate stated cognitive (un)certainty to risk preferences. Amador-Hidalgo et al. (2021) study the relation between risk-taking behavior and revealed cognitive abilities, and Enke and Graeber (2022) study the relation between revealed cognitive

uncertainty and biases and beliefs. Dohmen et al. (2018) find that cognitive ability correlates positively with avoidance of harmful risky situations, but negatively with risk aversion in advantageous situations. We contribute to this literature by studying differences between low and high literacy individuals.

Please note that the risk attitude of an individual is typically characterized by two factors: (i) risk preferences and (ii) risk capacity. In this paper, we do not attempt to measure risk capacity; instead, we solely focus on the measurement of risk preferences.[1] This is consistent with the three elicitation methodologies that we use, as they are developed to measure risk preferences.

This paper is structured as follows. Section 2 describes the formats of presentation plus the methods to elicit risk preference. Section 3 discusses the quality of the answers in terms of response and completion rates, and of cognitive certainty. Section 4 shows the risk preference estimates for both presentation formats and all three elicitation methods. Section 5 states conclusions.

---

1   We refer readers interested in risk capacity to the Netspar Theme Grant of Van Ooijen and Brou-
     wer: https://www.netspar.nl/en/project/health-and-labor-market-uncertainty-over-the-life-
     cycle-impact/

2. Method and data

Our survey contains two presentation formats and three elicitation methods. We use a traditional questionnaire and a "serious game" as formats of presentation. For each presentation format, we use three frequently used elicitation methods for measuring risk preferences, in the following fixed order: (i) a choice sequence list (Barsky et al., 1997), (ii) a single choice list (Eckel and Grossman, 2002 and 2008), and (iii) convex time budgets (Andreoni and Sprenger, 2012). Each elicitation method is followed by a question about how certain the individuals are regarding their answers. We conducted our survey in cooperation with one of the major pension funds in The Netherlands during July 2021 and September 2021. The pension fund handles the pensions for the construction industry. Among those participants who finished the game or the questionnaire, 25 gift vouchers of €20 each were raffled.

Our survey was not incentivized strictly on the basis of the answers given by the participants.[2] It was stated at the beginning of the survey that the results of the study would be used to make the participant's pension better and more personal. In that sense, participation in the survey could be perceived as consequential. Pension funds in The Netherlands will be required by law to elicit risk preferences from their populations at least every five years and use this for their asset allocation (under supervision of the Dutch Authority of the Financial Markets). In line with common practice in the literature (see e.g., Alserda et al., 2019; Meeren et al., 2019), we tailor the monetary amounts shown in the survey towards the income level of the pension fund's population, in our case the average income of the fund.[3]

2.1 Format of presentation
A principal goal of our paper is to assess whether, and how, features of the elicitation method affect outcomes. Specifically, we study how the format for presenting the task affects risk preference estimates and the quality of the answers.

We use two presentation formats: a serious game and a traditional questionnaire. Our hypothesis is that the game format is easier to comprehend and needs less

2   In general, in economic experiments financial incentives constitute the default for theoretical and empirical reasons (Smith, 1976 and 1982). However, there is a trade-off with external validity. In our study, it is important that the monetary amounts are meaningful for pension savings decisions. This makes the implementation of incentive-compatible choices financially infeasible. There is also recent evidence suggesting that hypothetical choice options lead to comparable results (e.g., Cohen et al. (2020) and Hackethal et al. (2023)).

3   Table 1 shows that the mean income of our sample is consistent with the mean income at fund level.

Figure 1: Decision screen for the serious game

The figure (in Dutch) shows the first question of the choice sequence method in the serious game presentation format. See Appendix B, 1st question, for the English translation and replace 'pension' by 'island'. The individual is asked to choose between the safer island (i.e., pension) A and the riskier island (i.e., pension) B, in which good (in green) and bad (in red) states of the world happen with 0.5 probability – like heads and tails.

cognitive capacity. In both formats we keep the architecture of the three elicitation methods the same. That is, pension benefits, probabilities and wording are fixed in both formats. However, we vary the graphical format in which the elicitation methods are shown. In the traditional questionnaire, individuals make choices between different pension options and are asked to select the pension that fits their preferences. In the serious game, individuals are placed in a context where they are the captain of their pension, represented graphically by a boat, and they need to steer the boat to an island (i.e., possible pension) that they prefer. All financial characteristics are the same as in the traditional questionnaire. However, in the game we show animations of the boat, including seagull and water sounds, and we use the word "island" rather than "pension". Both presentation formats could be completed on a computer, tablet, or smartphone.

Figure 1 shows the decision screen in the serious game for the first question in the choice sequence method. When comparing the figure with the traditional questionnaire, which can be found in Appendix B, we see that the graphical formats differ, but everything else is identical, except for the words "island" and "pension".

2.2 Risk elicitation methods

In this section we describe the risk preference elicitation methods.

Choice sequence list – In the choice sequence list, subjects are asked to choose between two pensions: a more risky and a less risky pension. A pension is presented as a lottery. Variation across lotteries is obtained through manipulations of the outcomes of each lottery, while keeping the probability of the two outcomes fixed at 50% (i.e., similar as a coin toss for heads and tails). Per lottery, one outcome consists of a high payout that is defined as the situation 'better than expected', while the other outcome constitutes a low payout that is defined as 'worse than expected'. Subjects are asked to choose one pension per question for a total of five sequential questions. The method is based on the original approach of Barsky et al. (1997). The pensions that an individual can choose from depend on the individual's previous choices, so that over the choice sequence risk aversion is narrowed down to a small interval. Table 10 in Appendix A shows the values used in the experiment.

Single choice list – In the single choice list, subjects are asked to choose a pension from an ordered set of pensions. A pension is defined as a lottery. We use a version based on the question proposed by Eckel and Grossman (2002 and 2008). Table 11 in Appendix A presents the values used in the experiment.[4] Subjects choose the preferred pension from a set of four lotteries characterized by a linearly increasing expected value as well as greater standard deviation (except for Pension 4). More risk-averse subjects choose low-risk low-return pensions (Pension 1 and 2); risk-neutral subjects choose Pension 3; risk-seeking subjects choose Pension 4.[5]

The variation in answer possibilities is obtained through manipulations of the outcomes of each lottery, while keeping the probability of the two outcomes fixed at 50% (i.e., similar as a coin toss for heads and tails). For each lottery, one outcome contains a high payout and is defined as the situation 'better than expected', while the other outcome contains a low payout and is defined as 'worse than expected'. Subjects are asked to choose one pension.

Convex time budgets – An important advantage of the CTB method is that it allows measurement of risk and time preferences simultaneously. We simultaneously

---

4   The range and cutoff points for the CRRA parameter values are based on insights from an earlier risk preference study at a Dutch pension insurer.

5   In our analysis, we take the average value of the interval as a proxy for the risk aversion value. On the bounds, we assume values of $\gamma = 5.5$ and $\gamma = -2$.

measure risk aversion and patience. We use a shortened version of the original approach of Andreoni and Sprenger (2012), as we do not attempt to measure present bias.

In this method we ask individuals to allocate an initial budget m = € 10,000 between payments, available at two points in time: an early payment at time t and a delayed payment at time $t + k$. The early payment is always one year from the survey date (to avoid interference with present bias). The late payment is delayed by either five years (k = 5) or ten years (k = 10). Subjects receive interest, or return, r on delayed payments, which varies between 0% to 21.06% interest on an annual basis. The allocations must be made such that their budget constraint is satisfied, i.e., the early payment and the present value of the delayed payment must equal the initial budget m. Early and late payments are certain.

Individuals make six consecutive CTB decisions between early and delayed payments. We implement two different decision sets, and within each set there are three different interest rate scenarios. The first choice set uses k = 10, and the three decisions within this set differ in the accrued return. The second choice set uses k = 5, and the three decisions within this set differ in the accrued return accordingly as well. Differences between the delayed payment dates $t + k$ (i.e., back–end delay) elicit long–term patience. Sensitivity to variation in the interest rates, or return, identifies curvature of the utility function. Table 12 in Appendix A presents an overview of our experimental CTB design.

To estimate risk preferences and time preferences, we identify the hypothetically allocated payments as solutions to standard intertemporal optimization problems. These solutions are assumed to be functions of our parameters of interest (discounting and risk aversion) and exogenously varied parameters (interest rates and delay periods). Given assumptions on the functional form of utility and the nature of discounting, this setup provides a natural context to jointly estimate individual risk and time preferences.

We assume that an individual's preferences can be represented by a standard CRRA utility function with curvature parameter γ and that they discount future payoffs exponentially with discount factor δ. We estimate risk and time preferences jointly, i.e., the CRRA risk aversion parameter γ and the long–term discount factor δ. Our preference estimates are based on OLS regressions. In line with the existing literature (Andreoni and Sprenger, 2012; Potters et al., 2016) and consistent with our two previous elicitation methods, we assume a background income close to zero. That is, we assume that participants do not integrate any other income sources with their CTB decisions, consistent with the choice sequence and single choice lists. Stated

Figure 2: Mapping of choices into the implied CRRA risk aversion parameter value by task
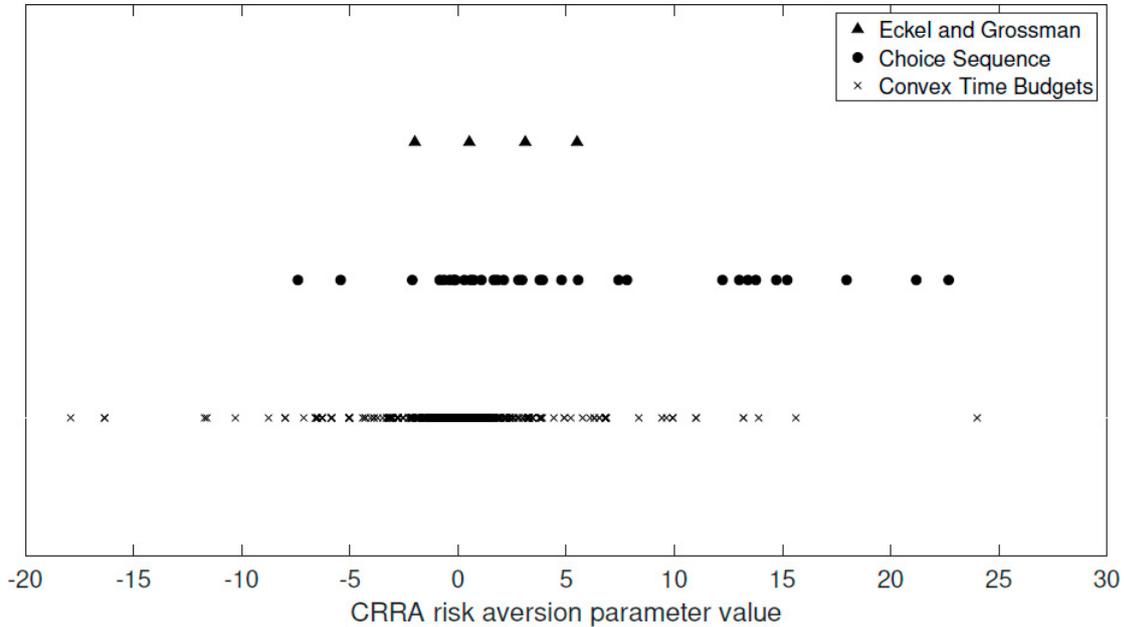
The figure assumes a CRRA power function $U(x) = \dfrac{x^{1-\gamma}}{1-\gamma}$. γ = 0 means risk neutral, γ > 0 implies risk-averse preferences, and γ < 0 implies risk-seeking preferences.

differently, we do not control for the risk capacity of an individual in the estimation of the risk preferences. See Goossens and Knoef (2022a) for more details on the estimation.

