Master Thesis
M.Sc. in Economics and Business
Specialisation in Financial Economics
Erasmus School of Economics
Erasmus University Rotterdam

Comparing risk preference elicitation methods for pensions

10 October 2017

Author: Gust van der Meeren

Supervisor: Dr. Alexandra van Geen

Abstract

This paper examines which risk preference elicitation methods performs best in the pension domain. Firstly, I select the most viable methods for the pension domain based on four criteria: it must produce CRRA parameters, use realistic pension scenarios, test rational risk preferences, and are easy and short. The methods are a single choice, based on Eckel and Grossman (2002), a choice list, based on Holt and Laury (2002), and a choice sequence, similar to Barsky et al. (1997). Secondly, I adapt the methods to the pension domain in four ways, i.e. with respect to their presentation, number of risk preference categories, CRRA ranges, and income-based pension scenarios. I include these three methods in a within-subject experiment ($N = 426$) through the *AFM Consumentenpanel*, to determine which performs best in the pension domain. I rank each elicitation methods on two levels to indicate good performance, i.e. similarity of their risk preferences to those of (1) the other methods and (2) two risk preference questionnaires. The experiment shows that the three methods' mean risk preferences are quite similar (from $r = 1.42$ to $r = 1.85$) and correlate strongly (from $\rho = 0.62$ to $\rho = 0.67$). The methods' risk preferences also correlate weakly to moderately with those of the two questionnaires, i.e. the DOSPERT scale (from $\rho = 0.08$ to $\rho = 0.11$) and the SOEP questions (from $\rho = 0.29$ to $\rho = 0.39$). The choice list performs best and the choice sequence performs second-best on the first comparison level and vice versa for the second comparison level; the single choice performs worst on both. Therefore, I conclude that the choice list and choice sequence together are the best-performing methods in the pension domain.

Keywords

Risk preference · Elicitation method · Private pension · Household finance · Experiment

JEL classifications

C81 · C91 · D1 · D81 · G11 · J32

1. Introduction

According to Dutch law, customers' risk preferences must be considered by pension funds when they invest their pensions. However, there is no academic consensus on how to properly elicit risk preferences for the pension domain. Several literature surveys find pros and cons for all risk preference elicitation methods without concluding which perform better (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). Numerous comparisons elicit risk preference with two or more methods, but are similarly inconclusive (e.g. Deck et al. 2010; Raynaud and Couture 2012; Crosetto and Filippin 2016). No studies take a bottom-up approach to select the most viable methods among all available alternatives. Little research examines risk preference elicitation methods in specific domains, specifically pensions Alserda et al. (2016) and Dellaert et al. (2016) each propose one method for risk preference elicitation in the pension domain, but make no extensive comparison with the alternative methods. This paper fills this gap in the literature It examines the viability of different risk preference elicitation methods for the pension domain and compares them in an experiment to determine which performs best.

To start with, I identify the predominant risk preference elicitation methods in the literature based on surveys of risk preference elicitation by Charness, Gneezy, and Imas (2013) and Holt and Laury (2014).[1] Subsequently, I determine which of these risk preference elicitation methods are viable for the pension domain. For this purpose, I develop four viability criteria. Firstly, the measurability criterion requires methods to produce a constant relative risk aversion (CRRA) parameter.[2] This allows for absolute risk preference that are easily applicable to pension funds' investment policies through modern portfolio theory (Holt and Laury 2014; Markowitz (1952a). Secondly, the appropriateness criterion requires methods to use prospects that form realistic pension scenarios. Given that risk preferences vary across domains (Weber, Bais, and Betz 2002), even between finance and pensions (Van Rooij, Kool, and Prast 2007), it is important to measure the appropriate risk preferences. Thirdly, the rationality criterion requires methods to avoid or minimise irrationalities. Because the method elicits rational risk preferences under an Expected Utility Theory (EUT) model, effects such as reference points, probability weighting, loss aversion, and framing can undermine proper measurement (e.g. Kahneman and Tversky 1979; Tversky and Kahneman 1981; 1992). Fourthly, the feasibility criterion requires methods to not require exceptional cognitive abilities or take more than a few minutes to complete. The complexity of methods can be prohibitive (Charness, Gneezy, and Imas 2013), which is especially troublesome in the pension domain where the population has heterogeneous cognitive abilities. Furthermore, they generally are not willing to spend much effort on their pension decisions (Van Rooij, Kool, and Prast 2007).

Based on the four viability criteria, I select those risk preference elicitation methods that are viable for the pension domain. There are three: the multiple lottery choice (hereafter the EG method, from Eckel and Grossman 2002); the choice list with only risky prospects (hereafter the HL method,

[1] The risk preference elicitation methods are derived from two surveys (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). I divide them in five categories and several subcategories: (1) single choices, namely multiple lottery choices (Eckel and Grossman 2002), investment games (Gneezy and Potter 1997), and certainty equivalents (Harrison 1986); (2) choice lists with certain payoffs on one side (Dohment et al. 2005; Murningham, Roth, and Schoumaker 1988; Bruner 2009) and with risky prospects on both sides (Holt and Laury 2002); (3) choice sequences with certain payoffs as on option (Barksy et al. 1997) and with only risky prospects as options; (4) dynamic methods, namely BART (Lejuez et al. 2002), BRET (Crosetto and Filippin 2013a) and DOND (Post et al. 2008); and (5) questionnaires, namely the DOSPERT scale (Weber, Blais, and Betz 2002) and SOEP questions (Dohmen et al. 2005).

[2] Risk preference elicitation methods often quantify risk preferences by modelling risk-taking behaviour with an EUT model and CRRA (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). To allow for such modelling, subjects are asked to choose between quantified prospects. They vary in riskiness due to different outcomes and probabilities. Based on a subject's preferences between prospects, it is possible to model their risk preferences and hence predict their preferences in other risk-taking decisions. This makes them absolute risk preferences. See for a more elaborate discussion of the EUT model with CRRA Subsection 2.1.

from Holt and Laury 2002); and the choice sequence with only risky prospects (hereafter the CS method). The other methods do not meet one or more criteria. Questionnaires do not meet the measurability and rationality criteria, because their risk preferences are qualitative and their questions are prone to framing. The investment game and certainty equivalent do not meet the rationality criterion because of several irrational effects and the latter also does not meet the feasibility criterion because of the complexity it requires to avoid the previous effects. Choice lists and sequences do not meet the rationality criterion when they have a certain payoff, mainly due to the certainty effect. The dynamic methods do not meet the appropriateness criterion because of their own game-like settings do not transfer to pension scenarios.

In order to test and compare the three viable risk preference elicitation methods in the pension domain, I make four adaptations with regards to the methods' presentation, risk preference categories, CRRA ranges, and custom pension scenarios. Firstly, I adapt the presentation of the prospects in two ways. The prospects are introduced as pension products (i.e. different pension incomes with their own probabilities) that subjects must choose between, to ensure I elicit risk preferences specific to the pension domain, since these may differ across domains (Weber, Bais, and Betz 2002; Van Rooij, Kool, and Prast 2007). I also employ pie charts to visualise the probabilities of the pension incomes, which makes them easier to understand and is common in risk preference elicitation (e.g. Lipkus and Hollands 1999; Deck et al. 2010). Secondly, I equalise the risk preference categories of all three methods to ten. The literature includes many variations on the number of categories, for instance, six by Binswanger (1981) and five by Eckel and Grossman (2002) in case of the EG method. By equalising the categories, all three methods have the same degree of precision in their measurement of risk preferences. Thirdly, I also equalise the CRRA ranges that correspond to the risk preference categories. For this purpose, I use the ranges of Alserda et al. (2016) for the HL method and import them to the other two methods. They also use it in the pension domain and it provides a broad spectrum from very risk averse ($r \geq 4.46$) to very risk seeking ($r \leq -4.82$) with a neutral midpoint ($r = 0$) to avoid framing (Anderson et al. 2006).[3] Fourthly, I compute pension products (i.e. pension incomes and their probabilities) based on subjects' incomes. Alserda et al. (2016) already provide multiplication factors for the HL method and I devise them for the other two methods. These, together with the known income categories of my subjects, provide considerable customisation and create realistic pension scenarios.

Now that the three viable risk preference elicitation methods are adopted to the pension domain, I compare their performance in an experiment. It is difficult to determine the best-performing elicitation method in the pension domain. Whereas earlier studies validate methods' elicited risk preferences to subjects' risk-taking behaviour (e.g. Anderson and Mellor 2009; Barsky et al. 1997), for pensions risk-taking behaviour does not reflect true risk preferences due to numerous intermediary influences. An alternative way of validating methods is to compare their elicited risk preferences to other measurements of risk preferences (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). Assuming that these measurements are accurate on average, similarity between risk preference measurements indicates good performaning elicitation methods. I compare the elicitation methods' risk preferences on two levels: a comparison between different risk preference elicitation method (e.g. Anderson and Mellor 2009; Deck et al. 2010) and a comparison to two questionnaires, the DOSPERT scale and the SOEP question (e.g. Dohmen et al. 2011; Kapteyn and Teppa 2011). For this purpose, I use four hypotheses, two per comparison level: (1) the elicitation methods produce the same risk preferences; (2) the elicitation methods produce perfectly correlated risk preferences; (3) the elicitation methods' risk preferences are perfectly correlated with the questionnaires'; and (4) the elicitation methods' risk preferences are

[3] I quantify risk preferences in a model based on the Expected Utility Theory (EUT) formula: $EU(x) = \sum_{i=1}^{n} \Pr(x_i)\, U(x_i)$ and constant relative risk aversion (CRRA) formula: $U(x_i) = (x_i^{1-r})/(1-r)$. Where an individual with risk preference $r > 0$ is risk averse, one with $r = 0$ is risk neutral, and one $r < 0$ is risk seeking. See for a more elaborate discussion of the model Subsection 2.1.

explained by the questionnaires. I subsequently rank the EG, HL, and CS methods according to their approximation of each hypotheses and aggregate their ranks for each comparison level. The sum of these ranks determines which method performs best in the pension domain.

To collect the data for this comparison, I run an experiment in the voluntary consumer panel of Dutch Authority of Financial Markets, i.e. the *AFM Consumentenpanel*, which strives to cover a representative sample of the Dutch population. Participants select their income out of five categories when they join the panel. I use this information to customise the pension incomes. Subjects (N = 426) complete all three risk preference elicitation methods in randomised order for a within-subject design. Subsequently, all subjects also complete the DOSPERT scale, the SOEP questions, two conventional methods (which pension fund currently employ to elicit risk preferences), and a numeracy test. Furthermore, some subjects' financial literacy is available from an earlier experiment.

The first comparison level involves the similarity between the three risk preference elicitation methods. Hypothesis 1 states that the elicitation methods produce the same risk preferences. Previous comparisons between the EG and HL methods find only limited similarity: e.g. Crosetto and Filippin (2016) in a between-subjects design (r = 0.09 and r = 0.66, respectively); and Reynaud and Couture (2012) in a within-subjects design (r = 0.36 and r = 1.02). In my experiment, I find more similar risk preferences (between r = 1.42 and r = 1.85) for the EG, HL, and CS method. The HL method's risk preferences are most similar to those of the other two and hence perform best on the first hypothesis. Hypothesis 2 states that the elicitation methods produce perfectly correlated risk preferences. Correlations in the literature vary: e.g. Deck et al. (2010) find weak a correlation between the EG and HL method (ρ = 0.21), while Reynaud and Couture (2012) find stronger ones (ρ = 0.4 and ρ = 0.67, for respectively low and high payoffs); for the HL and CS method, Anderson and Mellor (2009) find a weak correlation again (ρ = 0.18). The risk preferences of the elicitation methods in my experiment are moderately to strongly correlated (between ρ = 0.62 and ρ = 0.67). The differences between the elicitation methods' correlations are not statistically significant, thus the methods' performance is the same.

The second comparison level involves the similarity between the elicitation methods' risk preferences and those of two questionnaires, the DOSPERT scale and SOEP questions. Hypothesis 3 states that the elicitation methods' risk preferences are perfectly correlated with the questionnaires'. Other experiments do not always find such correlations and if they do the correlations are weak: e.g. Reynuad and Couture (2016) find weak correlations for the EG and HL methods and the DOSPERT scale (between ρ = 0.26 and ρ = 0.40) and none for the SOEP question; and Crosetto and Filippin (2016) find similar correlations (from insignificant correlations to between ρ = 0.23 and ρ =0.30). My results, although all significant, are not much different for the correlations of the EG, HL, and CS methods with the DOSPERT scale (from ρ = 0.08 to ρ = 0.11) and SOEP question (from ρ = 0.27 ato ρ = 0.39). The correlations are insignificantly different for the DOSPERT scale and for the SOEP question the ranking is (from high to low correlation): the CS, HL, and EG method. Hypothesis 4 states that the elicitation methods' risk preferences are explained by the questionnaires'. The explanatory power is determined by including each questionnaire in a regression of each methods' risk preferences on demographic variables and measuring the increase in adjusted R^2. Crosetto and Filippin (2016) also takes this approach and find increases of 1.6 and 2.7 percentage points for the EG method with respectively the DOSPERT scale and SOEP questions, but no increases for the HL method. Other studies find explanatory power based on the regression coefficient of the questionnaire (e.g. Deck et al. 2010; Dohmen et al. 2011). The increases in adjusted R^2 that I find lay between 2 and 14 percentage point. When I rank the increases in adjusted R^2, the EG method correlates the most with the DOSPERT scale and the least with the SOEP question. The opposite holds for the CS method and the HL method ranks second twice.

Overall, the HL and CS methods are thus both the best-performing risk preference elicitation method in the pension domain. Based on the four viability criteria, the EG, HL, and CS methods are the only viable methods for the pension domain. I adapt these methods to the pension domain and compare them in an experiment with a within-subject design. Its result indicate that the HL method performs best on my first comparison level, which compares the elicitation methods' risk preferences. The CS method performs best on my second comparison level, which compares the elicitation methods' risk preferences to those of the questionnaires. Together, the HL and CS share the first position for their performance, followed by the EG method, which comes in last on both comparison levels.