Note that the CTB method identifies risk preferences through the elasticity of inter-temporal substitution (i.e., smoothness of consumption path), whereas the CS and EG methods identify risk preferences through the risk aversion (i.e., uncertainty over different states of the world). All methods identify risk preferences as curvature of the utility function. Since we work with a standard CRRA utility function, the risk aversion parameter in the CTB method equals the inverse of the elasticity of intertemporal substitution.[6]

Theoretical possibilities – Our analysis uses EUT with CRRA as the model for normative choice. Figure 2 shows all theoretically possible outcomes for each task in the space of the CRRA coefficient γ. We assume throughout this paper that the utility function employs the CRRA form

6   To separate risk aversion from the elasticity of intertemporal substitution, one could assume recursive utility, for example using the preference specification of Epstein-Zin (1989).

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma}$$

where γ = 0 means risk neutral preferences, γ > 0 implies risk-averse preferences, and γ < 0 implies risk-seeking preferences. It is immediately clear from the figure that EG is a coarse measure with only four outcome possibilities. CS has 32 possible distinct values, and CTB has 4,096 possible distinct values. Therefore, CTB, and to a lesser extent CS, can be viewed as more continuous measures relative to EG.

Notably, CTB is well suited for distinguishing between slightly risk-seeking and slightly risk-averse behavior, because many possible parameters cluster around zero.

2.3 Sample

We sent invitations to 25,000 pension fund participants. Of the invitees, 60% is still employed in the construction sector and building up a pension at the pension fund. 40% is not actively building up a pension at the pension fund anymore; they are either working elsewhere or are retired. We invited the same proportion of active and inactive participants to the game and the questionnaire, respectively.

Table 1 shows that our experimental sample of N = 1,601 pension fund participants has similar characteristics as the total pension fund population. We consider respondents who completed all 15 questions of the questionnaire and the game, respectively. Most of the respondents, 95%, are male, which reflects the population in the pension fund. The average age and income in our experimental sample are slightly higher than at the pension fund level, but lie within one standard deviation. Panel A shows that the proportion of males, average age, and average income are all similar for game respondents and questionnaire respondents. From Panel B we observe that blue-collar workers are almost all male and earn less than white-collar workers.

About 50% of the respondents are active, meaning that they are working in the construction sector. The other half, the group of inactive respondents, consists for two-thirds of pensioners and for one-third of people who do not yet receive a pension but who also do not build up a pension at the pension fund for construction workers anymore. Individuals in this latter group, the so-called 'sleepers', may build up pension elsewhere, are self-employed, or have quit working for some unknown reason.

The pension administrator provided us with additional information for most of the respondents. For almost all respondents we know their age, gender and status (active, sleeper or pensioner). For roughly 80% of our sample, including all active respondents, we know whether individuals have or had a blue-collar job (simply put,

Table 1: Comparison of experimental sample and actual pension fund

Panel A	Pension fund	Sample Total	Questionnaire	Game
Male	0.92	0.95 (0.22)	0.94 (0.23)	0.95 (0.22)
Age (years)	55	59 (12)	59 (12)	60 (13)
Income (euros)	45,300	50,062 (24,502)	50,100 (25,750)	50,031 (23,477)
# Respondents		1,601	690	911

Panel B	Pension fund	Type of work available	Blue collar	White collar
Male	0.92	0.94 (0.23)	0.99 (0.08)	0.89 (0.32)
Age (years)	55	57 (12)	58 (13)	57 (12)
Income (euros)	45,300	50,062 (24,502)	34,369 (8,249)	67,238 (24,872)
# Respondents		1,308	705	603

Mean values with standard deviations between brackets. Annual before-tax income is only based on employed participants who actively build up a pension. Panel A shows statistics for subgroups related to survey type. Panel B shows statistics for subgroups related to type of work.

are working on a construction site) or a white-collar job (for example, in IT or administration). This allows us to analyze whether either of these characteristics affects the completion rate, risk attitude, or self-reported cognitive certainty when answering the survey.

3. Quality of the answers

3.1 Response and completion rates

We asked the invitees by email to complete a task using an online tool. Two-thirds of the invited participants received a link to the game, the remaining one-third received a link to the traditional questionnaire. From the invitation email, participants could not infer whether they would end up playing the game or answering the traditional questionnaire. This they learned only after clicking on the link.

Therefore, we do not expect a difference in response rate concerning opening the game link or the questionnaire link. However, there is a difference in completion rates, where a completion rate measures the fraction of invited participants that finished the survey. The median time to complete the survey is five minutes.

Table 2: Completion rate by mode of presentation and worker status

status	Game		Questionnaire		Total	
	invitations	completion (%)	invitations	completion (%)	invitations	completion (%)
active	10,000	4.4	4,998	7.2	14,998	5.3
inactive	6,667	7.0	3,333	10.0	10,000	8.0
total	16,667	5.5	8,331	8.3	24,998	6.4

The completion rate measures the percentage of invited participants who finished the survey, as given in the column "completion (%)".

Table 2 shows completion rates for all invitees, as well as for participants of the game and the questionnaire, and for active and inactive participants. The average completion rate is 6.4%. That is, 1,601 respondents completed the survey and answered all 15 questions covering the three elicitation methods (choice sequence, single choice list, and convex time budgets). Inactive respondents have a 50% higher completion rate than active respondents. This holds for respondents to both game and questionnaire. Also, invitees who received the questionnaire have a 50% higher completion rate than invitees who received the game. This holds for both active and inactive respondents.

Table 3 shows the number of respondents who answered the first question and the number of respondents who answered all questions.[7] Of the respondents who answered the first question, 87% (game) to 89% (questionnaire) completed the survey. Thus, conditional on answering the first question, there is only a minor difference in terms of completion between game and traditional questionnaire. The

---

7   Unfortunately, we do not have data on whether participants clicked on the invitation link. Therefore, we have used completion until question 1 as a measure for response.

Table 3: Completion rate by mode of presentation

	Game			Questionnaire		
completion	until Q1	until Q15	until Q15 | Q1	until Q1	until Q15	until Q15 | Q1
active	5.0%	4.4%	88.8%	7.9%	7.2%	91.1%
inactive	8.2%	7.0%	85.4%	11.4%	10.0%	87.6%
all respondents	6.3%	5.5%	87.0%	9.3%	8.3%	89.4%

Share of respondents who answered at least the first question (until Q1) and all 15 questions (until Q15) of the game and the traditional questionnaire, respectively. The third column (until Q15 | Q1) of each presentation mode shows the percentage of respondents who completed the survey, given that they answered the first question.

remaining 12% of the respondents who answered the first question quit along the way, with about 60 individuals quitting immediately after the first question, 50 more at the start of the second elicitation method, and 30 more at the start of the third elicitation method. Thus, across the three methods, completion rates do not appear to differ much.

When comparing respondents who answered the game or the traditional questionnaire, we see that 6% of the game respondents finished the first question, whereas this is the case for 9% of the respondents to the questionnaire. This difference could be due to more invitees to the game not opening the survey, although from the invitation they could not know whether they would receive a game or a traditional questionnaire. Another explanation could be that more invitees to the game started the survey but quit before answering the first question.

One might expect that younger participants would be more tempted to complete the game than the questionnaire. As survey participants have been selected based on working status rather than age, we distinguish completion rates of active and inactive respondents. We observe that fewer active than inactive individuals open the survey and answer the first question, but that, once started, active individuals complete the survey slightly more often. Both active and inactive respondents complete the traditional questionnaire more often than the game. Hence, active participants, who are younger on average, are not more motivated to complete the game.

When including only those respondents for whom additional characteristics are available, including type of work, similar completion percentages hold. In that case, however, we do observe an interesting difference between blue- and white-collar workers. A priori the game was developed to involve more blue-collar workers. Such higher involvement should be reflected by a higher response rate for this group, an improved measurement of risk attitude, more consistent choices, and less cognitive uncertainty. Unexpectedly, of the blue-collar workers almost 15% quit early in both game and questionnaire. Hence, involving more blue-collar workers by

Table 4: Completion rates, presentation mode, and individual characteristics

	Completion (model 1)	Completion (model 2)	Completion (model 3)	Completion (model 4)
Blue collar	−0.064*** (0.017)	−0.091*** (0.026)	−0.098*** (0.026)	
Game	−0.025 (0.017)	−0.052**a (0.026)	−0.054** (0.026)	−0.014 (0.018)
Blue collar and game		0.048b (0.034)	0,051 (0.034)	
Age <= 55			0.039** (0.018)	0.054** (0.026)
Male			0.069* (0.037)	
Age <= 55 and game				−0,031 (0.035)
Constant	0.929*** (0.016)	0.944*** (0.020)	0.871*** (0.038)	0.878*** (0.014)
Observations	1,488	1,488	1,488	1,819
R²	0.010	0.012	0.017	0.004
Adjusted R²	0.009	0.010	0.014	0.003

Note: The sum of a and b is not significantly different from zero
*p < 0.1; **p < 0.05; ***p < 0.01

Using OLS, we regress the dummy variable 'completion' (= 1 when the subject completed the survey, conditional on answering the first question, = 0 otherwise) on individual characteristics. We report standard errors between parentheses.

presenting them a game did not appear to increase completion rates. Moreover, of the white-collar workers roughly 5% quit early in the questionnaire and roughly 10% in the game.

Compared to younger individuals, we see that also a larger share (around 15%) of pensioners and older respondents stop early. This holds for both game and questionnaire. Still, given the higher overall completion rate for inactives, we may conclude that pensioners are on average more inclined to start the survey in the first place.

Table 4 shows that similar conclusions can be drawn from an Ordinary Least Squares (OLS) regression of the completion rate on some background variables. A survey is considered to be completed if a respondent has answered all 15 questions, conditional on answering the first question. A survey is not completed if a respondent quit the survey at some point between the first and the last question.

Blue-collar workers are less likely to complete the survey. Also, game participants are less likely to complete the survey (model 1), although this effect is mainly due to white-collar workers not completing the game (model 2). In other words, for blue-collar workers the combined effect of the survey type on the completion rate is not significantly different from zero, while for white-collar workers the game leads

to a lower completion rate. In model 3 we correct for age and gender: this hardly affects our observations regarding type of work and presentation format. We notice that respondents younger than 55 (roughly the younger half of the active participants) complete the survey more often than older respondents. And males are more likely to complete the survey than females, although the statistical significance for this is weak and hard to detect, given the small number of females in the sample (and in the population). Finally, model 4 shows that young respondents in the game are less likely to complete the questions, although this effect is statistically insignificant.

3.2 Cognitive certainty

There are many different methodologies to measure the risk attitude of individuals. Risk preferences are a latent variable; they might be instable over time or across methodologies (Pedroni et al., 2017; Schildberg–Hörisch, 2018). Respondents may find it difficult or confusing to answer a survey about risk attitude. For this reason we are interested in the quality of the answers given. To investigate this we look at different quality measures: self–reported cognitive certainty, the use of an automatic pilot (i.e., whether respondents have given the same answer, all A or all B, to every question), and the consistency of risk attitudes across methods. In this section, we focus on the first two measures.

First, we asked respondents how certain they are about their answers, on a 4–point Likert scale from 'very uncertain' to 'very certain', after each of the three methods. Overall, we consider a respondent certain if the person responded having been 'certain' or 'very certain' about each method. Table 5 shows the outcomes of our OLS regression analyses, using the same explanatory variables as in Table 4. We conclude that blue–collar workers are less certain about the answers given, and this effect is statistically significant. Respondents to the game tend to be less certain, but blue–collar workers in the game tend to be more certain than white–collar workers. However, these effects of the presentation format to explain certainty are not statistically significant.

Second, we defined an 'automatic pilot' dummy variable, indicating whether respondents gave the same answer (all A or all B) to each of the five choice sequence (CS) questions. This could relate to true underlying risk preferences, but it could also indicate that individuals answer without thinking well about the choice problem. Again, using the same set of explanatory variables as before in an OLS regression analysis, we observe that the type of work is statistically significant: blue-collar workers use the automatic pilot more often than white–collar workers. Respondents to the game use the automatic pilot less often (model 1). This also holds for blue–collar

Table 5: Self-reported certainty and auto-pilot subjects

	Certainty (model 1)	Certainty (model 2)	Certainty (model 3)	Certainty (model 4)	Auto pilot (model 1)	Auto pilot (model 2)	Auto pilot (model 3)	Auto pilot (model 4)
Blue collar	−0.051** (0.026)	−0.095** (0.039)	−0.100** (0.040)		0.172*** (0.022)	0.202*** (0.033)	0.208*** (0.033)	
Game	−0,014 (0.026)	−0,056 (0.039)	−0,057 (0.039)	−0,005 (0.028)	−0.063*** (0.022)	−0.034[a] (0.033)	−0,031 (0.033)	−0.086*** (0.025)
Blue collar and game		0.078 (0.052)	0.082 (0.052)			−0.054[b] (0.045)	−0,065 (0.044)	
Age <= 55			0.035 (0.028)	0.051 (0.039)			−0.103*** (0.023)	−0.149*** (0.034)
Male			0.046 (0.057)				−0.056 (0.049)	
Age <= 55 and game				−0,018 (0.053)				0,041 (0.046)
Constant	0.713*** (0.024)	0.737*** (0.029)	0.684*** (0.060)	0.654*** (0.022)	0.154*** (0.021)	0.137*** (0.025)	0.222*** (0.050)	0.308*** (0.019)
Observations	1,308	1,308	1,308	1,601	1,308	1,308	1,308	1,601
R^2	0.003	0.005	0.007	0.002	0.051	0.052	0.067	0.026
Adjusted R^2	0.002	0.003	0.003	0.000	0.050	0.050	0.063	0.024

Note: The sum of a and b is not significantly different from zero
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

In the columns 'Certainty' we regress the dummy variable 'certainty' (= 1 if self-reported 'certain' or 'very certain', = 0 otherwise) on individual characteristics. In the columns 'Auto pilot' we regress the dummy variable 'auto pilot' (= 1 if same answers in CS method, = 0 otherwise) on individual characteristics. We use OLS regressions and report standard errors between parentheses.

workers who participate in the game (model 2, where we tested that the combined effect of the presentation format on automatic pilot behavior is statistically significant for blue-collar workers). Finally, we observe that respondents younger than 55 use the automatic pilot less often than older respondents.

Additionally, we tested autopilot behavior for CTB. The CTB methodology, consisting of six questions each having four answer possibilities, is also potentially exposed to autopilot answers. However, for CTB we only find weak statistical significance that blue-collar workers use the automatic pilot more than white-collar workers. Statistical significance of worker type disappears when correcting for the background characteristics of age and gender. Besides, similar to CS, respondents aged 55 or younger use the autopilot less often. We do not observe any impact from the presentation format on autopilot CTB answers. All in all, the explanation for autopilot behavior is weaker for CTB than for CS. Possible reasons are that CTB questions vary across more dimensions (both in terms of risk and time horizon) or that CTB questions have four answer possibilities (instead of two with CS), thereby reflecting more closely the respondents' underlying risk preferences.

4. Risk preference estimates

Having studied the differences in quality of the answers between the serious game and the traditional questionnaire, we now compare the risk preference estimates between presentation formats and elicitation methods.

4.1 Summary statistics

Table 6 shows the summary statistics for the estimated risk preferences that result from the three methods, aggregated over the game and the traditional questionnaire. We observe that the estimated average risk aversion differs across methods. Choice sequence gives the highest average risk aversion parameter. Also the standard deviation in risk aversion is larger compared to the other methods.

Although the mean and median CRRA parameter values are in line with previous findings from several risk elicitation methods in the literature (Alserda et al., 2019; Meeren et al., 2019; Goossens et al., 2022b), we notice that the mean and median risk aversion levels differ substantially between methods:
8.74 and 5.55 for CS, 3.21 and 3.00 for EG, and 0.39 and 0.28 for CTB. Such differences would yield economically sizeable effects in, for example, optimal asset allocations of pension funds.

Tables 13 and 14 in Appendix A show that differences in elicited preferences between the game and the traditional questionnaire are small to negligible in terms of magnitudes, especially at the median. This is an interesting finding, as it indicates that the graphical format of presentation is irrelevant for risk preference elicitation.

Table 7 shows quantile regressions on the median for CS and CTB and an OLS regression for EG, with the elicited parameters as dependent variable and the same set of explanatory variables as in the previous sections, separately for each elicitation method. They show that blue-collar workers are more risk- averse than white-collar

Table 6: Overview of risk preference estimates

method	obs	mean	std	5%	25%	median	75%	95%
CS	1,308	8.74	8.41	−0.12	2.11	5.55	13.40	22.68
EG	1,308	3.21	1.87	0.50	3.00	3.00	5.50	5.50
CTB	1,308	0.39	1.87	−2.10	0.06	0.28	1.30	1.65
Composite	1,308	4.11	3.20	0.00	1.76	3.20	6.50	9.66
CTB delta	1,308	1.03	0.27	0.82	0.89	0.96	1.07	1.47

This table shows the summary statistics for the CRRA parameter values per methodology. 'Composite' is the unweighted average of the CRRA parameter values from CS, EG, and CTB. 'CTB delta' is the simultaneously estimated time preference parameter from the convex time budgets.

workers. The difference is statistically significant for two of the three methods, and the economic effect is sizeable, in particular for the choice sequence method. There we find that the *median* risk aversion parameter is more than 4 points higher for blue-collar workers compared to white- collar workers. Also, the *average* risk aversion is more than 4 points higher, and the variation in risk aversion levels is larger within the group of blue-collar workers. Reason for this is that blue-collar workers choose more often for the least risky pensions and thus display higher risk aversion. To be precise, almost 30% of the blue-collar workers have a CS risk aversion parameter larger than 20, compared to 10% of the white-collar workers. Just over 30% of the blue-collar workers have a CS risk aversion parameter between 0 and 5, versus almost half of the white-collar workers. Also for EG and (to a lesser extent) CTB we find that blue-collar workers choose more often for the less risky pensions.

Another factor affecting risk aversion in CS (and to a smaller extent also in the other two methods) is age. Older respondents have a higher risk aversion than younger individuals, which is in line with other findings (Schildberg-Hörisch, 2018). The effect of the presentation format on risk aversion is ambiguous. Based on the CTB method we observe that respondents to the game have a higher *median* risk aversion than respondents to the traditional questionnaire, although this effect is offset for

Table 7: Risk aversion parameters, presentation mode, and individual characteristics

	CS gamma	EG gamma	CTB gamma
Blue collar	4.627*** (1.081)	0.634*** (0.156)	0.123 (0.084)
Game	−0.499 (0.898)	−0.077 (0.171)	0.414*** (0.113)
Blue collar and game	−1.870 (1.548)	0.001 (0.207)	0.020 (0.129)
Age <= 55	−2.757*** (0.829)	−0.121 (0.161)	−0.129 (0.079)
Male	0.000 (0.655)	−0.071 (0.227)	0.046 (0.127)
Age <= 55 and game	0.499 (1.045)	−0.081 (0.219)	−0.400*** (0.125)
Constant	5.549*** (0.928)	3.036*** (0.238)	0.210* (0.125)
Observations	1,308	1,308	1,308
R^2		0.032	
Adjusted R^2		0.028	

Note: *p<0.1; **p<0.05; ***p<0.01

We regress the CRRA risk aversion parameter on individual characteristics. In the columns 'CS gamma' and 'CTB gamma' we use quantile regressions on the median. In the column 'EG gamma' we use an OLS regression. We report standard errors between parentheses.

younger respondents participating in the game. Moreover, we do not observe a statistically significant effect of the presentation format with the other methods, nor when we analyze *average* risk aversion levels with CTB.