Additionally, I make three analyses independent of determining the best risk preference elicitation method for the pension domain. Firstly, I compare the two conventional methods to the EG, HL, and CS methods. For these methods to replace the conventional methods, they should perform superiorly in pension risk preference elicitation. For the questionnaires, the conventional methods' correlations are similar to those of the three methods, albeit slightly higher for the DOSPERT scale ($\rho = 0.29$ and $\rho = 0.33$) and slightly lower for the SOEP question (both $\rho = 0.30$). Interestingly, the conventional methods are not correlated with the three methods. Furthermore, the conventional methods' internal correlation ($\rho = 0.47$) is also significantly weaker than those of the three methods (from $\rho = 0.62$ to $\rho = 0.67$). This suggests the EG, HL, and CS methods produce more stable risk preferences than the conventional methods. Secondly, I search for explanations of inconsistent answer – which do not follow the EUT model and are hence problematic, because they do not produce usable risk preferences – by running a logistic regression of inconsistency on several demographic variables. This only explains some of the inconsistency in the HL method: it is partly explained by subjects' educational level ($\beta = -1.04$ for the high-level dummy) and numerical literacy ($\beta = -0.53$). For the HL method inconsistency is thus due to difficulty, while it remains unexplained for the CS method. Thirdly, I also look at the stability of subjects' risk preferences across the EG, HL, and CS methods, which is the ideal under the EUT model. About 20% of subjects has completely stable risk preferences. About 60% (80%) of subjects changes an average of $1^1/_3$ (2) risk preference category or less between the three methods. These results are like those of other studies (e.g. Anderson and Mellor 2009; Reynaud and Couture 2012). To examine the causes of instability further, I run a linear and logistic regressions of respectively stability and a stable dummy on several demographic variables, but I find no significant explanators. Therefore, it remains unclear what causes risk preference instability across the EG, HL, and CS methods.

My results have several implications for future research and practice in risk preference elicitation for pension. The four viability criteria I develop present a framework for assessing whether methods are viable or not. They also present the EG, HL, and CS methods as good candidates for risk preference elicitation in the pension domain. My four adaptations to the methods for a comparison in the pension domain are useful to both future pension experiments and comparisons. As is the case with the determination of performance based on the two comparison levels. More practically, the within-subject experiment in the pension domain produces relatively similar and strongly correlated risk preferences compared to other comparison in the literature. This underscore the validity of both my experiment and the use of these methods for risk preference elicitation in the pension domain. The fact that the conventional methods, in contrast, produce less correlated risk preferences suggests that they are less suitable for elicitation in the pension domain. Finally, inconsistencies remain a problem for widespread application of the risk preference elicitation methods in the pension domain. In the HL method, it partly results from low educational levels and numerical literacy, which underlines the importance of ease of use in method design. However, both here and in the entirely unexplained inconsistencies in the CS method, the relatively high educational level and numerical literacy of my sample does not yet bode well for population-wide application of these methods.

The setup of this paper looks as follows. It starts with the EUT and CRRA modelling to derive risk preferences from choices (Subsection 2.1). Thereafter, I give an overview of five risk preference elicitation method categories (2.2). In the third section, I develop (3.1) and apply (3.2) the four criteria to determine the risk preference elicitation methods that are viable for the pension domain. Section 4 explains how I rank the viable methods on their performance in the pension domain based on an experiment. This starts four adaptations of the viable methods to the pension domain that I construct based on the literature (4.1). Subsequently, I present my methodology, including my ranking mechanism and hypotheses (4.2). The section concludes with my experimental design (4.3). The fifth section gives the results of my experiment. It begins with a description of my sample and a first look at my results (5.1). I test my hypotheses and rank the EG, HL, and CS methods in Subsection 5.2. Afterwards, I discuss additional analyses on the conventional methods and explanations for risk preference inconsistency and instability (5.3). I conclude my thesis with a summary of my research and its implications and limitations in Section 6.

2. The five categories of risk preference elicitation methods found in the literature

2.1. Four out of five categories model rational risk preferences with EUT and CRRA

2.1.1. Rational risk preferences under Expected Utility Theory (EUT)

Four out of the five elicitation method categories model risk preferences with EUT and CRRA. It is important to understand how this works, before I elaborate upon the differences between these four categories. Risk preferences can be measured by asking an individual's preferences between prospects with different degrees of risk (Meyer 2014). For example, consider a choice between a safe and risky prospect with the same expected value. When an individual prefers the safe (risky) prospect, they are risk averse (seeking), and when they are indifferent, they are risk neutral. By altering the riskiness and expected values, risk preferences can be further specified. The risk-taking behaviour in these choices is linked to risk preferences with the Expected Utility Theory (Bernoulli 1738; Von Neumann and Morgenstern 1944; Savage 1954). This theory is summarised by the following formula:

$$EU(x) = \sum_{i=1}^{n} \Pr(x_i)\, U(x_i)$$

where $EU(x)$ is the expected utility of prospect x, x_i is the i^{th} outcome of prospect x, and $\Pr(x_i)$ and $U(x_i)$ are the probability and the utility of outcome x_i, respectively. Expected Utility Theory (EUT) is also used to model the risk (i.e. variance) and return trade-off in portfolio theory (Markowitz 1952a). EUT risk preferences are therefore easily translated into pension funds' investment policies.

EUT provides a prescriptive behavioural model; it prescribes how rational people should behave. But people behave irrationally and other models – that incorporate irrationalities – frequently describe behaviour better.[4] The descriptive shortcomings of EUT were documented early on (Markowitz 1952b; Kahneman and Tversky 1979) and are largely explained by the following irrationalities: decision isolation, probability weighting, loss aversion,[5] and framing (Quiggin 2014).[6] These irrationalities lead

---

[4] See for a survey of descriptive "non-expected" utility models Quiggin (2014).

[5] The irrationality of loss aversion is debatable, but it is assumed here for the sake of brevity. See for a discussion on rationality and loss aversion Tversky and Kahneman (1991).

[6] Temporal discounting can also affect EUT, but this is outside the scope of this paper. Therefore, the model assumes exponential discounting with the same discount rate for everybody. This is formulated as $V_{i,t} = \sum_{t=0}^{\infty} \delta_i^t U_i(x_t)$, with δ_i the same for all subjects. By eliciting risk preferences for one specific point in time,

to unstable risk preferences across elicitation methods, since methods differ in their susceptibility to the various irrationalities (Holt and Laury 2014; Rabin and Thaler 2001). Irrational and unstable risk preferences should not prescribe pension risk-taking. Because these do not correspond to people's true risk preferences, using them for people's pension is not in their best interests. Therefore, models that incorporate irrationalities are not suitable for risk preference measurement. Instead, risk preference elicitation methods should measure underlying rational risk preferences. For this purpose, it is better to measure risk preferences with an EUT model while minimising irrationalities that may affect elicitation.[7] In this way, EUT's descriptive shortcomings in modelling irrationality do not undermine its prescriptive validity in modelling risk preferences (Hammond and Zank 2014).

2.1.2. An expected utility model with constant relative risk aversion (CRRA)

In EUT, the shape of an individual's utility function gives their risk preference. A risk averse individual has a concave utility function and a risk seeking individual has a convex utility function. The degree of concavity (convexity) gives the intensity of their risk aversion (seeking); a risk neutral individual has a linear utility function. Constant relative risk aversion (CRRA) is most common for risk preference elicitation (Holt and Laury 2014).[8] Therefore, I assume the following CRRA utility function (Pratt 1964; Arrow 1965):

$$U(x_i) = \frac{x_i^{1-r}}{1-r}$$

where $U(x_i)$ is the utility of outcome x_i and r is the risk preference parameter, a constant.[9] An individual with risk preference $r > 0$ is risk averse, one with $r = 0$ is risk neutral, and one $r < 0$ is risk seeking. Once an individual's risk preference is determined, it is possible to predict their preferences in any number of decisions, including pension fund investment policy.

Early critics already pointed out that (relative) risk aversion is not constant, since people can be risk averse in some decisions and risk seeking in others (Friedman and Savage 1948; Markowitz 1952b). In risk preference elicitation methods specifically, experiments show higher risk aversion for higher stakes, i.e. (Binswanger 1981; Holt and Laury 2002). CRRA is thus unlikely to hold for all decisions and all outcome levels. But it is a fair assumption for risk preference elicitation in the pension domain, at least when methods use realistic pension scenarios. The reason for this is that outcomes are at high levels and span relatively small ranges, e.g. between net monthly pension incomes of €1,080 and €1,470. Increasing relative risk aversion found in earlier studies, occurred only after outcomes were multiplied by ten or even one hundred (Binswanger 1981; Holt and Laury 2002). The above multiplications in the pension outcomes are much smaller and it is therefore fair to assume relative risk aversion is practically constant. Furthermore, realistic pension scenarios also insulate the link between risk preference elicitation and risk-taking decisions from the possibility of an incorrect CRRA assumption. Because of the similarity between the scenarios in which risk preference are elicited and those in which risk-taking decisions are made, the model is of relatively little influence. Thus, in case the model does not hold, it

---

different time preferences do not affect the elicited risk preferences. Alternatively, it is also possible to elicit and model time preferences. See for a survey of temporal discounting Doyle (2013) and for the simultaneous modelling of risk and time preferences Tanaka, Camerer, and Nguyen (2010).

[7] Nevertheless, there are attempts to include irrationalities in models that measure risk preferences. However, this is outside the scope of this paper. See for examples of the simultaneous modelling of risk preferences and probability weighting Harbaugh, Krause, and Vesterlund (2002) and Comeig, Holt, and Jaramillo-Gutiérrez (2015).

[8] See for a survey of sorts of risk aversion Meyer (2014).

[9] For $r = 1$, the utility function is: $U(x_i) = \ln x$.

is still reasonable to apply preferences to pension risk-taking if they are elicited with the same pension outcomes in the risk preference elicitation method.

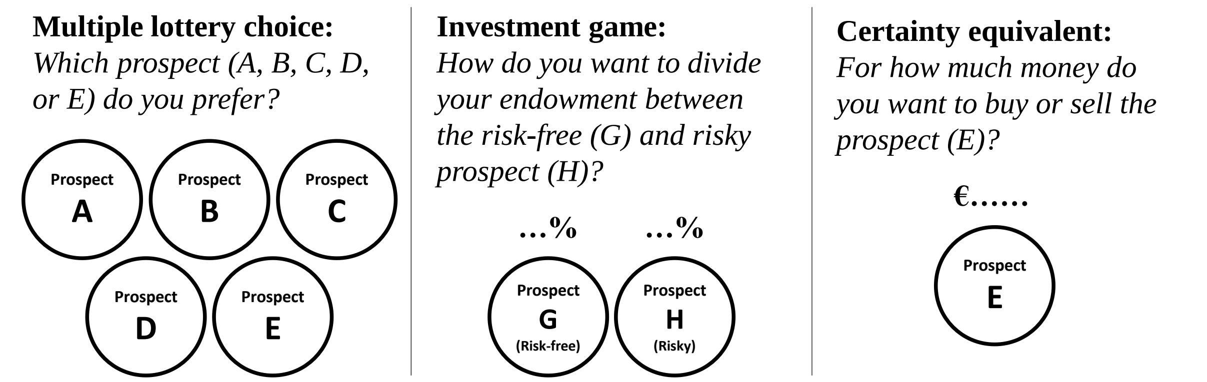
2.2. The five categories of risk preference elicitation methods found in the literature

Five categories encompass all the risk preference elicitation methods found in the literature (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). Four out of the five categories use EUT and the CRRA utility function to model individuals' risk preferences based on preferences between prospects with different degrees of risk. (Prospects are usually risky and consist of two payoffs with their own probabilities, e.g. a 50% chance of €100 and a 50% chance of €10.) Risk preference elicitation methods differ in how they present prospects and how prospects differ. The categories are based on the former (i.e. how they present prospects): single choices, choice lists, choice sequences, and dynamic methods; and within these categories, methods differ on the latter (i.e. how prospects differ, e.g. in payoffs, probabilities, or both). Questionnaires, which do not use EUT and CRRA, are the final category of risk preference elicitation methods (Charness, Gneezy, and Imas 2013).

2.2.1. Single choices: multiple lottery choice, investment game, and certainty equivalent

The single choice methods elicit risk preferences from the answer to one question. Based on their preferred prospect and its difference to the other prospects, it is possible to compute a risk preference range. There are three subcategories of single choice methods: multiple lottery choices, investments games, and certainty equivalents. Figure 2.1 below shows the typical setups for the different single choice methods. The multiple lottery choice asks individuals to choose the one lottery they prefer out a list of five, six, or more lotteries (Binswanger 1981; Eckel and Grossman 2002). The lotteries can range from a certain payoff on the one end to an all-or-nothing bet on the other end.

FIGURE 2.1 Typical setups for the different single choice methods



The investment game is comparable to the multiple lottery choice in that individuals again choose the one lottery they prefer.[10] But here they are asked to make that lottery themselves by dividing a risk-free and risky asset (Gneezy and Potter 1997; Charness and Gneezy 2012). Without the pre-determined discrete choices of the multiple lottery choice, individuals can now show their risk preference in a continuous manner with the investment game.

The certainty equivalent method asks individuals for a "selling price" or "buying price" for which they are willing to sell or buy a specific lottery (Harrison 1986; Kachelmeier and Shehata 1992;

[10] The investment game is comparable to the Distribution Builder (Goldstein, Johnson, and Sharpe 2008) and the Pension Builder (Dellaert et al. 2016), which elicit participants' risk preferences with a computer simulation. Individuals distribute 100 boxes (each representing a 1% probability) across different payoff levels. The distribution is limited by a budget, where a 1% probability at a higher payoff costs more than at a lower payoff.