4.2 Formats of presentation

This section specifically analyses risk preferences in the game and the traditional questionnaire per elicitation method. Our method of analysis is based on univariate statistics. We first discuss overall differences in risk preferences between game and traditional questionnaire, then we zoom in on individual characteristics such as worker type (i.e., blue- versus white-collar) and age.

Choice sequence – Figure 3 shows the distribution of CRRA parameter values from the CS method in both formats of presentation. In general, we observe only small differences between respondents to the game and to the traditional questionnaire. According to the choice sequence method, respondents to the questionnaire are on average slightly more risk-averse than respondents to the game, with a gamma of 9.3 versus 8.3; however, the median risk aversion levels (indicated by the white lines) are almost equal. The distribution across all possible risk attitudes is very similar. We observe the largest differences between the formats of presentation at the higher risk aversion levels. The variation in risk preferences is higher in the questionnaire than in the game. Relatively more individuals in the questionnaire (23% versus 18% of respondents to the game) have an estimated risk aversion larger than 20.

Figure 3: Distribution of CS risk aversion levels in serious game and traditional questionnaire
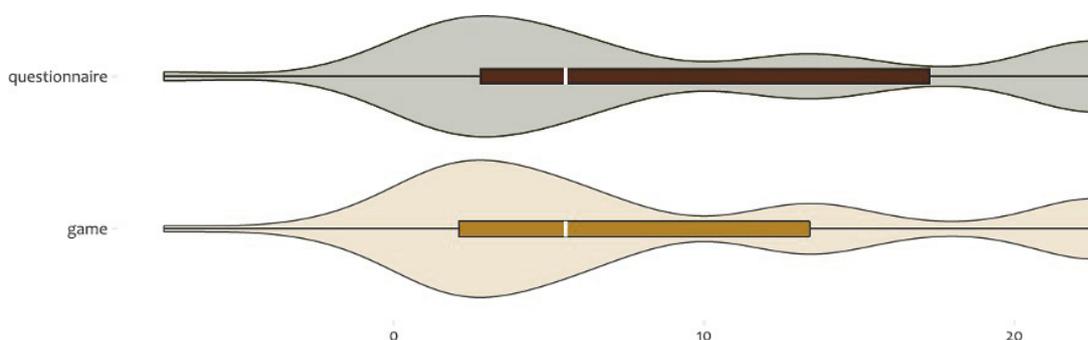
The upper chart presents the distribution of CRRA parameter values in the traditional questionnaire, while the lower chart is for the game. The horizontal bars in the figure run from the 25th to the 75th percentile, which means that at least 50 percent of the respondents has a risk aversion falling inside the bar. The white line equals the median.

Figure 4: Distribution of EG risk aversion levels in serious game and traditional questionnaire
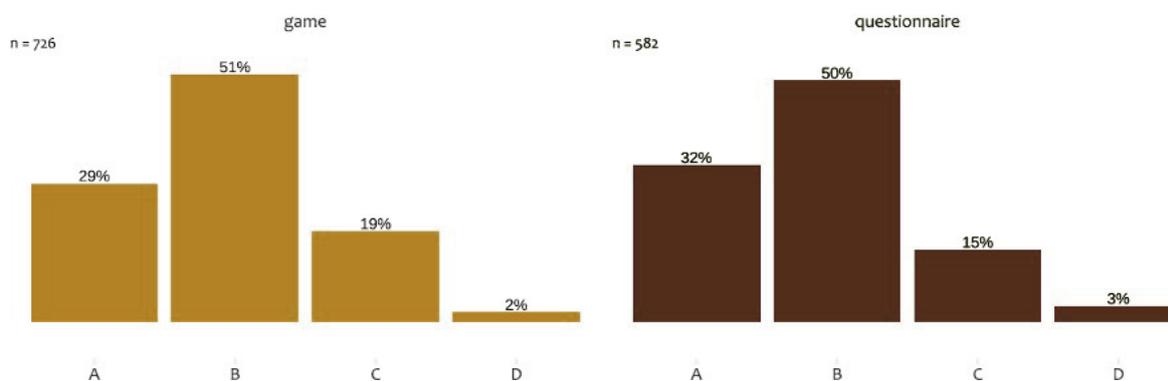
In both bar charts, the horizontal axes display the four EG options (A most risk averse, D risk neutral / risk seeking).

Eckel and Grossman – Also according to EG risk attitudes, respondents to the questionnaire seem to be slightly more risk-averse at the mean than respondents to the game (see Figure 4 for the distributions). A somewhat larger share of respondents to the questionnaire chooses the least risky pension A, mostly at the expense of the more risky pension C. However, the differences in risk preferences between game and questionnaire are small, and even smaller than with the CS method. Overall, the distributions of risk preferences in the game and questionnaire are similar.

Convex time budgets – According to the CTB method there is hardly any difference between respondents to the game and respondents to the questionnaire (see Figure 5 for the distributions of risk aversion levels). For both formats of presentation, average risk aversion levels are 0.4. The majority of individuals is risk-averse, i.e., they have a CRRA risk aversion value larger than zero.[8]

Individual characteristics – In Section 4.1, we already saw that blue-collar workers are more risk-averse than white-collar workers in the aggregate by about 4 units in CRRA parameter values for the choice sequence method. This observation also holds separately for respondents in the game and respondents in the questionnaire (see Figure 6 for the average CS risk aversion levels per worker type). This implies that the difference

---

8   Using the CTB also yields time preference estimates. In the game we estimate an average annual discount factor of 1.0, and in the traditional questionnaire we estimate an average annual discount factor of 1.1. This indicates that participants use subjective discount rates close to or just below zero, in line with prevailing market interest rates at the time of our survey.

Figure 5: Distribution of CTB risk aversion levels in serious game and traditional questionnaire
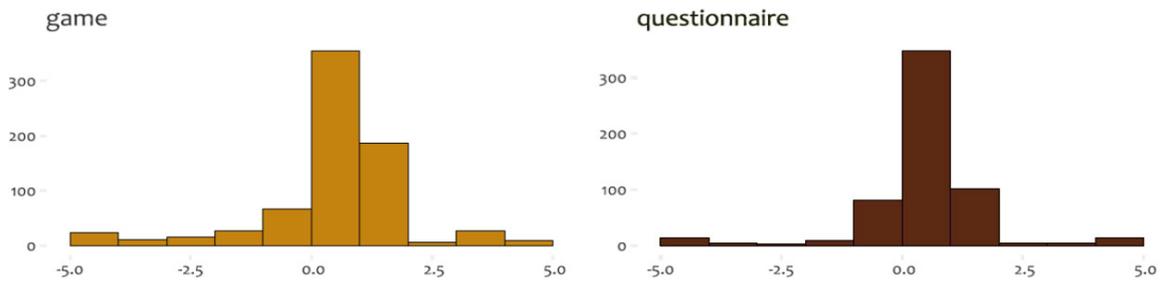
In both bar charts, the horizontal axes display the CRRA parameter values, while the vertical axes display the frequency. All risk–aversion levels smaller than −5, or larger than +5, are included in the respective boundary buckets.

Figure 6: Average CS risk aversion levels in the serious game and the traditional questionnaire by worker type
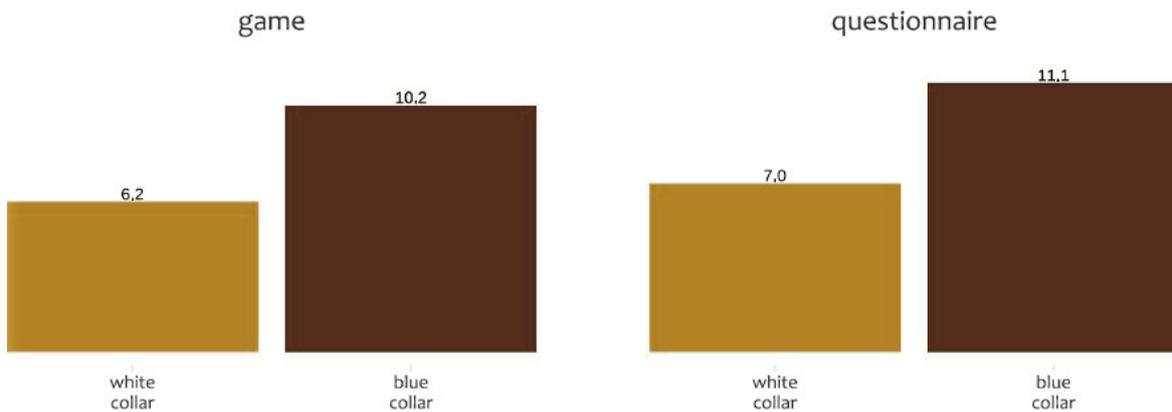
The height of the bars represents the average CRRA parameter values.

Figure 7: Distributions of CS risk aversion levels per presentation format and worker type
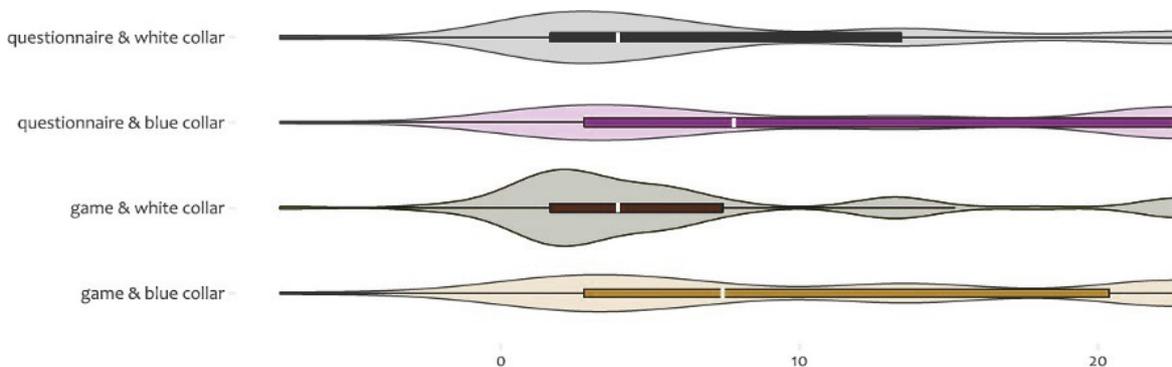
The horizontal bars in the figure run from the 25th to the 75th percentile, which means that at least 50 percent of the respondents has a CRRA risk aversion falling inside the bar. The white line equals the median.