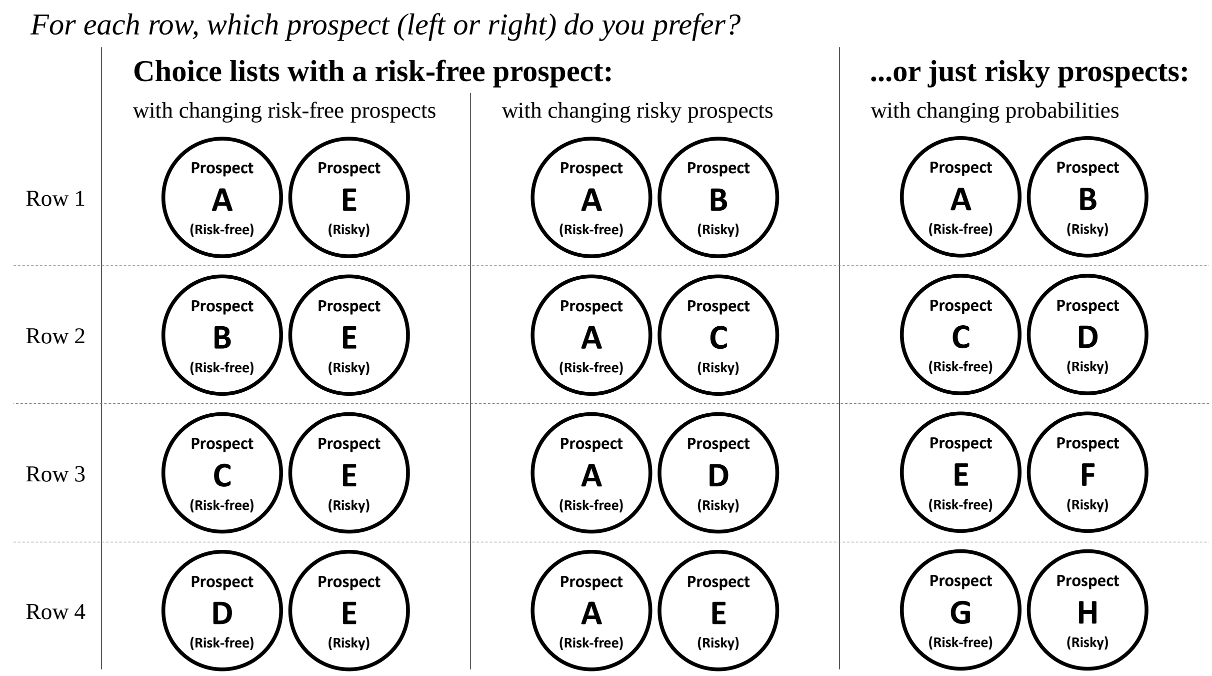
Isaac and James 2000). At the price, or "certainty equivalent", they are thus indifferent between the certain payoff and the lottery. Here, individuals can also show their risk preferences in a continuous manner.

2.2.2. Choice lists: multiple price lists with a risk-free prospect or risky prospects

Choice list methods, often called "multiple price lists", elicit risk preferences from the answers to a list of binary questions. Every binary question has a safe prospect on one side and a risky prospect on the other side. Moving downwards on the list, the riskier prospect becomes more attractive. This induces a switch from the safe to the risky prospect somewhere down the list and this "switching point" corresponds to a risk preference range. There are two subcategories of choice list methods, which either use a risk-free or risky prospect for the safe side. Figure 2.2 below shows the typical setups for the different choice list methods.

The first subcategory always has risk-free prospects on one side and risky prospects on the other side. There are three examples of such a choice list. One has risk-free prospects with changing in payoffs on one side and a constant risky prospect on the other side (Dohmen et al. 2005; 2011; Fehr-Duda, De Gennaro, and Schubert 2006; Abdellaoui, Driouchi, and l'Hardion 2011). This is essentially the same as the certainty equivalent method, but with pre-determined discrete choices. A second and third example have the same setup, but here the risky prospects change down the list (rather than the risk-free prospects, which are now constant). The risky prospects can either change with respect to their probabilities (Murningham, Roth, and Schoumaker 1988; Millner and Pratt 1991) or payoffs (Bruner 2009). The second subcategory has risky prospects on both as the safe and risky prospect, albeit safer and riskier lotteries, respectively. Payoffs of the lotteries are constant, but the probabilities change such that the riskier lottery becomes more attractive down the list (Holt and Laury 2002).

FIGURE 2.2 Typical setups for the different choice list methods



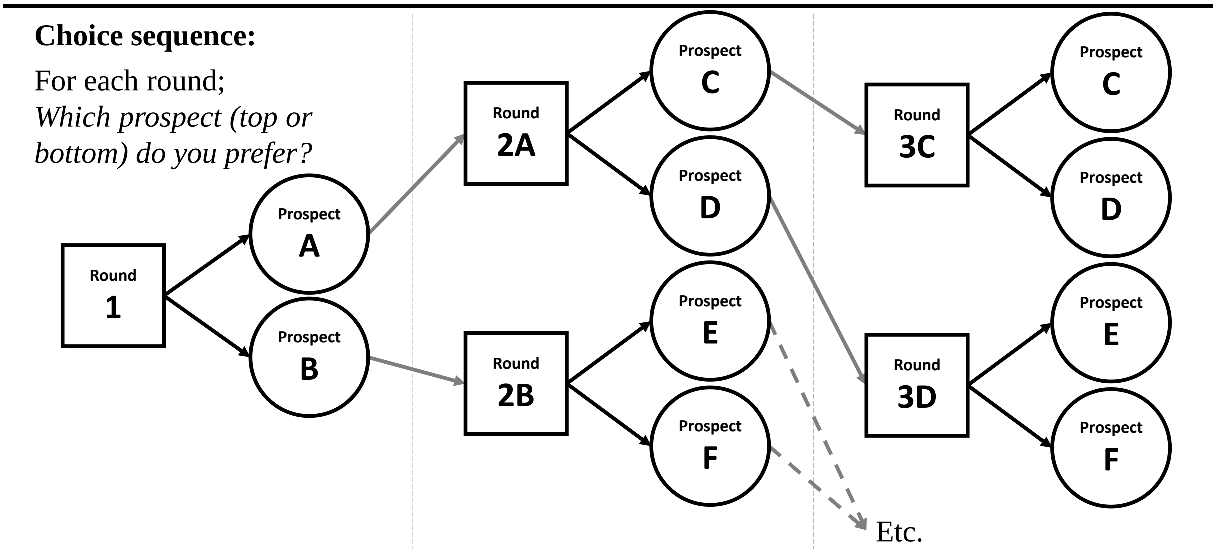
There are three choice lists in this figure (from left to right): a choice list with one risk-free prospect that changes down the list and one constant risky prospect; a choice list with one constant risk-free prospect and one risky prospect that changes down the list; and a choice list with two risky prospects that both change down the list in their probabilities.

2.2.3. Choice sequences: series of binary questions with certain payoffs or just lotteries

Choice sequence methods elicit risk preferences from the answers to a series of subsequent binary questions. It combines features of the previous two categories: a variety of prospects that are presented in a series of binary questions. After the first question, subsequent questions are based on previous answers. In this way, an individual's risk preference range is gradually narrowed down. Figure 2.3 below shows the typical setup for the choice sequence methods.

The literature usually frames choice sequences as decisions over lifetime income (Barsky et al. 1997; Arrondel and Calvo Pardo 2002; Kapteyn and Teppa 2011). That is, individuals are asked to choose between a current job with a certain wage and a new job with a wage that is higher on average, but risky. When they choose the certain (risky) wage, they are asked subsequent similar questions with higher (lower) risky wages as the alternative to their current job. There are also choice sequences with different binary questions, which include two risky outcomes (rather than one risky and one certain outcome).[11]

FIGURE 2.3 Typical setup for the choice sequence methods



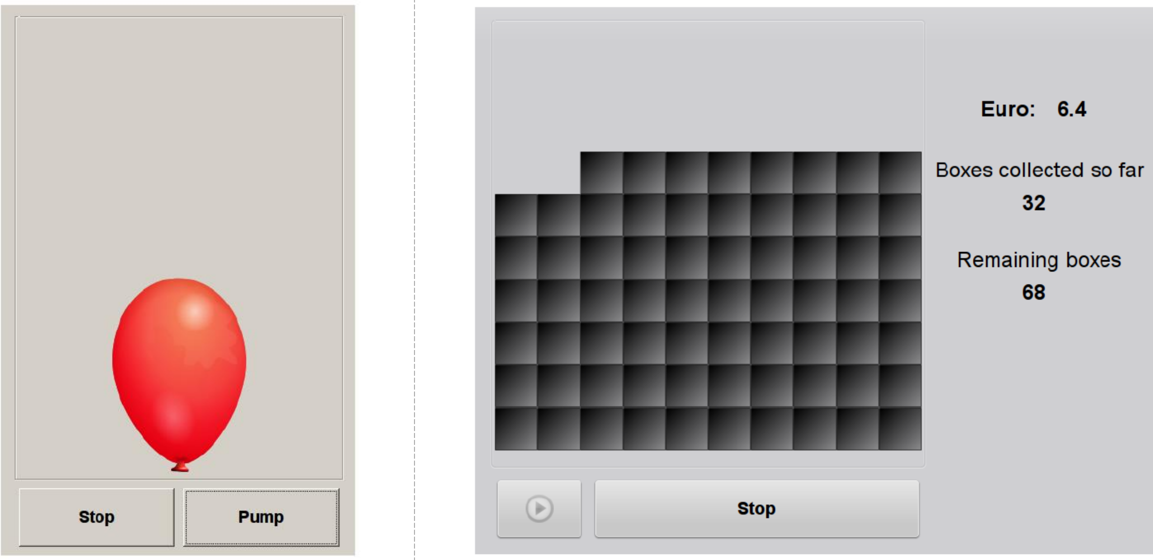
2.2.4. Dynamic methods: the balloon and bomb tasks and the Deal or No Deal show

Dynamic methods elicit risk preferences with a wide variety of dynamic simulations. All methods essentially let individuals choose between a certain payoff and continuing to play a game with a risky payoff. Based on the moment they stop playing the game, one can compute a risk preference range. The balloon analogue risk task (BART) lets individuals inflate a balloon, where the potential payoff increases with the size of the balloon, as does the chance of popping, which results in no payoff at all (Lejuez et al. 2002; 2003). The "bomb" risk elicitation task (BRET) lets individuals collect boxes, where the potential payoff increase with the collected boxes, as does the chance of collecting a bomb, which results in no payoff at all (Crosetto and Filippin 2013a). Figure 2.4 below shows the typical setup for the dynamic methods BART and BRET. The "Deal or No Deal" (DOND) television game show also lets individuals choose between a certain payoff and continuing to play a game with risky payoffs (Post et al. 2008).

[11] An application of the choice sequence method in this way is AEGON's Risicotool (in Dutch only), which elicit individuals' risk preferences through a series of binary choices between two pensions with incomes of different riskiness.

FIGURE 2.4 Typical setup for the dynamic methods BART and BRET

Balloon analogue risk task: Pump or stop?	**Bomb risk elicitation task:** Collect or stop?



The BART and BRET screenshots are taken from Crosetto and Filippin (2013b).

2.2.5. Questionnaires: general and domain-specific risk-taking attitude and behaviour

Questionnaires elicit risk preferences based on qualitative questions about risk preferences. Questions (either direct or through positions with respect to statements) are usually answered on a rating scale and can take many different forms. For example, the German socio-economic panel (SOEP) survey asks directly for one's risk preference (Dohmen et al. 2005). The domain-specific risk-taking (DOSPERT) scale, which includes a financial domain, asks how likely one is to conduct certain risky behaviours (Weber, Blais, and Betz 2002; Blais and Weber 2006).

3. Literature-based viability of risk preference elicitation methods for pensions

With five categories and even more subcategories, there is a wide variety of risk preference elicitation methods. I develop four viability criteria to determine which methods are viable for the pension domain. I apply these criteria to the risk preference elicitation methods in Subsection 3.2. Afterwards, Table 3.1 gives an overview of the results.

3.1.Four criteria for viable risk preference elicitation methods for pensions

I use four criteria to determine the most viable risk preference elicitation methods for the pension domain. They are based on both the general and pension domain-specific viability of methods. The four criteria concern the measurability, appropriateness, rationality, and feasibility of the risk preference elicitation methods. For each criterion, I summarise the criterion's general relevance and its importance to the financial and pension domain. Subsequently, I formulate a concrete rule for every criterion to determine whether a risk preference elicitation method meets it or not.

3.1.1. Criterion I: the elicitation method measures absolute risk preferences

For application in the pension domain, it is important that the risk preference elicitation methods measure absolute risk preferences. Elicitation methods can produce absolute and relative risk preferences. Relative risk preferences rank individuals. Absolute risk preferences quantify risk preference. When risk preferences are measured absolutely, they are comparable across different elicitation methods. In contrast, relative risk preferences are only comparable within the same elicitation method. Another benefit of absolute risk preferences is that they can provide preferences for other risky decisions as well. This makes it possible to use a subject's measurements in the elicitation method to prescribe, for instance, investment policies for their pension funds

Absolute risk preferences are based on EUT model with constant relative risk aversion (CRRA) throughout the literature (Holt and Laury 2014). Such measurement creates risk preferences that are easily transferable to other financial decisions. For example, modern portfolio theory can use CRRA parameters for the risk-return trade-offs in investing decisions (Markowitz 1952a). This measurement of risk preferences can thus be directly applied to pension funds' investment decisions. Therefore, for an elicitation methods to meet the measurability criterion, it must produce a CRRA parameter.

Measurability Criterion (I): the elicitation method produces a CRRA parameter.

To meet this requirement, the method must elicit preferences between quantified risky prospects in terms of payoffs and probabilities. This information is necessary to compute the required CRRA parameters.

3.1.2. Criterion II: the method elicits appropriate risk preferences for pensions

To measure risk preferences for pensions, elicitation must take place within realistic pension scenarios. The domain-dependency of risk preferences is well-documented in the literature (Holt and Laury 2014) Differences in risk preference occur between, among others, the social, health, and financial domains (Weber, Blais, and Betz 2002; Blais and Weber 2006). It is not necessarily problematic when risk preferences differ between domains that are far apart. An individual may very well be risk averse in the health domain and simultaneously be risk seeking in the social domain. But within the financial domain, which involves large-stake economic decisions, one expects stable risk preferences under the EUT. Unfortunately, this is not always the case: e.g. Van Rooij, Kool, and Prast (2007) find significant differences between self-assessed risk preferences between the financial and pension domain in a Dutch sample. Since this variation lacks adequate explanation, it is best to use risk preferences elicited in the appropriate domain (Charness, Gneezy, and Imas 2013; Holt and Laury 2014).[12]

The domain-dependency of risk preferences means it is important to elicit risk preferences that are appropriate to the pension domain, which is where they are applied. Given the sensitivity of risk preferences to the elicitation method, a statement referring to the pension domain is probably insufficient to achieve this aim. Subjects should feel like their decisions concern their real pension. Therefore, for an elicitation methods to meet the appropriateness criterion, it must use realistic pension scenarios.

Appropriateness Criterion (II): the elicitation method uses realistic pension scenarios.

To meet this requirement, the method must introduce a pension context and elicit risk preferences between different realistic pension incomes. In this way, elicited risk preferences are appropriate for the pension domain.

[12] Several studies have shown different factors affect the stability of risk preferences, for example, subject's visceral states (Loewenstein 2000), emotional evaluations (Loewenstein et al. 2001), and risk perception (Weber, Blais, and Betz 2002). To the extent that these factors depend on the domain, they are addressed by Criterion II, which requires pension risk preferences to be elicited in the pension domain. And to the extent that these factors depend on the method and its prospects, they are addressed by Criterion III, which requires risk preference elicitation method to use, for example, intermediate probabilities and neutral framing.