Figure 8: Average CS risk aversion levels in the serious game and the traditional questionnaire by age bucket
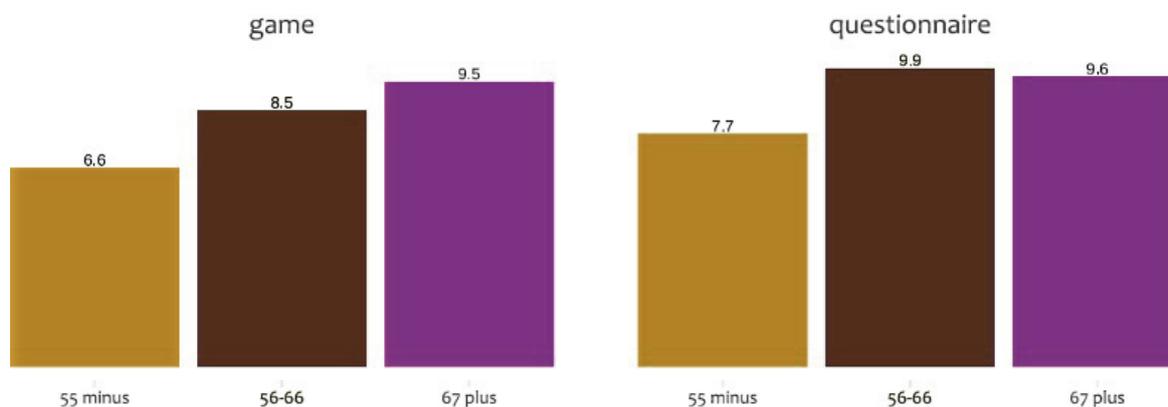
The height of the bars represents the average CRRA parameter values.

in risk aversion is much larger between types of workers (blue- or white-collar) than the difference due to the type of presentation format.

Figure 7 shows the distribution of risk aversion levels (including the median risk aversion) per worker type and presentation format. Note that Figures 6 and 7 indicate that respondents to the questionnaire have on average a 1 point higher risk aversion parameter value, while median risk aversion levels are nearly identical. This holds for both blue- and white-collar workers.

For the EG method we observe similar results (see Figure 11 in Appendix A). Blue-collar workers show on average higher risk aversion than white-collar workers, irrespective of the presentation method.

Respondents to both game and questionnaire have a 0.7 higher EG risk aversion parameter if they have a blue-collar job than when having a white-collar occupation. Interestingly, for CTB risk aversion (see Figure 12 in Appendix A) we only find a clear difference between blue- and white-collar for respondents to the questionnaire. Respondents to the game have a similar average risk aversion, irrespective of the type of work.

Finally, we focus on age in Figure 8. We already noticed that older respondents have a higher risk aversion than younger individuals. This holds for respondents to both game and questionnaire.

However, for respondents to the questionnaire, risk aversion does not increase anymore with age for the two oldest age buckets (i.e., ages 56–66 and 67 or older). For participants aged 67 or older, the average risk aversion is the same for game and questionnaire respondents. For the younger age buckets we again observe that respondents to the questionnaire are on average slightly more risk-averse (roughly

one point in terms of the CRRA risk aversion parameter) than respondents to the game.

For the other two methods we observe less pronounced differences in average risk aversion between age buckets. See Figures 13 and 14 in Appendix A. In general we conclude that older respondents tend to be more risk-averse than younger respondents, and that age has a larger impact on CRRA risk aversion values than the type of presentation format.

4.3 Consistency between methods

There are many measures to determine the consistency between risk preference elicitation methods. We use two measures in our analysis: (i) correlations and (ii) rankings.

First, we use the simple metric of Spearman rank correlation coefficients between the three risk aversion methods. Table 8 shows our results. We observe that the risk aversions following from CS and EG are positively correlated. Despite small differences between game and questionnaire and between blue- and white-collar workers, the correlation is sizable and significantly different from zero for all these subgroups. The correlations are lower when we compare the CTB risk aversion with the two other methods. Nevertheless, for the full sample (as well as for some subsamples) the correlation is still significantly different from zero. Interestingly, we observe that the correlations for blue-collar workers are generally lower than for white-collar workers.

In Table 9 we use an OLS regression analysis to explore whether the difference in consistency between presentation formats is significant.[9] We regress risk aversion parameters of one method on risk aversion parameters of another method. The regression coefficients between CS and EG risk aversion parameters are statistically significant for all specifications. That is, CS and EG risk aversion parameters are

Table 8: Consistency of risk preferences

correlations	all	game	questionnaire	white collar	blue collar
CS vs EG	0.49***	0.53***	0.44***	0.52***	0.44***
CS vs CTB	0.07**	0.05	0.10**	0.09**	0.03
EG vs CTB	0.06**	0.07**	0.04	0.06	0.04

Note: * p<0.1; ** p<0.05; *** p<0.01

This table shows the Spearman rank correlation coefficients between the CRRA risk preference elicitation methods. The columns indicate subgroups.

9   Our results are almost identical when using ordered probit and ordered logit regressions.

Table 9: Consistency of risk preferences

	CS gamma	CS gamma	CS gamma white collar	CS gamma blue collar	CS gamma	CS gamma
EG gamma	2.043*** (0.111)	1.865*** (0.163)	1.641*** (0.237)	1.857*** (0.214)		0.026 (0.028)
Game		−1.793** (0.829)	−2.958** (1.304)	−0.417 (0.982)		
EG gamma * game		0.318 (0.222)	0.628* (0.328)	−0.052 (0.288)		
CTB gamma					0.033 (0.124)	
Constant	2.175*** (0.412)	3.182*** (0.618)	5.299*** (0.954)	1.519** (0.743)	8.727*** (0.237)	0.302*** (0.103)
Observations	1,308	1,308	705	603	1,308	1,308
R^2	0.206	0.210	0.177	0.216	0.000	0.001
Adjusted R^2	0.206	0.208	0.173	0.212	−0.001	0.000

Note: *p<0.1; **p<0.05; ***p<0.01

The first four columns of this table show OLS regressions of CRRA parameter values elicited from CS on CRRA parameter values elicited from EG risk aversion parameters. The impact of presentation format and type of worker on CS versus EG consistency is shown in columns 2, 3, and 4 respectively. The last two columns show consistency of CS and CTB, and of CTB and EG, respectively.

significantly positively correlated. This also holds for white- and blue-collar workers separately. As measured by the interaction variable in Table 9, there is a slightly higher correlation between CS and EG risk aversion among game participants, although not (or at most, weakly) statistically significant.

The consistency between CTB risk aversion and the risk aversion following from the other two methods is not significantly different from zero (despite Table 8 showing that ranks are slightly positively correlated).

A second measure for consistency between methods is directly comparing the outcomes for (groups of) respondents in terms of ranking of the CRRA parameter values. For instance, do the respondents who choose the least risky pension in the EG method (pension A) also have the highest average reported CS risk aversion? Figure 9 shows that this is indeed the case. Respondents who choose EG pension A have an average CS risk aversion of 13. For respondents who choose the increasingly risky EG pensions B, C and D, the average CS risk aversion is 8.5, 2.5 and 0, respectively. Both methods are thus consistent in terms of ordering of risk attitude, although we note that the average risk aversion parameter from the CS method is higher than the implied risk aversion parameter from the EG method.

Figure 17 in Appendix A shows a similar chart for the average CTB risk aversion parameter for each of the four possible EG answers. We observe a similar pattern,

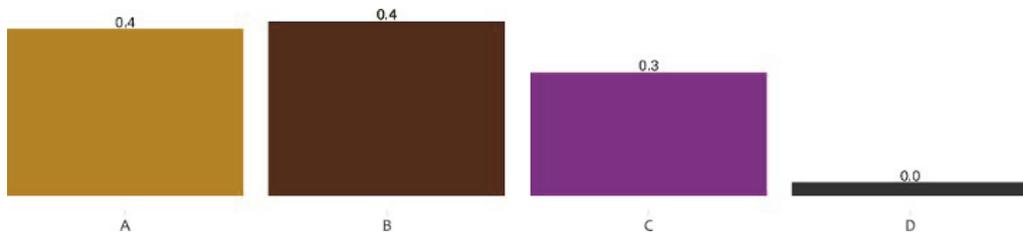Figure 9: Consistency between CS and EG, on average
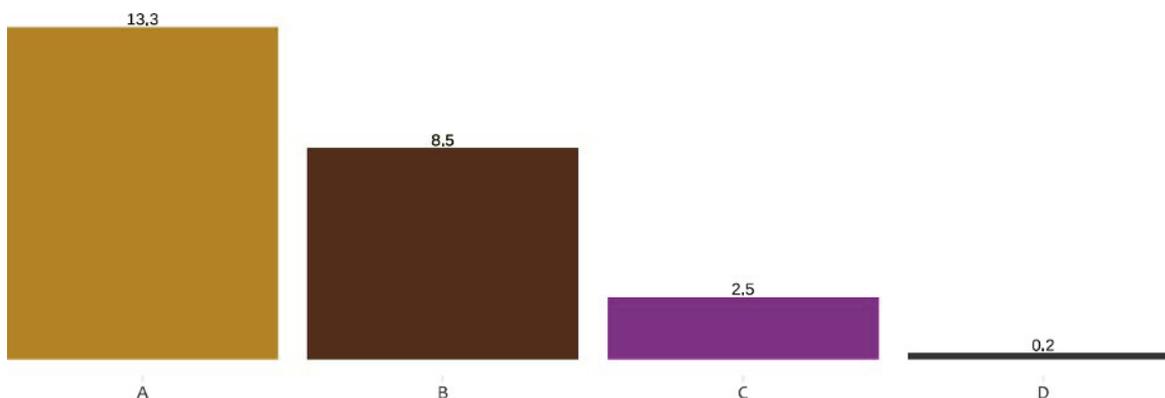
This figure presents on the horizontal axis the four EG options, where A is least risky and D is most risky, and above the charts the average CRRA parameter values elicited from CS.