3.1.3. *Criterion III: the elicitation method tests underlying rational risk preferences*

The elicitation method should test underlying rational risk preferences to provide accurate preferences in other decisions, such as pensions.[13] Irrational effects in risk preference elicitation abound. Decision isolation leads people to assess gains and loss from a reference point rather than through asset integration (Kahneman and Tversky 1979). This causes implausible risk preferences under EUT, when they are transferred to other wealth levels (Rabin 2000).[14] Probability weighting leads people to overweigh small probabilities and underweigh large probabilities (Kahneman and Tversky 1979; Prelec 1998). Loss aversion leads people to overweigh losses relative to gains (Tversky and Kahneman 1992). Framing of elicitation methods affects risk preferences due to several irrational biases and heuristics (Baltussen, Post, & Van den Assem 2008). For instance, choices in a health context are reversed when its prospects are rephrased from "saved" to "lost lives" (Tversky and Kahneman 1981). All these irrationalities affect risk preference elicitation differently across methods and subjects. This makes it important to avoid them and test for the underlying rational risk preferences instead.

The ideal of testing underlying rational risk preference is approximated by avoiding or minimising the above irrational effects. There are several ways to do this for the risk preference elicitation methods in the pension context. Decision isolation is minimised by presenting prospects with integrated outcomes (Holt and Laury 2014). In practice, this means subjects see their total pension income rather than only the part of the income for which their risk preference is elicited (Alserda et al. 2016). Probability weighting is smallest for the intermediate probabilities (where there is only a slight under weighing). Therefore, risk preference elicitation methods can avoid most probability weighting by using intermediate probabilities (Holt and Laury 2014).[15] This is possible with a variety of methods that use prospects with two equally probable payoffs (that is, with 50% chance) or choice lists with expected switching points in the five middle rows (that use probabilities between 30% and 70%). Loss aversion's effect on risk preference elicitation depends on the framing of the method and its prospects (Holt and Laury 2014).[16] Gain or loss frames are avoided by integrated outcomes without reference points. Additionally, neutral framing of the method and its prospects also avoids most other framing effects.

The above measures do not account for all irrational effects, but they do address the most troublesome and well-documented ones (Holt and Laury 2014). Therefore, these are leading in the assessment of this criterion. Other irrational effects are only included in the assessment when the literature indicates they may be of interest to a specific method. For an elicitation methods to meet the rationality criterion, it must thus avoid or minimise irrational effects on risk preference elicitation.

Rationality Criterion (III): the method avoids or minimises irrational effects on elicitation.

To meet this requirement, the method must avoid or minimise irrational effects of decision isolation, probability weighting, loss aversion, framing, and other irrational effects. This means it elicits risk preference with integrated outcomes, intermediate probabilities, no reference points, neutral framing, and possibly other solutions.

---

[13] See for a discussion of the (underlying) rationality of risk preferences in the pension domain Subsection 2.1.

[14] Some experimental evidence indicates people neither fully integrate their assets or isolate their decisions, but engage in "partial asset integration" (Cox and Sadiraj 2006; Anderson et al. 2011). Such a model can give plausible risk preferences and it mitigates much of Rabin (2000)'s "calibration" critique (Holt and Laury 2014).

[15] Alternatively, see for risk preference elicitation methods that incorporates probability weighting Harbaugh, Krause, and Vesterlund (2002) or Comeig, Holt, and Jaramillo-Gutiérrez (2013).

[16] Experimental evidence on loss aversion's effect in risk preference elicitation methods is mixed: some find loss aversion (Kahneman and Tversky 1979; Tversky and Kahneman 1992) and others do not or only partially (Harbaugh, Krause, and Vesterlund 2002, 2010; Bosch-Domenech and Silvestre 2006, 2010; Laury and Holt 2008). See for an overview Holt and Laury (2014).

3.1.4. Criterion IV: the elicitation method is feasible for broad application

Elicitation methods are constrained by the subjects' cognitive ability and their willingness to partake in it. This limits methods' complexity and duration. An inability to comprehend the elicitation method leads to inaccurate or inconsistent risk preferences.[17] The cognitive ability of subjects is an important aspect of method design (Charness, Gneezy, and Imas 2013). Several studies find that elicitation methods differ significantly in their complexity and this creates a trade-off between accuracy and noise (Dave et al. 2010; Crosetto and Filippin 2016). The willingness of subjects is also important. They should comprehend and truthfully complete the elicitation method to accurately determine their underlying rational risk preference. Both cognitive ability and willingness are considerably constrained in the pension context. Subjects come from a variety of socio-economic statuses, since everyone must be able to take the risk preference elicitation method. This may be problematic. It includes a broad spectrum of educational backgrounds and intellectual capacities, which limits methods' complexity. Neither may all subjects grasp the importance of the elicitation method and hence be willing to (seriously) partake. For instance, a Dutch sample indicates they are unwilling to spend time on pension decisions (Van Rooij, Kool, and Prast 2007). This also limits methods' duration.

Feasible risk preference elicitation methods for the pension context should be accessible to people from a broad range of cognitive abilities and willingness to partake. This constrains its complexity and duration. Therefore, for an elicitation methods to meet the feasibility criterion, it must not require exceptional cognitive ability and take only a few minutes to complete.

Feasibility Criterion (IV): the elicitation method must not require exceptional cognitive ability and take only a few minutes to complete.

To meet the first part of this requirement, the methods' ease (or lack of complexity) must follow from the literature. This can be based on explicit discussion of easiness or complexity of methods, comparisons between them, or validation of them in relevant subject pools. And to meet the second part of the requirement, the method must not take more than a few minutes to complete. This is easily deductible from the elicitation method, which, of course, can be supplemented with practical trial data.

3.2. Selection of the most viable methods for pensions based on the four criteria

I use the four criteria to select the most viable risk preference elicitation methods for pensions. Methods' performance on each criterion is based on evidence from the literature. This determines how well each method scores on each criterion. Predominantly positive evidence means it meets a criterion; predominantly negative evidence means it does not. I also allow two alternative scores for mixed evidence and for when evidence is not available. Methods that do not meet one or more of the four criteria, I deem unviable. The remaining methods – those that meet all four criteria or those that have no or mixed evidence – I deem viable. The latter thus receive the benefit of the doubt. This approach is a result of the ambiguity in the literature on risk preference elicitation methods. No method receives unanimous praise and most positive comparisons include numerous caveats. I therefore only dismiss those methods that are definitively unviable for application in the pension domain, i.e. those that perform negatively on one or more of my criteria. The resulting methods are the most viable for usage with pensions. Below, I discuss the performance of the risk preference elicitation methods per criterion. Table 3.1 subsequently provides on overview of the scores as well as the resulting viability of each method.

[17] See for a discussion of the role of inconsistencies in risk preference elicitation methods Subsection 5.1.3.

3.2.1. *Criterion I: measurable through a CRRA parameter*

The elicitation method must measure risk preferences with a CRRA parameter. They meet this criterion by eliciting risk preferences between quantified risky prospects in terms of payoffs and probabilities. Questionnaires do not meet this criterion because they use qualitative questions (Charness, Gneezy, and Imas 2013). The other elicitation methods can produce CRRA parameters because they use quantified risky prospects in terms of payoffs and probabilities (Holt and Laury 2014; Crosetto and Filippin 2016). However, the initial studies of the investment game, BART, and DOND methods do not measure their risk preferences through a CRRA parameter (Charness, Gneezy, and Imas 2013; Lejuez et al. 2002; 2003; Post et al. 2008). Thus, all methods, except for the questionnaires, meet the measurability criterion.

3.2.2. *Criterion II: appropriate because of realistic pension scenarios*

The method must elicit the appropriate risk preferences through realistic pension scenarios. They meet this criterion by introducing a pension context and eliciting risk preferences between realistic pension incomes. Questionnaires' format is relatively free and can therefore be adapted to different domains, including pensions (Dohmen et al. 2005; Weber, Blais, and Betz 2002; Van Rooij, Kool, and Prast 2007). The single choice, choice list, and choice sequence methods can also be adapted to the pension domain. Payoffs can be scaled up (Binswanger 1981; Holt and Laury 2002), probabilities and payoffs altered, and introductory texts can put the questions into a certain context, for instance, lifetime income (Barsky et al. 1997). The dynamic methods are not adaptable because of specific game-like setups in both its format as well as payoffs and probabilities (Charness, Gneezy, and Imas 2013; Buelow and Suhr 2009). Because of this, only the dynamic methods do not meet the appropriateness criterion.

3.2.3. *Criterion III: rational by avoiding or minimising irrational effects*

The method must elicit underlying rational risk preferences by avoiding or minimising irrational effects, such as decision isolation, probability weighting, loss aversion, and framing. This means they elicit risk preference with integrated outcomes, intermediate probabilities, no reference points, and neutral framing. If it cannot do this or it faces other unavoidable irrational effects, following from the literature, it does not meet the criterion. Note that all methods are influenced by some sort of irrationality and hence none fully meets the rationality criterion. Since mixed evidence suffices, however, only those methods with considerable and unavoidable irrational effects do not meet this criterion. I discuss the viability of the methods per category.

Questionnaires are qualitative and hence prone to irrational framing effects that can affect answers in numerous ways (Tversky and Kahneman 1981). They also do not use the quantitative EUT that may measure an underlying rational risk preference in the form of a CRRA parameter. As such, questionnaires do not meet the rationality criterion.

The single choice, choice list, and choice sequence methods can integrate their outcomes and refrain from reference points. Besides, the single choice methods – as well as most choice list and choice sequence methods – can avoid probability weighting by using gambles with two equally likely outcomes (Holt and Laury 2014). Only framing poses a problem for the single choice methods. Multiple lottery choices and investment games may have a bias towards the central options, given that the mean and median answers lay here (Binswanger 1981; Eckel and Grossman 2002; Charness and Gneezy 2012). This is unavoidable because any order creates a framing effect and I do not consider it a significant irrationality. The investment game may, however, also be prone to naïve diversification, where subjects divide their equally between the number of alternatives they are offered (Benartzi and Thaler 2001). The certainty equivalent method is prone to framing via the "response mode effect" – individuals who are asked for a "selling price" give higher certainty equivalents than those who are asked for a "buying price” – while neutral framing makes the method too complex (Kachelmeier and Shehata 1992; Isaac and James 2000; Holt and Smith 2009) Therefore, the investment game and certainty equivalent methods do not meet the rationality criterion, while the multiple lottery choice does.

Choice lists with a certain payoff on one side and a gamble on the other side are also prone to several irrationalities. Most importantly, the “certainty effect” leads individuals to overweigh the certain payoff, resulting in too risk-averse preferences (Kahneman and Tversky 1979; Holt and Laury 2002, 2014). Simple payoffs can lead people to “nice” focal points, e.g. €10 or €100 (Dohment et al. 2005) and may also induce risk-neutral answers because it simplifies expected-value calculations (Sprenger 2010). Choice lists with gambles on both sides avoid the above issues. However, they use changing probabilities, which makes them vulnerable to probability weighting. But its effect is minimised in choice lists where the common switching points lay around the middle rows, which have intermediate probabilities (Holt and Laury 2014). Therefore, the choice lists with a certain payoff do not meet the rationality criterion.

Choice sequences face many of the same issues as the other choice methods, although these are less well researched. It can avoid these with integrated outcomes and intermediate probabilities, without reference points and with neutral framing. In its “lifetime income” format, it is especially vulnerable to the certainty effect or “status quo bias,” since individuals may be unwilling to risk their current job and wage level (Barsky et al. 1997). In practice, this problem can be solved by using two gambles and avoiding a reference point such as the status quo.[18] The choice sequences thus meet the rationality criterion.

The dynamic methods’ vulnerability to irrationalities is also little researched. But there are numerous irrational effects that likely affect the BART, BRET, and DOND method. For example, they employ extreme probabilities, which makes them prone to probability weighting. Furthermore, certain payoffs can cause the “certainty effect” and reference points can trigger loss aversion. And their complex format probably creates framing effects. Because of the lack of evidence, however, I do not report the dynamic methods’ performance on the rationality criterion.

3.2.4. Criterion IV: feasible for all socio-economic backgrounds in a few minutes

Partaking in the risk preference elicitation method must be feasible for subjects of all socio-economic backgrounds and take no longer than a few minutes. With respect to the duration, completion of all methods seems feasible within a few minutes. The methods thus meet the feasibility criterion if their ease follows from the literature; when the predominant literature finds the method complex, it does not meet the feasibility criterion. Questionnaires generally pose simple and straightforward questions (Charness, Gneezy, and Imas 2013) and their length is easily adjusted to an appropriate duration. Single choices vary in difficulty. Multiple lottery choices and investment games are simple (Charness, Gneezy, and Imas 2013; Crosetto and Filippin 2016). Certainty equivalents are hard to understand, especially with neutral terminology (Holt and Laury 2014), although this seems not to be a problem for higher-educated individuals (Kachelmeier and Shehata 1992). Multiple price lists also vary in difficulty. Lists with a certain payoff on one side are easier (Holt and Laury 2014), whereas lists with gambles on both sides are more complex (Charness, Gneezy, and Imas 2013; Crosetto and Filippin 2013b). They require more mathematical abilities (Dave et al. 2010), but are widely used in the literature (Charness, Gneezy, and Imas 2013) and do not seem problematic for a sample of French farmers (Reynaud and Couture 2012). Choice sequences seem less complex than both choice list and single choice methods, although there are no comparisons. It is also used with “a large cross section of households” in the US Health and Retirement Study (Barsky et al. 1997). The dynamic methods can create a familiar and easy-to-grasp risk-taking context (Charness, Gneezy, and Imas 2013), but they can also be relatively complex, in case

[18] This is similar to the choice sequence application of AEGON’s Risicotool (Dutch). See also footnote 10.

of the BART for instance (Crosetto and Filippin 2013b). In sum, most methods meet the feasibility criterion or have mixed evidence on it (see also Table 3.1 below). Only the certainty equivalent method does not meet the criterion and there is no available evidence for the DOND methods.

3.2.5. Viability of methods based on aggregate performance over the four criteria

The performance of the risk preference elicitation methods on the four criteria is depicted in Table 3.1. Based on the literature, I conclude whether there is positive evidence (+), mixed evidence (=), negative evidence (-), or no evidence available (n/a) for each method on the criteria. In case of any no negative evidence (-), I deem the method unviable. The remaining methods – those with either positive (+), mixed (=), or no available evidence (n/a) – I deem viable. My selection methodology produces three viable methods: the multiple lottery choices, the choice lists with two gambles, and the choice sequences.

Table 3.1 Performance of the risk preference elicitation methods on the pension context criteria

Category	Subcategory	Criterion				Viable
		Measurability	Appropriateness	Rationality	Feasibility	
Questionnaire		-	+	-	+	No
Single choice						
	Multiple lottery choice	+	+	=	+	Yes
	Investment game	+	+	-	+	No
	Certainty equivalent	+	+	-	-	No
Choice list						
	With a certain payoff	+	+	-	+	No
	With two gambles	+	+	=	=	Yes
Choice sequence						
	With a certain payoff	+	+	-	+	No
	With two gambles	+	+	=	+	Yes
Dynamic method						
	BART	+	-	n/a	=	No
	BRET	+	-	n/a	+	No
	DOND	+	-	n/a	n/a	No

Performance on the criteria is based on the literature and indicated as: criteria met (+), mixed evidence (=), criteria not meet (-), or evidence not available (n/a). Viability is based on the elimination of the unviable methods, which I define as those that do not meet one or more of the four criteria.

4. Experiment-based comparison of risk preference elicitation methods for pensions

The four literature-based criteria produce three viable risk preference elicitation methods for the pension context. Hereafter, I refer to these methods with abbreviations. The multiple lottery choice and choice list with two gambles become the EG and HL methods, respectively, based on the authors that introduced or popularised these methods, i.e. Eckel and Grossman (2002) and Holt and Laury (2002). The choice sequence is abbreviated as the CS method. Based on an experiment, I compare these methods to determine which performs best in this domain. The literature does not provide any such comparisons yet, i.e. does not compare all the three methods simultaneously or in similar setups (e.g. presentation, number of categories, and CRRA ranges). Neither are there many experiments, let alone comparisons, of these risk preference elicitation methods in a pension context. I fill this gap with an experiment that compares the EG, HL, and CS methods simultaneously, with similar setups, and in a pension context. For this purpose, I first adapt these risk preference elicitation methods to the pension domain with respect to presentation, risk preference categories, CRRA ranges, and pension scenarios.

4.1. Adaptations of the risk preference elicitation methods for the pension domain

The three risk preference elicitation methods follow the literature standards as much as possible. The EG method is adapted from Eckel and Grossman (2002), the HL method is adapted from Holt and Laury (2002), and the CS method is adapted from several papers and practical applications, e.g. Barsky et al. (1997) and Kapteyn and Teppa (2011). (Tables A.4, A.5, and A.6 and Figure A.1 in Appendix C show the original design of respectively the EG, HL, and CS methods, as well as the latter's decision tree.) I make four adaptations to adapt the methods for risk preference elicitation in the pension domain. To start with, I present the methods in a pension context with illustration to improve comprehensibility. Subsequently, I equalise their risk preference categories at ten. Thirdly, I set the methods' prospects such that their risk preference categories correspond to the same CRRA ranges. Finally, I create pension scenarios that make for realistic decisions.

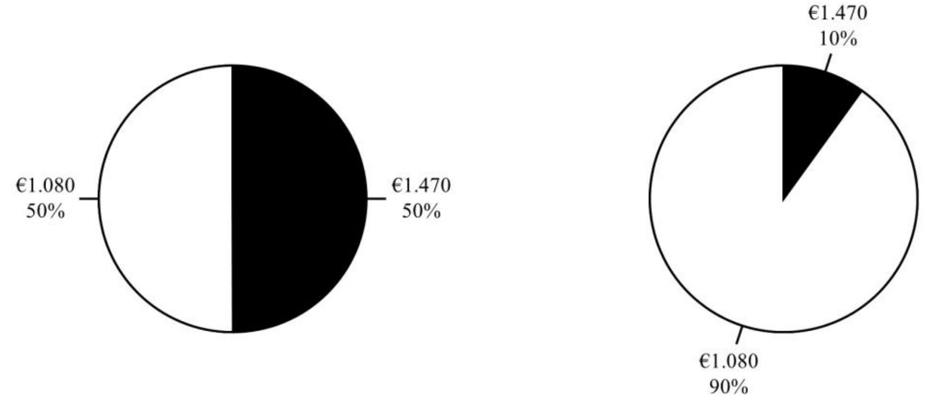
4.1.1. Adaptation I of the risk preference elicitation methods: presentation

The three methods are presented in such a way that they are easily linked to the pension context as well as understandable. It is important to evoke the pension context because of the context-dependency of risk preferences (see also Criterion II in Subsection 3.1.2.). The introduction advises subjects to imagine that their pension fund is asking them these questions and that it is using their answers to decide how to invest their pension savings. It explains they will be asked to choose between different pensions. And continues with an elaborate explanation of these pensions and their illustration. Stating, these "pension" are gambles that represent different net monthly pension incomes (i.e. including the public pension and after taxes) subjects can expect at the age of 75.[19] Every pension consists of two scenarios with different pension incomes that sometimes change and can also have different probabilities. There are two example pensions with an explanation of the corresponding incomes and their probabilities. The high and low incomes are related to a positive and negative scenario. The probabilities of these scenarios are translated to a natural frequency (e.g. for a scenario with 50%, this scenario occurs 5 out of 10 times) to aid subjects' understanding of the probabilities (Hoffrage et al. 2000; Gigerenzer and Edwards 2003). Presenting subjects with the interpretation of incomes as scenarios and probabilities as natural in the two examples, should improve their comprehension of pension incomes and probabilities in the subsequent risk preference elicitation methods.

To further improve understandability, pensions are presented with illustrations. Every pension consists of possible payoffs and corresponding probabilities. Illustrating these two variables simultaneously is more complex than illustrating none. Therefore, only one variable is illustrated, i.e. the probabilities. People find probabilities hard to understand (Hoffrage et al. 2000; Gigerenzer and Edwards 2003; Gigerenzer et al. 2005) and the benefits of illustrating probabilities are well documented (Edwards, Elwyn, and Mulley 2003; Hess, Visschers, and Siegrist 2011). Besides, it is likely they understand net monthly (pension) income levels, especially in relation to their current net monthly income. In illustrating probabilities, pie charts are one of the most effective ways to do so (Lipkus and Hollands 1999; Lipkus 2007). For these reasons, I present subjects with the pensions through pie charts that indicate the probabilities of the pension incomes. The pension incomes themselves are presented as euro amounts and linked to the relevant pieces of pie. Gambles are frequently presented through pie charts in the literature (Von Gaudecker, Van Soest, and Wengstrom 2011), including for risk preference elicitation in the EG method (Deck et al. 2010) and HL method (Reynaud and Couture 2012). Figure 4.1 shows two examples of the pie charts that represent pensions.

[19] This specific point in time ensures decisions are based on risk preferences rather than intertemporal preferences (see also footnote 6). The age is chosen to provide ample sample size in the *AFM Consumentenpanel*.

FIGURE 4.1 Two examples of pie charts representing pensions



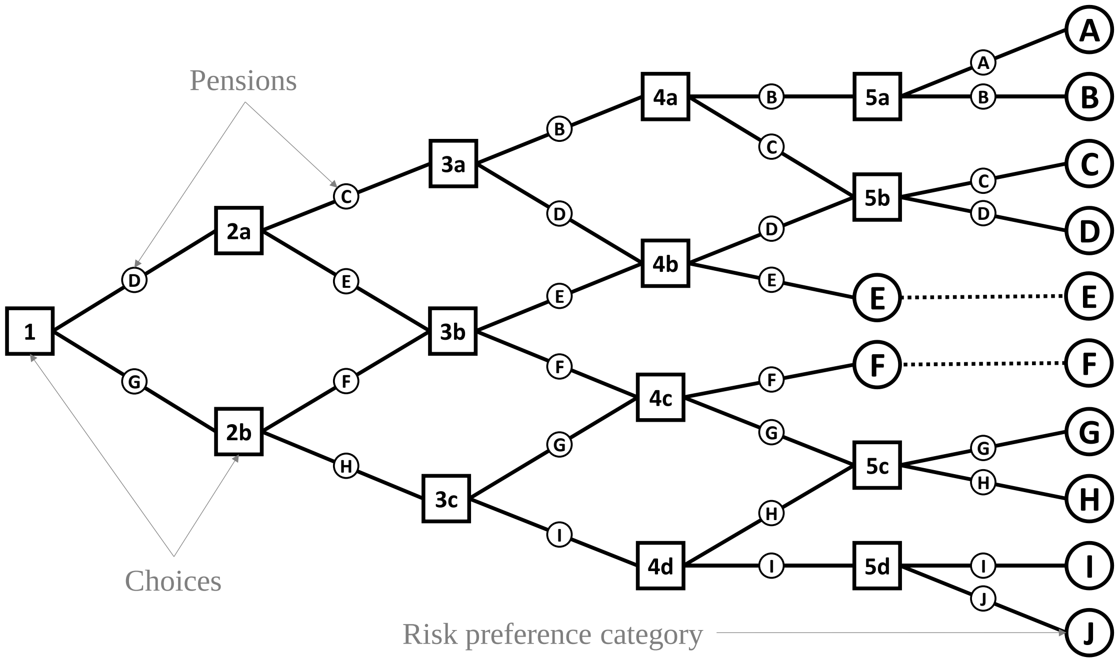
The pension on the left side has a 50% chance of an income of €1,080 and a 50% chance of an income of €1,470; the pension on the right side has a 90% chance of an income of €1,080 and a 10% chance of an income of €1,470. N.B. the euro amounts in the above examples use different punctuation, following Dutch standards (i.e. €1.080 instead of the English €1,080).

Before subjects are presented with the three risk preference elicitation methods, they are asked a test question. Its aim is to familiarise subjects with the pie chart presentation of pension incomes. After they answer, subjects see whether they are correct and receive an explanation of the test question. Before every one of the three risk preference elicitation methods, subjects also get another short introduction. It informs them there are no correct or incorrect answers to the questions and they should follow their own preferences, it also explains the particularities of the specific method and it reminds them that the pensions represent net monthly incomes at age 75, including the public pension.

4.1.2. Adaptation II of the risk preference elicitation methods: risk preference categories

The second adaptation sets all three elicitation methods' risk preference categories to ten. This is the standard number for the HL method, but higher than the common numbers for the EG and CS methods. Ten categories allow for both a more detailed risk preference elicitation as well as analysis. Reynaud and Couture (2012), for instance, compare the EG and HL method and equalise the number of categories for both at nine. For the EG method, the number of gambles to choose from (originally set at five) is thus simply increased to ten. The literature includes many increases in the EG method's categories, for instance, to six gambles (Binswanger 1981) or even a continuous choice in the investment game (Gneezy and Potters 1997). For the CS method, the number of consecutive questions is set at four to five to achieve the desired ten categories after the final questions. Although Barsky et al. (1997) and Arrondel and Calvo Pardo (2002) use two questions, but others use three (Kapteyn and Teppa 2011) or five (AEGON Risicotool), which also result in more risk preference categories. Whether one receives four or five questions depends on the extremeness of one's risk preference: more extreme preferences require more questions to determine the correct category since the first question starts at the middle. This approach is also taken by Kapteyn and Teppa (2011), who use two to three questions. Figure 4.2 shows the decision tree of the pension-adopted CS method.

FIGURE 4.2 Decision tree of the pension-adopted CS method



4.1.3. Adaptation III of the risk preference elicitation methods: CRRA ranges

The ten risk preferences categories also need the same CRRA ranges to create better comparable risk preferences based on the EUT-model. In the literature, the CRRA ranges differ widely across methods and experiments. For example, in the EG method, there are six categories ranging from below 0 to over 7.5 (Binswanger 1981), five categories ranging from below $r = 0.2$ to over $r = 2$ (Eckel and Grossman 2002), and even nine categories ranging from below $r = -0.95$ to over $r = 1.37$ (Reynaud and Couture 2012). The latter number of categories and range is also used for the HL method, as are, for example, nine categories ranging from below $r = -1.85$ to over $r = 1.37$ (Holt and Laury 2002) or ten categories ranging from below $r = -4.82$ to over $r = 4.46$ (Alserda et al. 2016). For the CS method, one can see a range from below $r = 1$ to over $r = 3.76$ with four categories (Arrondel and Calvo Pardo 2002) as well as wider ranges with more categories (Van Rooij, Kool, and Prast 2007; Kapteyn and Teppa 2011). In general, the CRRA range is larger when there are more categories. I follow the CRRA ranges of Alserda et al. (2016) for this experiment, which also uses ten risk preference categories. More importantly, it provides neutral and broad CRRA ranges for its categories, i.e. they are equally divided between risk averse and risk seeking and their span on either side ensures comprehensive elicitation. The CRRA ranges of a method influence the elicited risk preferences through framing, e.g. when a range is more risk averse it also leads to more risk averse answers and vice versa (Anderson et al. 2006). Furthermore, the extreme ends of the range are generally avoided in answers, irrespective of their value (List 2007; Bardsley 2008; Crosetto and Filippin 2014). This and little further information on actual risk preferences in the pension context, makes the neutral and broad range desirable. Finally, the Alserda et al. (2016) CRRA ranges are used in the context of Dutch pensions. They find a relatively high mean CRRA coefficient ($r = 1.92$)[20] and attribute this to higher risk aversion in the pension context. Given this high level of risk aversion and risk preference heterogeneity (i.e. they find a standard deviation of 2.15), broader ranges are desirable to accurately elicited subjects' risk preference.

[20] Their mean CRRA coefficient is based on the HL method combined with two other risk preference elicitation methods: a pension-related self-description question and a portfolio choice method. The three methods are combined in a composite score using factor analysis (with factor loadings of 0.87, 0.82, and 0.50 for the HL method and the other two methods, respectively). See for more details Alserda et al. (2016).

4.1.4. Adaptation IV the risk preference elicitation methods: custom pension scenarios

One of the most important adaptations is to present subjects with gambles that have realistic pension scenarios (see also Criterion II in Subsection 3.1.2.).[21] The gambles contain pension incomes that are based on subjects' income and simultaneously represent the desirable CRRA ranges. The pension incomes are customised by multiplying the reference income by a certain factor for each method and gamble. Because of CRRA, the multiplication factors are the same for all income levels, i.e. risk aversion is constant relative to income. After multiplying the current income by the factor, pension incomes are rounded to tens. This strikes a balance between unnecessary complex and overly simple payoffs, which may lead to "nice" focal points or expected-value calculations (Dohmen et al. 2005; Sprenger 2010). Public pensions and taxes are included in the payoffs "to keep as close to the actual situation as possible." (Alserda et al. 2016) To end up with net monthly pension incomes, the reference incomes should be net monthly income as well. Participants in the *AFM Consumentenpanel* provide their incomes when they join the panel. These data consist of five ranges of gross yearly income. The centre of these ranges is taken, divided by twelve, and subtracted by taxes to calculate the net monthly income of subjects.[22]

For the gambles of the HL method, I use the multiplication factors of Alserda et al. (2016), which is convenient because we use the same CRRA ranges (see Subsection 4.2.4). The multiplication factors are applied to current net monthly income and calculate the projected net monthly pension incomes. They are 0.6 and 0.7 for the low and high payoffs in the safe gamble and 0.4 and 0.9 for the low and high payoffs in the risky gamble. For example, given a current net monthly income of €1,960, the safe gamble pays €1,170 or €1,370 and the risky gamble pays €780 or €1,760. Table 4.1 shows the pension-adapted HL method design for the €1,960 income level. (Tables A.1, A.2, and A.3 in Appendix B show the multiplication factors and the pension incomes for all methods and income categories.)