Figure 10: Consistency between CS and EG, full distribution.
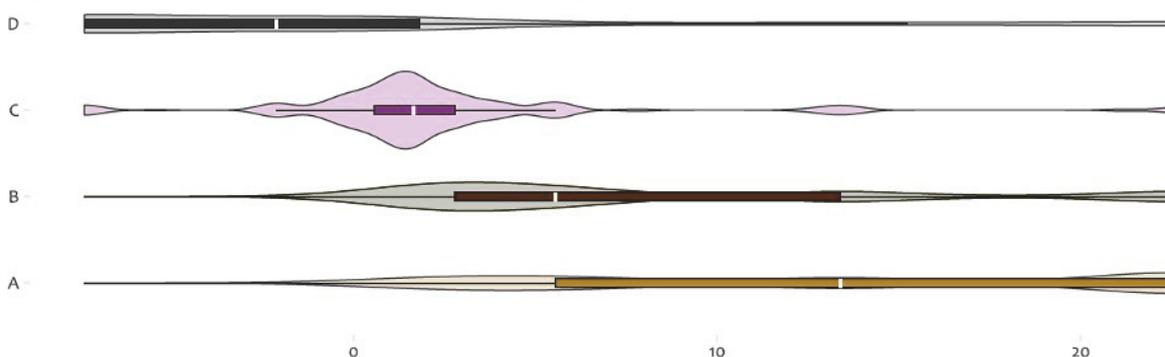
This figure presents on the rows the four EG options, where A is least risky and D is most risky, and on the horizontal axis the CRRA parameter values elicited from CS. The horizontal bars in the figure run from the 25th to the 75th percentile, which means that at least 50 percent of the respondents has a risk aversion falling inside the bar. The white line equals the median.

where respondents who choose the most risky EG pension D (only a small group) also have the lowest average CTB risk aversion. Between the groups answering A, B or C in the EG method, the difference in average CTB risk aversion is small.

If we plot not only the average risk aversion but also the full distribution of CS risk aversion conditional on the respondents' EG choice, we see a similar pattern. Figure 10 shows the results of this analysis. The median risk aversion by CS (indicated by the white line) is higher whenever the risk aversion according to EG is higher. Moreover, CS risk aversion is most concentrated around the median for respondents opting for EG pension C. For the other EG pensions, A, B or D, the dispersion in CS risk aversion is larger.

In an additional analysis, we examine the differences in consistency per presentation format. Specifically, we look at CS risk aversion conditional on the respondents' EG choice. We find that differences between game and questionnaire are small to negligible, in terms of both average risk aversion levels and distributional characteristics (see Figure 15 and Figure 16 in Appendix A).

4.4 Differences between risk preference methodologies

The previous section showed that the risk preference parameters from the three methodologies are generally positively related in terms of correlations and rankings, but that the relations are not perfect. For a completely rational individual (i.e., homo economicus), different elicitation methods should yield the same risk preference since the method of elicitation should not matter for the estimated risk preferences (Pedroni et al., 2017). However, we find evidence that participants' preferences for risk appear to change when measured using different methods, because correlation coefficients are not perfectly equal to one (see Table 8), nor are the rankings perfectly stable on an individual and aggregate level (see Figure 10). Imperfect correlations are to be expected due to decision and measurement errors. For this reason, Pedroni et al. (2017) and Gillen et al. (2019) recommend using more than one risk elicitation method within the same sample. Moreover, our observations are consistent with Dave et al. (2010) and Crosetto and Filippin (2016), who also compare the outcomes of several risk elicitation methods.

We thus find that different risk preference elicitation methods yield varying risk preference estimates at the individual and aggregate level. This finding is since recently referred to as the risk elicitation puzzle (Pedroni et al., 2017; Holzmeister and Stefan, 2021). Stated differently, in line with previous research, we observe heterogeneity in risk preferences across different elicitation methods. What is particularly challenging about the risk elicitation puzzle is the question how to properly interpret the observed variation in risk preferences: (i) are the varying risk preferences the result of unstable preferences, or (ii) do different methods stimulate distinct preferences, or (iii) are the participants' preferences inconsistent (Holzmeister and Stefan, 2021)? Another interesting question is whether the inconsistencies observed are stronger than what could be explained by decision and measurement errors. It is beyond the scope of the current paper to study these mechanisms. However, using the same sample as in the current paper, Goossens et al. (2023) follow up on these questions. Their study provides preliminary evidence that part of the observed heterogeneity in risk preferences can be explained by independent choices (i.e., random behavior)

between the three risk elicitation methods, which is in line with the findings of Holzmeister and Stefan (2021).

5. Conclusion

Given the importance of the tentative new pension legislation, which requires pension funds and insurers to elicit risk preferences at least every five years, we study whether the presentation format of a risk preference method influences elicited risk aversion levels. We find that risk aversion levels are similar in a traditional question-naire and a serious game, both at the aggregate level and per risk elicitation method.

We compare three well-known risk elicitation methods – a single choice list, a choice sequence list, and convex time budgets – and find that elicited risk aversion levels can differ substantially depending on the method used. This raises questions about the determination of the optimal investment strategy. Older respondents and blue-collar workers are more risk-averse than young and white-collar workers, an observation that could be taken into account in the investment strategy for these groups. If we assume that risk attitude is a function of both willingness and ability to carry risk, it would be worthwhile to study the relative importance of willingness and ability in explaining risk attitude. We leave this question for further research.

The type of presentation format does not have a significant influence on non-re-sponse after the first question. However, the response rate until the first question was lower among participants of the serious game (maybe because of technical issues, as the invitation email was the same for both the game and the questionnaire). The format of presentation has no effect on self-reported cognitive (un)certainty. We do observe, however, that respondents to the game use an automatic pilot less often when answering the choice sequence questions (i.e., answering all questions with 'A' or all 'B').

In general, blue-collar workers show lower completion rates, less self-reported cer-tainty, and more automatic pilot behavior than white-collar workers. These effects are only partly offset by offering a game to blue-collar workers. Also, consistency between risk elicitation methods is lower for blue-collar workers. The effect on consistency of offering a game to blue-collar workers is ambiguous. However, white-collar workers show somewhat higher consistency when offered the game.

Although we find evidence for varying risk preferences across elicitation methods, we still believe it is fruitful to elicit risk preferences in order to better understand pension fund participants. The current literature (e.g., Pedroni et al., 2017) suggests using more

than one risk preference elicitation method to arrive at a robust risk preference measure. As the relations between risk preferences across different methods are generally positive, there appears to exist a general risk preference factor. In the case of collective delegated decision-making, policymakers and board members can use these individually elicited preferences to construct a collective risk preference measure. Balter and Schweizer (2021) provide guidance on how to create a collective preference measure for a group of heterogeneous participants.

References

Alekseev, A., G. Charness and U. Gneezy (2017). "Experimental methods: When and why contextual instructions are important". In: Journal of Economic Behavior & Organization 134, pp. 48–59.

Alserda, G., B. Dellaert, L. Swinkels and F. van der Lecq (2019). "Individual pension risk preference elicitation and collective asset allocation with heterogeneity". In: Journal of Banking and Finance 101, pp. 206–225.

Amador-Hidalgo, L., P. Brañas-Garza, A. Espín, T. García-Muñoz and A. Hernández-Román (2021). "Cognitive abilities and risk-taking: Errors, not preferences". In: European Economic Review 134.

Andreoni, J. and C. Sprenger (2012). "Estimating time preferences from convex budgets". In: American Economic Review 102.7, pp. 3333–3356.

Balter, A. and N. Schweizer (2021). "Robust decisions for heterogeneous agents via certainty equivalents". In: Working paper.

Barsky, R., F. Juster, M. Kimball and M. Shapiro (1997). "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study". In: Quarterly Journal of Economics 112.2, pp. 537–579.

Cohen, J., K. Ericson, D. Laibson and J. White (2020). "Measuring time preferences". In: Journal of Economic Literature 58.2, pp. 299–347.

Crosetto, P. and A. Filippin (2016). "A theoretical and experimental appraisal of four risk elicitation methods". In: Experimental Economics 19, pp. 613–641.

Dave, C., C. Eckel, C. Johnson and C. Rojas (2010). "Eliciting risk preferences: When is simple better?", Journal of Risk and Uncertainty 41, pp. 219–243.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2018). "On the relationship between cognitive ability and risk preference." In: Journal of Economic Perspectives 32.2, pp. 115–134.

Eckel, C. and P. Grossman (2002). "Sex differences and statistical stereotyping in attitudes toward financial risk". In: Evolution and Human Behavior 23.4, pp. 281–295.

Eckel, C. and P. Grossman (2008). "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices". In: Journal of Economic Behavior & Organization 68, pp. 1–17.

Enke, B. and T. Graeber (2022). "Cognitive uncertainty". In: NBER Working Paper.

Epstein, L. and S. Zin (1989). "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework". In: Econometrica 57.4, pp. 937–969.

Frey, R., A. Pedroni, R. Mata, J. Rieskamp and R. Hertwig (2017). "Risk preference shares the psychometric structure of major psychological traits". In: Science Advances 3.10.

Friedman, D., S. Habib, D. James and B. Williams (2022). "Varieties of risk preference elicitation". In: Games and Economic Behavior 133, pp. 58–76.

Gillen, B., E. Snowberg and L. Yariv (2019). "Experimenting with measurement error: Techniques with applications to the Caltech cohort study." In: Journal of Political Economy 127.4, pp. 1826–1863.

Goossens, J. and M. Knoef (2022a). "Time-varying risk and time preferences: Relation with trading behavior". In: Netspar Academic Paper DP 06/2022-023.

Goossens, J., M. Knoef and E. Ponds (2022b). "Can estimated risk and time preferences explain real-life financial choices?". In: Netspar Academic Paper DP 06/2022-024.

Goossens et al. (2023). "The risk elicitation puzzle: A potential resolution". In: Working paper.