TABLE 4.1 Overview of the pension-adapted HL method design

	Option A				Option B				EV(A)-EV(B)	CRRA range
	Prob.	Pension	Prob.	Pension	Prob.	Pension	Prob.	Pension		
1	10%	€1,370	90%	€1,170	10%	€1,760	90%	€780	€312	$-\infty < r < -4.82$
2	20%	€1,370	80%	€1,170	20%	€1,760	80%	€780	€234	$-4.82 < r < -3.00$
3	30%	€1,370	70%	€1,170	30%	€1,760	70%	€780	€156	$-3.00 < r < -1.82$
4	40%	€1,370	60%	€1,170	40%	€1,760	60%	€780	€78	$-1.82 < r < -0.86$
5	50%	€1,370	50%	€1,170	50%	€1,760	50%	€780	€0	$-0.86 < r < 0.00$
6	60%	€1,370	40%	€1,170	60%	€1,760	40%	€780	-€78	$0.00 < r < 0.85$
7	70%	€1,370	30%	€1,170	70%	€1,760	30%	€780	-€156	$0.85 < r < 1.76$
8	80%	€1,370	20%	€1,170	80%	€1,760	20%	€780	-€234	$1.76 < r < 2.85$
9	90%	€1,370	10%	€1,170	90%	€1,760	10%	€780	-€312	$2.85 < r < 4.46$
10	100%	€1,370	0%	€1,170	100%	€1,760	0%	€780	-€390	$4.46 < r < \infty$

Whereas the HL method uses four payoffs with different probabilities in every gamble, the EG and SC method use two different payoffs for every one of the ten gambles. The EG method commonly ranges from a certain payoff in the safest gamble to an all-or-nothing bet in the riskiest gamble (Binswanger 1981; Eckel and Grossman 2002). A certain payoff for the safest gamble is a realistic choice for the pension context, since pensioners can choose a risk-free annuity for their monthly pension income in the Netherlands. I use a 0.65 multiplication factor for the certain payoff, which is comparable to the HL method (i.e. equal to the mean of its multiplication factors). In contrast, an all-or-nothing bet for the

[21] Ideally these pension scenarios are based on pension funds' actual risk and return forecasts. Such computations, however, are outside the scope of this paper and more rudimentary calculations are used instead.

[22] I use the *Loonwijzer*-website, an affiliate of the Wage Indicator Foundation, to transform gross incomes to net incomes (https://www.loonwijzer.nl/home/salaris/brutonetto).

riskiest gamble is an unrealistic choice for the pension context, since pensioners are guaranteed a risk-free public pension that functions as a minimum. Therefore, to determine the multiplication factors for the remaining nine gambles, I incrementally subtract 0.03 from the multiplication factor of the low payoff and subsequently use the appropriate CRRA range to calculate the multiplication factor for the high payoff. An incremental subtraction of 0.03, together with the given CRRA ranges, creates multiplication factors of 0.38 and 0.887 for the riskiest gamble, which are comparable to those of the risky gamble in the HL method. Furthermore, the 0.38-low multiplication factor ensures the lowest possible net monthly pension income is well above the minimum of the public pension scheme. For example, take a current net monthly income of €1,960. The first gamble pays a certain payoff of €1,270, based on a 0.65 multiplication factor. The second gamble pays either €1,350, based on a 0.62 multiplication factor, or €1,210, based on indifference between the first and second gamble with a CRRA coefficient of 4.46. (Recall that in the EG method, each gamble has a 50% chance of the low and high payoff.) Table 4.2 shows the pension-adapted EG method design for the €1,960 income level.

TABLE 4.2 Overview of the pension-adapted EG method design

Option	Probability	Pension	Prob.	Pension	EV	CRRA range
A	50%	€1,270	50%	€1,270	€1,270	$4.46 < r < \infty$
B	50%	€1,350	50%	€1,210	€1,280	$2.85 < r < 4.46$
C	50%	€1,440	50%	€1,160	€1,300	$1.76 < r < 2.85$
D	50%	€1,540	50%	€1,100	€1,320	$0.85 < r < 1.76$
E	50%	€1,620	50%	€1,040	€1,330	$0.00 < r < 0.85$
F	50%	€1,680	50%	€980	€1,330	$-0.86 < r < 0.00$
G	50%	€1,710	50%	€920	€1,315	$-1.82 < r < -0.86$
H	50%	€1,730	50%	€860	€1,295	$-3.00 < r < -1.82$
I	50%	€1,740	50%	€800	€1,270	$-4.82 < r < -3.00$
J	50%	€1,740	50%	€740	€1,240	$-\infty < r < -4.82$

The CS method uses the same gambles as the EG method (see also the decision tree in Figure 4.3). For example, given a current net monthly income of €1,960, its first question poses the choice between a safe gamble of €1,100 or €1,540 and a risky gamble of €920 or €1,710. (Recall that in the CS method, each gamble also has a 50% chance of the low and high payoff.) Table Y shows the gambles of the EG method for the €1,960 income level.

TABLE 4.3 Overview of the pension-adapted CS method design									
	Option A				Option B				Next step†
	Prob.	Pension	Prob.	Pension	Prob.	Pension	Prob.	Pension	
1	50%	€1,540	50%	€1,100	50%	€1,710	50%	€920	2a / 2b
2a	50%	€1,440	50%	€1,160	50%	€1,620	50%	€1,040	3a / 3b
2b	50%	€1,680	50%	€980	50%	€1,730	50%	€860	3b / 3c
3a	50%	€1,350	50%	€1,210	50%	€1,540	50%	€1,100	4a / 4b
3b	50%	€1,620	50%	€1,040	50%	€1,680	50%	€980	4b / 4c
3c	50%	€1,710	50%	€920	50%	€1,740	50%	€800	4c / 4d
4a	50%	€1,350	50%	€1,210	50%	€1,440	50%	€1,160	5a / 5b
4b	50%	€1,540	50%	€1,100	50%	€1,620	50%	€1,040	5b / E
4c	50%	€1,680	50%	€980	50%	€1,710	50%	€920	F / 5c
4d	50%	€1,730	50%	€860	50%	€1,740	50%	€800	5c / 5d
5a	50%	€1,270	50%	€1,270	50%	€1,350	50%	€1,210	A / B
5b	50%	€1,440	50%	€1,160	50%	€1,540	50%	€1,100	C / D
5c	50%	€1,710	50%	€920	50%	€1,730	50%	€860	G / H
5d	50%	€1,740	50%	€800	50%	€1,740	50%	€740	I / J

† The next step is either the follow-up question or the final risk preference category (with the letter corresponding to the choices in the EG method, see Table 4.2 above). Left of the slash-sign is the next step for subjects who choose option A and right of the slash-sign for option B.

4.2.Methodology for determining the best-performing elicitation method

4.2.1. Determining the best-performing risk preference elicitation method

In theory, the best risk preference elicitation method is the one that measures true risk preferences. Risk-taking behaviour is often used as an indicator for true risk preferences. But validation of risk preference elicitation methods based on risk-taking behaviour can be difficult. First, it is debatable whether behaviour shows true risk preferences, since numerous other factors influence decision-making. Second, it is rare to simultaneously observe risk-taking behaviour and elicit risk preference. However, some studies attempt to validate risk preference elicitation methods based on risk-taking behaviour. They show, for instance, a relationship between the HL method and entrepreneurship, smoking, and seatbelt usage (Elston et al. 2005; Anderson and Mellor 2009). The CS method can similarly be related to smoking, drinking, and several financial decisions (Barsky et al. 1997; Lusardi 1998; Lahiri and Song 2000; Rosen and Wu 2004; Dave and Saffer 2007). Risk-taking behaviour in the pension domain is harder to observe. The data are private and, more importantly, current differences in pension risk-taking are based on conventional methods of risk preference elicitation. I therefore cannot use risk-taking behaviour to compare the risk preference elicitation methods in the pension domain.

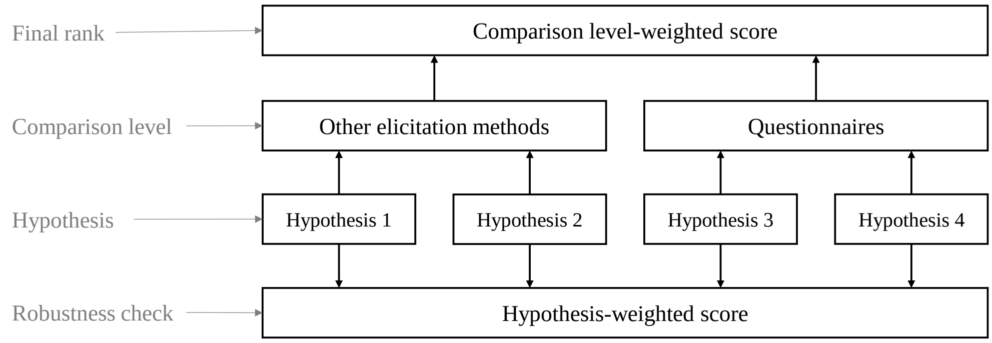
The literature presents other ways to compare risk preference elicitation methods (Charness, Gneezy, and Imas 2013; Holt and Laury 2014). They examine the relationship of an elicitation methods' risk preferences to risk preferences from other sources. In such comparison, similarity indicates accurately performing risk preferences, because they (presumably) measure the same underlying true risk preferences. (The next paragraph discusses a shortcoming of this assumption.) Risk preferences for such comparisons can come from other elicitation methods or from questionnaires. This thus creates two comparison levels to assess a risk preference elicitation methods' performance. Firstly, a comparison between risk preferences from different elicitation methods (like e.g. Anderson and Mellor 2009; Deck et al. 2010; Reynaud and Couture 2012; Crosetto and Filippin 2016). Secondly, a comparison between risk preferences from an elicitation method and from a questionnaire, most often the DOSPERT scale and SOEP question (like e.g. Van Rooij, Kool, and Prast 2007; Deck et al. 2010;

Dohmen et al. 2011; Kapteyn and Teppa 2011; Reynaud and Couture 2012; Crosetto and Filippin 2016). I discuss the results of these studies simultaneously with the relevant hypotheses in the next Subsections, 4.2.2 and 4.2.3.

This comparison mechanism has a shortcoming. It assumes that similarity in risk preferences indicates good performing elicitation methods. This is a fair assumption for risk preference that follow an EUT model or just when elicitation methods are accurate on average. But it does not hold when irrationalities similarly affect decision-making in the elicitation methods and hence bias risk preferences in a specific direction. In this case, a dissimilar method can actually do better than the other, more similar, methods. For instance, consider two methods that elicit risk preference in a gain-and-loss frame and one method that does so in a neutral frame. The former two elicitation methods probably produce more similar risk preferences and one of them is therefore considered the best. But loss aversion affects their risk preference elicitation, which increases risk aversion and produces biased risk preferences. Therefore, the dissimilar elicitation method with neutral framing would actually produce more accurate risk preferences. Although it is impossible to eliminate the chance of this happening, the minimisation of irrational effects makes it less likely. Because I use the elicitation methods that are least prone to irrationalities (see also the Rationality Criterion in Subsections 3.1.3 and 3.2.3), I think it is reasonable to assume the elicitation methods produce on average accurate risk preferences. Nevertheless, I also determine the best-performing method based on an additional comparison levels: the similarity between risk preferences of the elicitation methods to those of two questionnaires (i.e. the DOSPERT scale and the SOEP question). This creates more robustness in my comparison.

To determine the best-performing risk preference elicitation method in the pension context, I compare the methods based on four hypotheses. These are based on perfectly accurate risk preference elicitation and the more closely a method resembles this, the higher it scores. First, for every hypothesis, methods are thus ranked by closeness to the hypothesis. (The precise ranking methodology is different per hypothesis and discussed below.) Methods can rank first, second, or third and receives those respective scores. Note that a lower score thus equals a higher rank and hence is better. Second, for all hypotheses, the scores are aggregated per comparison level. Third, for all comparisons levels, the scores are aggregated for a final score. The second step ensures equal weight for the two comparison levels, irrespective of their number of hypotheses. (Note that the hypotheses on the questionnaires comparison level always consist of two rankings, one for each questionnaire.) Bypassing this step skews the final rank towards the comparison level with the most hypotheses. This is undesirable because the number of hypotheses per comparison level is quite arbitrary. Nevertheless, as a robustness check, I also compute a final score directly from the hypotheses scores. Such scoring gives equal weight to every hypothesis, but possible unequal weights to the two comparison levels. Figure 4.3 shows the two scoring mechanisms for the risk preference elicitation methods, based on either the hypotheses or the comparison levels.

FIGURE 4.3 Method scoring mechanism based on the hypotheses and comparisons levels



Scoring firstly takes place for each hypothesis. These are aggregated in two steps: at the comparison levels – the comparisons between different methods and with the questionnaires – and for the final rank. The hypothesis scores are also aggregated directly as a robustness check.

4.2.2. Hypotheses on the relationships between the elicitation methods' risk preferences

The first comparison of the risk preference elicitation methods tests whether the EG, HL, and CS method produce the same risk preferences. In theory, a person's risk preferences are the same in every elicitation. In practice, a person's risk preferences vary greatly between methods and experiments. For example, Crosetto and Filippin (2016) find mean risk preferences (i.e. CRRA coefficients) of r = 0.09 and r = 0.66 for the EG and HL methods, respectively. They subsequently compare their results to larger studies from the literature, which find r = 0.47 for the EG method (Eckel and Grossman 2008a) and r = 0.60 for the HL method (Filippin and Crosetto 2016).[23] An important problem of such comparisons is that their methods have different CRRA ranges, in number and size, which affect the elicited risk preferences, because people adjust to the provided range (Andersen et al. 2006) and generally avoid the extreme ends of the range (List 2007; Bardsley 2008; Crosetto and Filippin 2016). Good comparisons should avoid these pitfalls and hence produce more similar risk preferences. This is what Reynaud and Couture (2012) do. They use similar CRRA ranges for the EG and HL methods in their comparison, but still find different mean risk preferences for the EG and HL methods, r = 0.36 and r = 1.02, respectively. Despite these contrary findings, uniform elicitation methods (with respect to presentation, categories, and CRRA ranges) should produce similar risk preferences to indicate accuracy.