Hackethal, A., M. Kirchler, C. Laudenbach, M. Razen and A. Weber (2023). "On the role of monetary incentives in risk preference elicitation experiments". In: Journal of Risk and Uncertainty 66.2, pp. 189– 213.

Holzmeister, F. and M. Stefan (2021). "The risk elicitation puzzle revisited: Across-methods (in) consistency?" In: Experimental Economics 24, pp. 593–616.

Meeren, G. van der, H. de Cloe-Vos and A. van Geen (2019). "Meet risicobereidheid met een kwantitatieve methode". In: ESB 4773, pp. 222-225.

Pedroni, A., R. Frey, A. Bruhin, G. Dutilh, R. Hertwig and J. Rieskamp (2017). "The risk elicitation puzzle". In: Nature Human Behavior 1, pp. 803–809.

Potters, J., A. Riedl and P. Smeets (2016). "Towards a practical and scientifically sound tool for measuring time and risk preferences in pension savings decisions". In: Netspar Industry Paper 59.

Schildberg-Hörisch, H. (2018). "Are Risk Preferences Stable?" In: Journal of Economic Perspectives 32.2, pp. 135–154.

Smith, V. (1976). "Experimental economics: Induced value theory" In: American Economic Review 66.2, pp. 274-279.

Smith, V. (1982). "Microeconomic systems as an experimental science" In: American Economic Review 75.5, pp. 923-955.

Appendix A

Table 10: Choice sequence risk aversion task

Question	Sequence	Risk aversion	Risky (R) High	Risky (R) Low	Non-risky (N) High	Non-risky (N) Low	Risk aversion after risky	Risk aversion after non-risky
1		14.50	3270	2050	2380	2080	2.49	10.07
2	N	−0.50	2640	1040	2050	1800	4.05	11.60
2	R	−0.50	2640	1040	2050	1800	−0.12	3.52
3	NN	11.60	3550	2210	2580	2260	3.98	16.72
3	NR	4.05	3550	2020	2580	2260	−0.43	13.68
3	RN	3.52	3550	1970	2580	2260	1.44	6.67
3	RR	−0.12	3550	1220	2580	2260	−1.90	1.65
4	NNN	16.72	2980	1880	2170	1900	14.11	20.53
4	NNR	3.98	2980	1690	2170	1900	1.64	8.11
4	NRN	13.68	2980	1870	2170	1900	6.96	17.71
4	NRR	−0.43	2980	820	2170	1900	−2.19	1.10
4	RNN	6.67	2980	1796	2170	1900	4.11	10.35
4	RNR	1.44	2980	1450	2170	1900	0.44	2.65
4	RRN	1.65	2980	1480	2170	1900	0.01	4.46
4	RRR	−1.90	2790	1070	2370	2100	−4.92	−0.29
5	NNNN	20.53	3810	2420	2780	2430	17.96	22.68
5	NNNR	14.11	3810	2400	2780	2430	12.22	15.21
5	NNRN	8.11	3810	2330	2780	2430	4.78	13.77
5	NNRR	1.64	3810	1900	2780	2430	0.58	2.96
5	NRNN	17.71	3810	2420	2780	2430	14.71	21.18
5	NRNR	6.96	3810	2300	2780	2430	1.82	13.00
5	NRRN	1.10	3810	1800	2780	2430	−0.19	3.78
5	NRRR	−2.19	3520	1770	3090	2730	−5.43	−0.38
5	RNNN	10.35	3810	2370	2780	2430	7.42	13.40
5	RNNR	4.11	3810	2170	2780	2430	2.79	5.55
5	RNRN	2.65	3810	2050	2780	2430	1.65	3.91
5	RNRR	0.44	3810	1590	2780	2430	−0.12	1.08
5	RRNN	4.46	3810	2200	2780	2430	2.11	7.81
5	RRNR	0.01	3810	1400	2780	2430	−0.65	0.72
5	RRRN	−0.29	3810	1190	2780	2430	−0.85	0.29
5	RRRR	−4.92	3410	1690	3190	2830	−7.41	−2.12

Subjects choose a pension per question. Each question involves a choice between a risky and a less risky pension. A pension, whether risky or less risky, involves a 50/50 chance of a low or high payoff. "Non-risky" in the table below should be understood as a less risky pension. The implied risk aversion is based on the power utility function $U(x) = \dfrac{x^{1-\gamma}}{1-\gamma}$.

Table 11: Eckel–Grossman risk aversion task

Choice	Low payoff	High payoff	Exp. Return	St. Dev.	Implied CRRA range
Pension 1	1970	2050	2010	57	$\gamma > 4.37$
Pension 2	1900	2150	2025	177	$1.84 < \gamma < 4.37$
Pension 3	1500	3000	2250	1061	$-0.8 < \gamma < 1.84$
Pension 4	1100	3200	2150	1485	$\gamma < -0.80$

Subjects choose a pension, every pension involving a 50/50 chance of a low or high payout. The implied Coefficient of Relative Risk Aversion (CRRA) range is based on the power utility function $U(x) = \dfrac{x^{1-\gamma}}{1-\gamma}$. Each range is calculated by equalizing the gamble to its neighbors, and computing the value of γ that makes the individual indifferent in utility between each adjacent gamble.

Table 12: Convex time budgets risk aversion task

Decision	t	k	c_t	c_{t+k}	1+r	Annual
1	1	10	10000	10000	1.00	0.00
2	1	10	10000	16000	1.60	4.81
3	1	10	10000	26000	2.60	10.03
4	1	5	10000	10000	1.00	0.00
5	1	5	10000	16000	1.60	9.86
6	1	5	10000	26000	2.60	21.06

Choice sets in the convex time budgets. t and k are front and end delays in years, and c_t and c_{t+k} are allocated amounts in euros. $1 + r$ is the implied gross interest rate. Annual r is the yearly interest rate in percentage terms, calculated as $((1 + r)^{1/k} - 1) \times 100$.

Table 13: Overview of risk preference estimates for serious game

method	obs	mean	std	5%	25%	median	75%	95%
CS	726	8.28	8.14	−0.12	2.11	5.55	13.40	22.68
EG	726	3.16	1.84	0.50	3.00	3.00	5.50	5.50
CTB	726	0.36	1.80	−2.10	0.06	0.46	1.30	2.21
Composite	726	3.93	3.09	0.02	1.62	3.12	6.36	9.61
CTB delta	726	1.01	0.23	0.78	0.89	0.95	1.07	1.47

This table shows the summary statistics for the CRRA parameter values per methodology, in the serious game. 'Composite' is the unweighted average of the CRRA parameter values from CS, EG, and CTB. 'CTB delta' is the simultaneously estimated time preference parameter from the convex time budgets.

Table 14: Overview of risk preference estimates for traditional questionnaire

method	obs	mean	std	5%	25%	median	75%	95%
CS	582	9.31	8.70	−0.12	2.79	5.55	17.27	22.68
EG	582	3.28	1.91	0.50	3.00	3.00	5.50	5.50
CTB	582	0.42	1.96	−0.62	0.05	0.21	0.81	1.65
Composite	582	4.34	3.32	0.01	1.94	3.56	6.92	9.66
CTB delta	582	1.06	0.30	0.82	0.90	0.97	1.07	1.47

This table shows the summary statistics for the CRRA parameter values per methodology, in the traditional questionnaire. 'Composite' is the unweighted average of the CRRA parameter values from CS, EG, and CTB. 'CTB delta' is the simultaneously estimated time preference parameter from the convex time budgets.

Figure 11: Average EG risk aversion levels in the serious game and the traditional questionnaire, by worker type
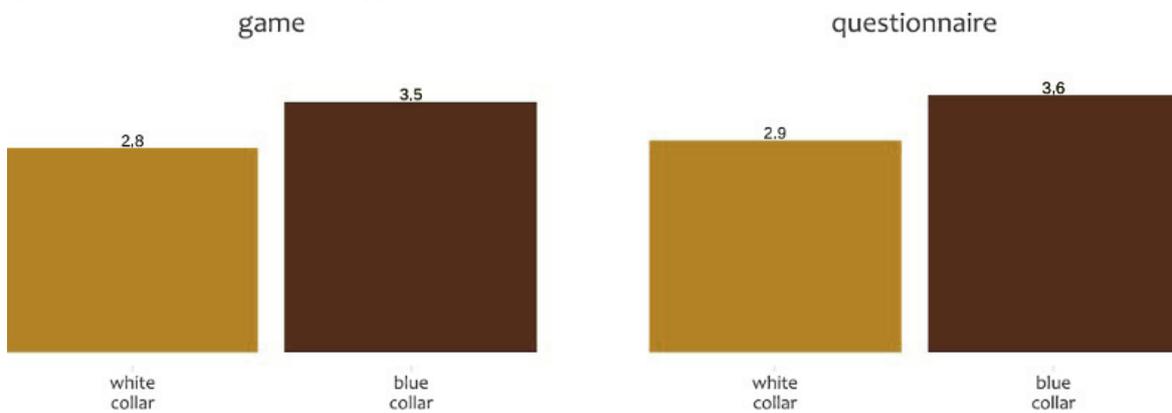
The height of the bars represents the average CRRA parameter values.

Figure 12: Average CTB risk aversion levels in the serious game and the traditional questionnaire, by worker type
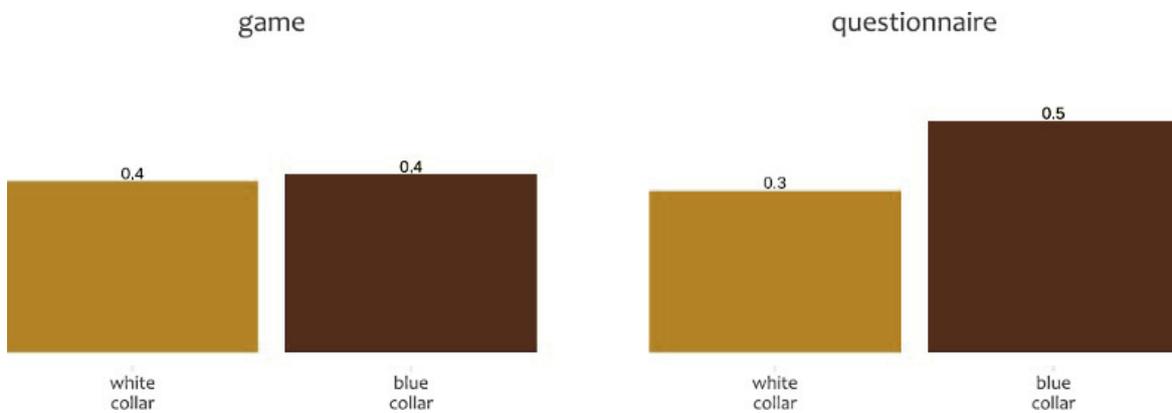
The height of the bars represents the average CRRA parameter values.