Hypothesis 1: the EG, HL, and CS methods produce the same risk preferences.

Risk preference elicitation methods that are more similar, are ranked higher for this hypothesis. This is computed by comparing each method's difference from the other two methods. In effect, three sub-hypotheses thus test the similarity between the different pairs of the EG, HL, and CS methods. The resulting absolute differences, if statistically significant, subsequently sum to total differences for each method. I rank elicitation methods according to these differences, from large to small differences, for the best to worst methods.

The second comparison of the risk preference elicitation methods tests whether the EG, HL, and CS methods' risk preferences are perfectly correlated. Higher positive correlations are generally accepted to indicate the methods' accuracy (Holt and Laury 2014). Overall, the literature only finds moderately positive correlations and these also vary greatly between methods and experiments (Crosetto and Filippin 2013b). Deck et al. (2010) find that the EG and HL methods are positively correlated (ρ = 0.21). These methods, however, are not correlated with two dynamic methods, the BART and DOND

[23] Crosetto and Filippin (2016) also investigate the investment game and BRET method methods and find risk preferences of respectively r = 0.73 and r = 0.78 themselves and r = 0.70 and r = 0.81 in the literature (Charness and Gneezy 2012; Crosetto and Filippin 2013a).

methods, although these are also positively correlated with each other (ρ = 0.27). Reynaud and Couture (2012) also find that the EG and HL methods are correlated (Spearman ρ = 0.4 and ρ = 0.67, for respectively low and high payoffs). Anderson and Mellor (2009) find that the HL and CS method are only weakly positively correlated (Spearman ρ = 0.175). Some of these results are promising, but truly accurate methods produce perfectly correlated risk preferences.

Hypothesis 2: the EG, HL, and CS methods correlate perfectly with each other.

Higher correlated elicitation methods are thus ranked higher. Three sub-hypotheses test the correlation of each pair of the EG, HL, and CS methods. All pairwise correlations are subsequently sorted from high to low, if the differences are statistically significant. I rank elicitation methods in accordance with the height of their correlations with the other two methods. Here, higher correlations mean a higher rank and thus better method.

4.2.3. Hypotheses on the relationships between questionnaires and elicitation methods

I also compare the elicitation methods' risk preferences to those of the DOSPERT scale and SOEP question, together referred to as the questionnaires. Positive and high correlations between questionnaires' and elicitation methods' risk preferences are, again, an important indication of accuracy. In the literature, correlations vary greatly, but overall elicitation methods are weakly to moderately correlated to questionnaires Reynaud and Couture (2012) use the EG and HL methods and compare their results to the DOSPERT scale and SOEP questions.[24] They find the EG and HL methods correlate with the DOSPERT scale (by respectively ρ = 0.40 and ρ = 0.33 in general and ρ = 0.28 and ρ = 0.26 in the financial domain). The correlations with the SOEP questions are insignificant, apart from the EG method in the financial domain which is very strong (ρ = 0.42). Crosetto and Filippin (2013b, 2016) take a similar approach but get slightly different results.[25] They focus on the general DOSPERT scale and the general SOEP question and find positive correlations for the EG (ρ = 0.30 and ρ = 0.30, respectively) and HL methods (ρ = 0.25 and ρ = 0.23). For the DOSPERT scale's domains of investment and gamble (i.e. financial), they only find significant correlation for the EG method and not for the HL method, like Reynaud and Couture (2012) in case of the SOEP questions' financial domain. Van Rooij, Kool, and Prast (2007) find a modest correlation for the CS method and the SOEP question (ρ = 0.24), as do Kapteyn and Teppa (2011) for another questionnaire (ρ = 0.15-0.21). These results are partly caused by fundamental differences between the questionnaires and elicitation methods. Nevertheless, accurate methods should still produce strongly (although perhaps not perfectly) correlated risk preferences.

Hypothesis 3: the EG, HL, and CS methods correlate perfectly with **a)** the DOSPERT scale and **b)** the SOEP questions.

Elicitation methods with risk preferences that are higher correlated to the questionnaires are thus ranked higher. In total, there are six sub-hypotheses for every correlation between, on the one hand, the EG, HL, and CS methods and, on the other hand, the DOSPERT scale and SOEP questions. Ranks are

[24] Reynaud and Couture (2012) experiment with low and high payoffs, as well as several domains for the questionnaires, but only the high payoffs and general and financial domains are reported here because they are most relevant to pensions.
[25] Crosetto and Filippin (2013b, 2016) also use the investment game, BART, and BRET methods – the only significant correlations are between the investment game and the financial-domain (i.e. the investment and gamble-domain) SOEP questions (ρ = 0.36 and ρ = 0.33, respectively) and the BART method and the general SOEP question (ρ = 0.37) – but these are omitted since they are unviable methods.

computed separately for the two questionnaires. The three correlations, between the elicitation methods and questionnaire, sort from high to low (i.e. from best to worst), if the differences are statistically significant.

The fourth comparison compares the elicitation methods to the questionnaires based on the latter's explanatory power in a regression. Several papers take this approach with moderately positive, but mixed results. Deck et al. (2010) regress the general and domain-specific DOSPERT scales on the EG and HL methods but do not find significant relations for either.[26] Dohmen et al. (2011) use a choice list and the SOEP question and do find a significant regression coefficients (β = 0.401 and β = 0.611, with and without controls, respectively). Crosetto and Filippin (2013b, 2016) also examine the relationship between the EG and HL methods and the DOSPERT scale and SOEP question with a regression. They look at the increase in adjusted R^2 (i.e. the increase in variance of risk preference that is explained) from inclusion of the questionnaires in an existing regression of the methods' risk preferences on age and gender. For the EG method, inclusion of the general DOSPERT scale or the SOEP question increases the adjusted R^2 with 1.6 and 2.7 percentage points, respectively, and this is 1.9 for the financial-domain DOSPERT scale. For the HL method, inclusion of neither questionnaire increases the adjusted R^2 significantly. Overall, like the correlations, one expects considerable (albeit not full) explanatory power for the questionnaires in case of accurate risk preference elicitation methods.

Hypothesis 4: the EG, HL, and CS methods are explained by **a)** the DOSPERT scale and **b)** the SOEP questions.

Elicitation methods with more explanatory power from the questionnaires thus rank higher. Again, there are six sub-hypotheses, one for every pair of the EG, HL, and CS methods and the DOSPERT scale and SOEP questions. Ranks are computed separately for the two questionnaires once more. They follow the "increase in adjusted R^2"-methodology of Crosetto and Filippin (2013b; 2016). The increases in adjusted R^2 sort in size (from high to low) to rank the elicitation methods (from best to worst) per questionnaire.

4.3. Experimental design for comparing the risk preference elicitation methods

4.3.1. Experimental design of the survey in the AFM Consumentenpanel

The experiment is conducted in cooperation with the Dutch Authority for Financial Markets (AFM). They manage an online voluntary consumers panel, the *AFM Consumentenpanel*, which I use for the experiment. The AFM approaches consumers through their website and other communication channels and these are supplemented by other consumers recruited through a survey company in order for the panel to better represent the Dutch population. Participants in the *AFM Consumentenpanel* are volunteers who themselves decide in which surveys they participate. They are presented with new surveys weekly and the survey is open to participants for about two weeks.

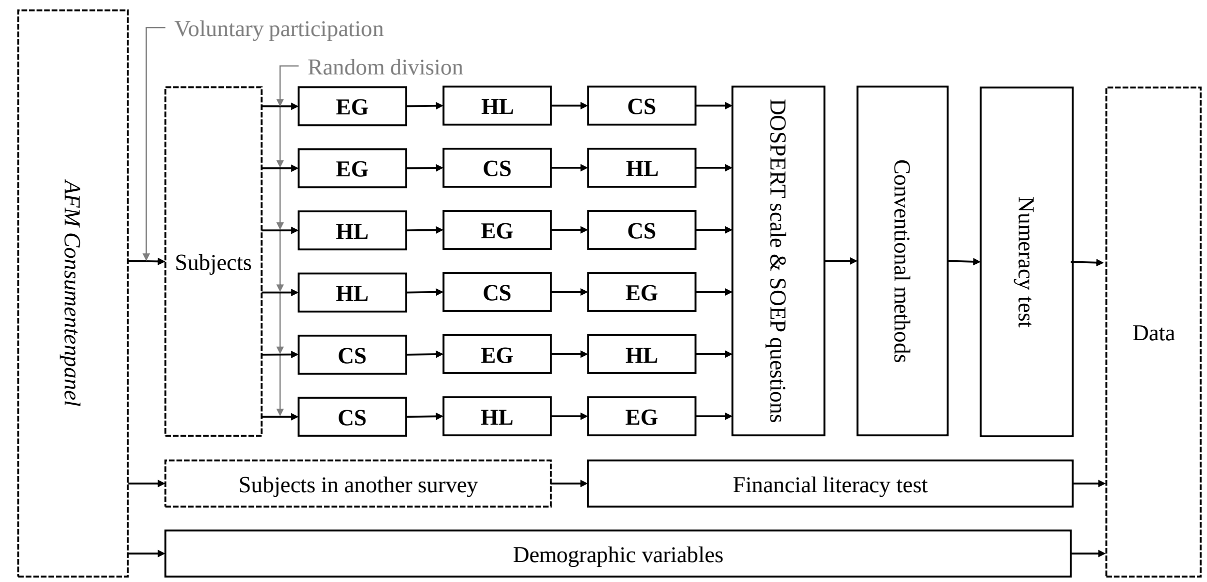
I apply two selection criteria on participants from the *AFM Consumentenpanel* to take part in my survey. Firstly, subjects over the age of 70 are not included. The risk preference elicitation methods ask subjects to imagine pension incomes at age 75. People close to this age will face unrealistic pension scenarios. while people over this age will be puzzled by the question no matter the incomes. Secondly, subjects in the lowest two income ranges (with gross yearly incomes between €0 and €12,500 or between €12,500 and €26,500) are not included. The multiplication factors of the custom pension scenarios

[26] Deck et al. (2010) also do this for the unviable DOND and BART methods as well as other domains of the DOSPERT scale. The only significant relation they find is small and negative, i.e. for the DOND method and investment-domain DOSPERT scale (β = -0.135).

produce pension incomes for these categories that are lower than the public pension. This makes such pension incomes unrealistic.[27] Hence, I do not use them in my experiment.

Figure 4.2 shows the experimental design of my survey with the *AFM Consumentenpanel*. It starts with the risk preference elicitation methods. I elicit subjects' risk preference with all three methods. (A detailed account of the risk preference elicitation methods is given in Subsection 4.1 on the adaptations of the EG, HL, and CS method.) The three methods can be presented in six different orders. I randomly divide subjects across these alternatives to control for possible order effects. The collection of secondary information comes after the risk preference elicitation methods. It comprises: two questionnaires for measuring risk preferences, i.e. the DOSPERT scale and SOEP questions; two conventional methods for measuring risk preferences; and a numeracy test. An earlier survey in the *AFM Consumtenpanel* conducted a financial literacy test (Cox, Kamolsareeratana, and Kouwenberg 2017). I use the financial literacy of those subjects that participated both in this earlier survey and in mine. The demographic variables of the subjects are already available because subjects provide this information when they join the *AFM Consumentenpanel*. (Appendix A.1 and A.2 show the complete experiment in respectively the English translation and the original Dutch version.)

FIGURE 4.2 Experimental design of the survey in the *AFM Consumentenpanel*



The data is collected from participants in the *AFM Consumentenpanel*. Subjects complete the EG, HL, and CS methods in random order first. Subsequently, they complete the DOSPERT scale, SOEP questions, conventional methods, and numeracy test. The financial literacy test data is collected from another survey. Participants in the *AFM Consumentenpanel* provide their demographic variables data when they join the panel.

4.3.2. Specification of the questionnaires, conventional methods, and tests

For the domain-specific risk-taking (DOSPERT) scale I use one of its five domains: financial (Weber, Blais, and Betz 2002; Blais and Weber 2006).[28] It present subjects with six risky financial activities (i.e.

[27] I use the public pension as a minimum because pension incomes lower than this are highly unrealistic. (Note that the pension incomes presented to subjects are net monthly pension incomes that already include the public pension.) Theoretically, private pensions could generate negative returns by leveraging their investments, which would result in pension incomes lower than the public pension. But, in practice, no pension fund will allow such a degree of risk in its pension products.

[28] These domains are based on Blais and Weber (2006). Weber, Blais, and Betz (2002) use six domains, because the financial domain is split into an investment and gambling domain. For both, the other five domains are ethical, gambling, health/safety recreational, recreational, and social. There is a Dutch version of the scale

three on investment and three on gambling) and ask them for the likelihood of engaging in them. I transform the original versions' five to seven-point scales, from "extremely unlikely" to "extremely likely," to a similar scale with ten points for a better comparison with the elicitation methods' ten risk preference categories. The DOSPERT-scale risk preference is calculated by taking the mean score for all six statements.

The German socio-economic panel (SOEP) asks subjects directly for their risk preference, both in general and in specific contexts, in my case financial and pension matters (Dohmen et al. 2005).[29] The original version uses an eleven-point scale, from "not at all willing to take risks" to "very willing to take risks," For a better comparison, I, again, transform it to a scale of ten points. The SOEP-question risk preference results directly from the self-assessed risk preference that subjects give as an answer.

The two conventional methods reflect common current practices in the industry.[30] My specific examples come from risk preference measurement for investing purposes, so I slightly adapt the framing to fit a pension context. The first question asks how subjects' sleeping behaviour is affected by a sudden significant drop in their expected monthly pension income. The four multiple-choice answers range from "sleepless nights" to "no effect at all." The second question asks subjects which decline in expected monthly pension income they would find acceptable. The six multiple-choice answers range from no decline to a decline of over 25%.

Numerical literacy is tested with Cokely et al. (2012)'s Berlin Numeracy Test, which "quickly assesses statistical numeracy and risk literacy."[31] It consists of two to three questions and adapts to subject's skill level. Based on this test, subjects are divided into four numerical literacy categories.