Figure 13: Average EG risk aversion levels in the serious game and the traditional questionnaire, by age bucket
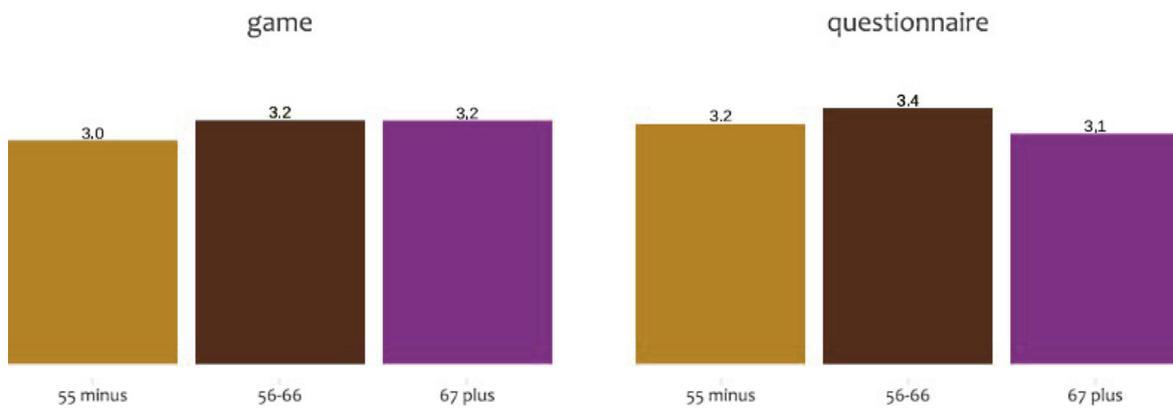
The height of the bars represents the average CRRA parameter values.

Figure 14: Average CTB risk aversion levels in the serious game and the traditional questionnaire, by age bucket
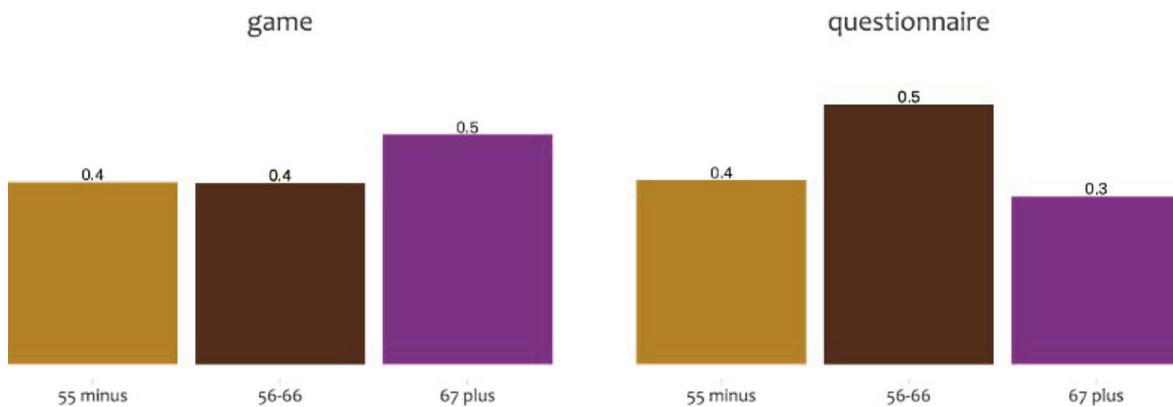
The height of the bars represents the average CRRA parameter values.

Figure 15: Consistency between CS and EG for each presentation format, on average
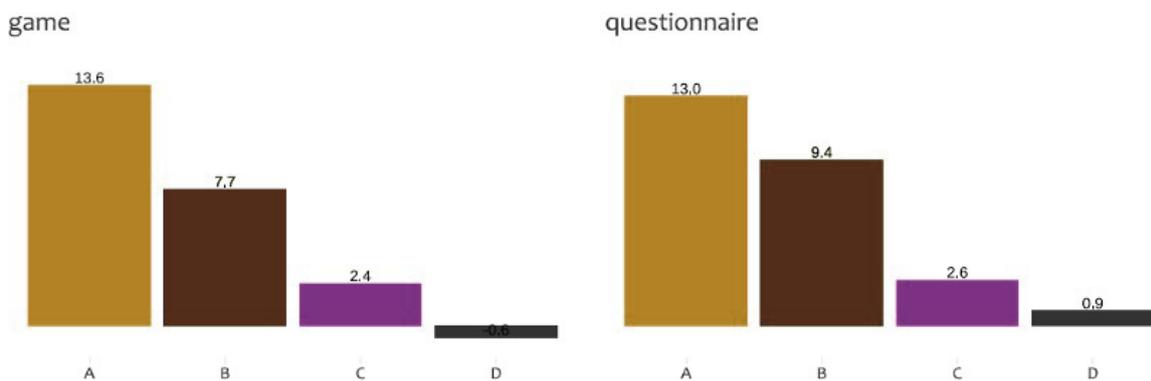
This figure presents on the horizontal axis the four EG options, where A is least risky and D is most risky, and above the charts the average CRRA parameter values elicited from CS. We distinguish both presentation formats.

Figure 16: Consistency between CS and EG for each presentation format, distribution

This figure presents on the rows the four single choice options, where A is least risky and D is most risky, and on the horizontal axis the CRRA parameter values elicited from CS. The horizontal bars in the figure run from the 25th to the 75th percentile, which means that at least 50 percent of the respondents has a risk aversion falling inside the bar. The white line equals the median. We distinguish both presentation formats.

Figure 17: Consistency between CTB and EG for each presentation format, on average

This figure presents on the horizontal axis the four EG options, where A is least risky and D is most risky, and above the charts the average CRRA parameter values elicited from CTB.

Appendix B

Intro text (new page):
Thank you for participating in this bpfBOUW survey.
You are in charge of your pension and decide the level of risk!
We are going to look at pensions, and you make your own choice. There are no right or wrong answers.

Your choices help to make your pension better and more personal.
In this study, assume that the prices of products and services will not change in the future. Click on the arrow to start.

You now start with block 1. *(new page, CS)*

Question 1 (new page):
Which pension would you choose?

With pension A you receive a total of €2,380 on your bank account every month if things turn out better than expected, or €2,080 if things are worse than expected. With pension B you will receive a total of €3,270 in your bank account every month if things are better than expected, or €2,050 if things are worse than expected. The chance that it will be better or worse is the same (50% each), just like with heads or tails.

	Tails	Heads
A	€2,380	€2,080
B	€3,270	€2,050

Question 2 (new page):
We ask you to choose pension A or B 4 more times. Note: the amounts change. Which pension would you choose?

[INSERT NEXT CHOICE TABLE, depends on previous choice]

Question 3 (new page):
Which pension would you choose?

[INSERT NEXT CHOICE TABLE, depends on previous choice]

Question 4 (new page):
Which pension would you choose?

[INSERT NEXT CHOICE TABLE, depends on previous choice]

Question 5 (new page):
Which pension would you choose?

[INSERT NEXT CHOICE TABLE, depends on previous choice]

Question 6 (new page):
How sure are you about the answers you have just given? [Very unsure, unsure, sure, very sure]

You have completed block 1. You now start with block 2. *(new page, EG)*

Question 7 (new page):
There are now 4 pensions to look at. You can only choose one. The pension amounts differ when things are better than expected and worse than expected. Which pension would you choose?

	Tails	Heads
A	€ 2,050	€ 1,970
B	€ 2,150	€ 1,900
C	€ 3,000	€ 1,500
D	€ 3,200	€ 1,100

Question 8 (new page):
How sure are you about the answers you have just given? [Very unsure, unsure, sure, very sure]

You have completed block 2. You now start with block 3. *(new page, CTB)*

Question 9 (new page):

You are given a voucher. You can buy everything with it in the coming years. With voucher A you can spend €10,000 in 1 year and €0 in 10 years. With the other vouchers you will receive less in 1 year, but more later. Which voucher would you choose?

	Valid in 1 year	Valid in 10 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 3,000
Voucher C	€ 3,000	€ 7,000
Voucher D	€ 0	€ 10,000

Question 10 (new page):

We ask you to choose a voucher 2 more times. Note: the amounts change.

Which voucher would you choose?

	Valid in 1 year	Valid in 10 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 4,800
Voucher C	€ 3,000	€ 11,200
Voucher D	€ 0	€ 16,000

Question 11 (new page):

Which voucher would you choose?

	Valid in 1 year	Valid in 10 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 7,800
Voucher C	€ 3,000	€ 18,200
Voucher D	€ 0	€ 26,000

Question 12 (new page):

You will receive new vouchers. Now you can spend an amount **in 1 year and in 5 years**. Which voucher would you choose?

	Valid in 1 year	Valid in 5 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 3,000
Voucher C	€ 3,000	€ 7,000
Voucher D	€ 0	€ 10,000

Question 13 (new page):

We ask you to choose a voucher 2 more times. Note: the amounts change. Which voucher would you choose?

	Valid in 1 year	Valid in 5 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 4,800
Voucher C	€ 3,000	€ 11,200
Voucher D	€ 0	€ 16,000

Question 14 (new page):

Which voucher would you choose?

	Valid in 1 year	Valid in 5 years
Voucher A	€ 10,000	€ 0
Voucher B	€ 7,000	€ 7,800
Voucher C	€ 3,000	€ 18,200
Voucher D	€ 0	€ 26,000

Question 15 (new page):

How sure are you about the answers you have just given? [Very unsure, unsure, sure, very sure]

End (new page):

You have reached the end of this study. Thanks for answering the questions!

Your answers will help us to provide you with the best possible service in the future. You can now close this screen.

OVERZICHT UITGAVEN
IN DE DESIGN PAPER SERIE

Network for Studies on Pensions, Aging and Retirement