Financial literacy is tested by two multiple-choice questions and five true-or-false questions based on several similar tests in the literature (Van Rooij, Lusardi, and Allessie 2011a; 2012; Lusardi and Mitchell 2007; 2011; Cox, Kamolsareeratana, and Kouwenberg 2017).[32] The financial literacy score is determined by summing the number of correct answers, which minimises the score at 0 and maximises the score at 7. Note that the financial literacy data is not collected directly. Instead it is collected from a previous survey in the *AFM Consumentenpanel* and afterwards linked to the data for the subjects who completed both that and my survey.

---

available from the DOSPERT website (https://www8.gsb.columbia.edu/decisionsciences/research/tools/dospert), which I use.

[29] The contexts originally are car driving, financial matters, leisure and sports career, and health (Dohmen et al. 2005). Van Rooij, Kool, and Prast (2007) use a similar question for the context "pension matters." There is no Dutch version of the SOEP questions available. I therefore translate the question to Dutch from the German and English versions, which are available through the SOEP website (http://www.diw.de/en/soep; see also Appendix A.3 for the German version)

[30] Note that risk preference elicitation methods in the industry do vary significantly. Some parties use advanced methods in line with the academic literature, albeit a minority. Most parties, however, use similar to the examples or even simpler ones.

[31] Berlin's Numeracy Test is validated by 21 studies and takes less than three minutes to complete (Cokely et al. 2012). I use the computer adaptive test format. (See Figure A.2 in Appendix C for the decision tree of the computer adaptive Berlin Numeracy Test.)

[32] The first two questions are taken from Van Rooij, Lusardi, and Allessie (2011a), they are the second and third of their basic literacy questions, on respectively interest compounding and inflation. Their financial literacy questions are in turn based on the 2004 U.S. Health and Retirement Survey and other surveys on financial literacy (Lusardi and Mitchell 2007; 2011) and are already used with a Dutch sample (Van Rooij, Lusardi, and Allessie 2011a; 2012). The latter five questions are taken from (Cox, Kamolsareeratana, and Kouwenberg 2017) and are more tailored to investing.

5. Experimental results of the risk preference elicitation method comparison

5.1.Description of the sample's demographic characteristics and risk preferences

5.1.1. Description of the sample's demographic characteristics

Table 5.1ab describes the subjects in the sample of my experiment. It consists of 426 subjects. The majority of subjects, over four-fifth, is male; the remainder, less than one-fifth is female. My sample is relatively old. Only a quarter of subjects is younger than 50 years, a quarter is between 51 and 60 years old, and half between 61 and 70 years old. Subjects are highly educated: only 5% and 28% are in respectively the low and intermediate educational levels. About two-third of subjects are in the highest education level. Income and wealth are also high. The average income is about €66,000 (gross yearly) and a third of subjects are in the highest income category (with a gross yearly income over €78,500). Similarly, almost a third are in the highest wealth category (with wealth of more than €150,000). The six lower wealth categories have roughly equal proportions of subjects (between 10% and 20%). Numerical literacy is quite well divided among the four categories (28%, 27%, 19%, and 26% from low to high). But financial literacy is skewed to the higher end: 80% of subjects have intermediate or higher financial literacy. The sample is thus not entirely representative of the Dutch population. Fortunately, this is not a problem for my comparison, because I am primarily concerned with the within-subject differences and similarities between the risk preference elicitation methods and questionnaires.

TABLE 5.1 Sample descriptions

Variable	Category	Mean	Median	St.Dev.	*N*	Share
Gender (male=0; female=1)		0.17	0	0.38	426	
	Male				353	83%
	Female				73	17%
Age		57.84	60	9.75	426	
	21-50 years				104	24%
	51-60 years				113	27%
	61-65 years				92	22%
	66-70 years				117	27%
Education (1-3)		2.61	3	0.58	426	
	1 (low)				22	5%
	2 (intermediate)				122	28%
	3 (high)				282	66%
Income (1-5)		3.52	3	1.30	426	
	€26,500 - €33,000				37	9%
	€33,000 - €39,500				53	12%
	€39,500 - €66,000				130	31%
	€66,000 - €78,500				64	15%
	> €78,500				142	33%
Wealth (1-6)		3.89	4	1.87	403	
	< €10,500				66	16%
	€10,000 - €25,000				45	11%
	€25,000 - €50,000				67	17%
	€50,000 - €80,000				44	11%
	€80,000 - €150,000				52	13%
	> €150,000				129	32%

TABLE 5.1 Sample descriptions (continued)						
Variable	Category	Mean	Median	St.Dev.	*N*	Share
Numerical literacy (1-4)		2.42	2	1.15	426	
	1				120	28%
	2				117	27%
	3				79	19%
	4				110	26%
Financial literacy (0-7)		4.69	5	1.59	367	
	0 (lowest)				5	1%
	1				8	2%
	2				24	7%
	3				34	9%
	4				88	24%
	5				86	23%
	6				71	20%
	7 (highest)				49	13%

5.1.2. Preliminary description of the sample's elicited risk preferences

Table 5.2 shows the results of the risk preference elicitation methods. The mean risk preference categories of the EG, HL, and CS methods are very similar (between 3.39 and 3.76) and the medians are the same (3). This similarly holds for the CRRA risk preferences, with means between $r = 1.42$ and $r = 1.85$ and a median of $r = 2.31$. These are considerably higher than those of earlier large studies, which, for example, find $r = 0.47$ for the EG method (Eckel and Grossman 2008a) and $r = 0.60$ for the HL method (Filippin and Crosetto 2016). However, in the large-stake domain of pension risk, where risk aversion is likely higher, one study finds very similar risk preferences ($r = 1.92$) to mine (Alserda et al. 2016),[33] while another finds considerably higher ones (Van Rooij, Kool, and Prast 2007).[34] Considering this, my risk preferences look realistic. (Subsection 5.2.1. analyses the differences between the methods' risk preferences in more detail.)

TABLE 5.2 Overview of the risk preference elicitation methods			
	EG	HL	CS
Category			
Mean	3.39	3.60	3.76
Median	3	3	3
St.Dev.	2.88	2.46	2.59
CRRA (*r*)			
Mean	1.85	1.42	1.59
Median	2.31	2.31	2.31
St.Dev.	3.06	3.33	3.20
% risk averse	85.68	83.18	77.55
N consistent	426	321	294
% inconsistent	-	24.65	30.99

[33] Alserda et a. (2016) do not report the HL method's risk preference directly, see for details footnote 20.
[34] Van Rooij, Kool, and Prast (2007) do not provide CRRA risk preferences, however, it is possible to approximate them based on the data they do provide. There is a description of the CS method, which can be used to compute CRRA ranges fort the different risk preference categories, and a figure of the distribution along these categories, which can be used to estimate a mean CRRA risk preference. (I find a mean CRRA preference in this way of well-above $r = 2$.)

An important note before the further analyses though, is that the EG, HL, and CS methods produce risk preferences in two formats. Risk preference categories and CRRA risk preferences, which are based on the computed CRRA range of the categories with EUT. The former allows for easy comparison with the questionnaires, that have the same number of risk preference categories, and makes the results easily interpretable. The most important benefit of the latter is that they can directly translate into investment policies through portfolio theory (Markowitz 1952a), but these benefits are less relevant in the comparisons. Furthermore, risk preference categories are often preferred in comparisons (e.g. Crosetto and Filippin 2016). Therefore, I henceforth use the risk preference categories for my analysis, unless I explicitly state otherwise. Nevertheless, as a robustness check, I run the comparisons between the risk preference elicitation methods and their regressions on the demographic variables also with the CRRA ranges. For the regression analyses I use interval regression to account for the fact that I have CRRA ranges for my subjects. Interval regressions allow the dependent variable to be a range of values. For the other analyses, I use midpoints of the CRRA ranges. The results for the risk preference categories and the CRRA risk preferences are very similar.

5.1.3. Dealing with inconsistent answers in the HL and CS method

A significant share of subject gives inconsistent answers in the HL and CS methods. (The EG method requires a single choice and has no opportunity for inconsistent answers.) People act inconsistently in the HL method when they switch multiple times, make “backwards” choices, choose the dominated pension income, or follow no pattern whatsoever (Charness, Gneezey, and Imas 2013; Holt and Laury 2014). This occurs with almost a quarter of the subjects, as Table 5.2 shows. Some have proposed to interpret multiple switching points (i.e. going back-and-forth) as a longer period of indifference and hence a broader CRRA range (Harrison, Lau, and Ruttstrom 2006). As such, they may be included in the data. I look for this in my experiment, but it is only the case for 5 of the 105 inconsistent subjects. A remedy to inconsistent answers in the HL method is to force subjects to switch at a single point (Andersen et al. 2006; Tanaka, Camaerer, and Nguyen 2010), but I allow inconsistent answers in order to measure their occurrence. For the CS method, inconsistency means contradicting preferences over different rounds, which is the case for 31% of my subjects. For example, when they choose the risk averse pension income in the first question, but in later questions persistently choose risky pension incomes and hence end up with a risk seeking pension income. The literature usually does not allow inconsistent answers in this way (Barksy et al. 1997; Arrondel and Calvo Pardo 2002; Kapteyn and Teppa 2011), but I again do allow for it to measure inconsistency in subjects. Besides, forcing consistent answers results in misleading risk preferences, since subjects are also included if they, for instance, do not understand the instruction (Charness, Gneezy, and Imas 2013; Crosetto and Filippin 2013b). Consistent answers can to a considerable extent signal subjects’ comprehension of an elicitation method and hence make their risk preferences more reliable.

Dealing with inconsistencies thus poses a difficult problem. Allowing for them ensures the consistent risk preferences are much more reliable. It also identifies inconsistent subjects, which for some reason where unable to correctly complete the elicitation method. But it does not provide risk preferences for these subjects. Forcing consistent answers produces risk preferences for all subjects, but historical data of considerable inconsistency in elicitation methods, suggests these are not always reliable. In practical terms, a couple of inconsistent answers may be a good signal of subjects requiring extra attention for proper risk preference elicitation. But a large share of inconsistent answers makes an elicitation method unworkable due to the number of subjects requiring extra attention.

For comparisons between elicitation methods, this also poses a problem. Initially, a method with little inconsistency looks preferable. But consistency is not necessarily the result of a better method. It may mean that some subjects have consistent risk preferences while in the meantime not fully comprehending the elicitation method. It is thus difficult to draw any conclusions from the consistency of an elicitation method without an understanding of the underlying explanators. I therefore do not include consistency in my comparison and ranking of risk preference elicitation methods. But I do analyse possible explanation for inconsistency in Subsection 5.3.2.

5.2.Comparison and ranking of the risk preference elicitation methods

5.2.1. Comparison of the relationships between risk preference elicitation methods

Hypothesis 1 states that all three elicitation methods produce the same risk preference. To determine if subjects' risk preferences are the same across the EG, HL, and CS methods, I use the Skillings-Mack test, similar to Reynaud and Couture (2012).[35] It finds subjects' risk preferences significantly different across methods at a 1% significance level. A post-hoc pairwise comparison with Wilcoxon signed-rank tests and Bonferroni adjustments finds that all methods are different from each other. That is, the EG and HL methods are significantly different at a 5% significance level, the EG and CS methods at a 1% level, and the HL and CS methods at a 5% level as well. The EG method thus produces the most risk averse category at 3.39, followed by the HL method at 3.60 and then the CS method at 3.76. Since the methods produce significantly different risk preferences, I rank the methods by summing their differences to the other two methods. In this manner, the HL method ranks first, the CS method second, and the EG method third for the first hypothesis.

Hypothesis 2 states that all three elicitation methods are perfectly correlated with each other. Table 5.3 shows the Pearson correlations between the EG, HL, and CS methods. All correlations are significantly different from zero at a 1% significance level. But the hypothesis states perfect correlation (rather than no correlation), so I use a Fisher Z transformation to determine if the correlations are significantly different from $\rho = 0.9$. This level of correlation is generally considered as extremely strong and I hence interpret it as perfect correlation. I find that all correlations are also significantly different from 0.9 at a 1% significance level and reject the hypothesis. To rank the methods' correlations, I use Hotelling's t tests and Steiger's Z tests to determine if they significantly different from each other. Such comparisons require an observation for every variable, I therefore omit any subjects with missing observations, i.e. subjects with an inconsistent risk preference in any of the three elicitation methods. This decreases the sample size ($N = 224$) and increases the correlations slightly (from $\rho = 0.67$ to $\rho = 0.68$). I do not find any significant differences between the elicitation methods' correlations with the statistical tests. Therefore, all methods are ranked equally for the second hypothesis.

TABLE 5.3 Pearson correlations of the risk preference elicitation methods

	EG	HL	CS
EG	1		
	(426)		
HL	0.6153***	1	
	(321)	(321)	
CS	0.6691***	0.6702***	1
	(294)	(224)	(294)

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level) for the correlation's difference from zero. Sample size is indicated between brackets under the correlations

[35] Reynaud and Couture (2012) use the Kornbrot test, an alternative to the Wilcoxon signed-rank test, because they only compare two methods (Kornbrot 1990; Haidous and Sawilowsky 2013). The Skillings-Mack test is a nonparametric test for the hypothesis that multiple variables are the same for a within-subject design (Skillings and Mack 1981). It is an alternative to the Friedman test in case of missing data, which is the case here due to the inconsistent answers in the HL and CS methods.

All correlations are moderately high. The EG and HL method have a correlation coefficient of 0.62 and the CS method has correlations coefficients with the EG and HL methods of 0.67 both. Such correlations are comparable to those of Reynaud and Couture (2012) between the EG and HL methods in their high payoffs version (Spearman ρ = 0.67), but higher than those of most other studies (Anderson and Mellor 2009; Deck et al. 2010; Crosetto and Filippin 2013a). As a robustness check, I also look at Spearman correlations (see Table A.7 in Appendix D). Whereas normal (Pearson) correlations assess the linear relationship between variables, Spearman correlations assess the monotonic relationship, i.e. the stability of ranks. This follows the approach of Crosetto and Filippin (2016) and several authors which also use this measure in other comparisons (Anderson and Mellor 2009; Reynaud and Couture 2012). The Spearman correlations are as significant as the normal (Pearson) correlations, albeit slightly lower (from Spearman ρ = 0.47 to ρ = 0.61).

5.2.2. Comparison of the relationships between questionnaires and elicitation methods

Another comparison level uses the risk preference questionnaires. The DOSPERT scale encompasses risk preferences in the financial domain, i.e. it is a combination of the investment and gamble domains. Responses to the domains statements have a Cronbach's alpha, a measurement of internal consistency, which is considered good (α = *0*.83). The SOEP questions ask for self-reported risk preferences in general and in the financial and pension domains in specific. Table 5.4 shows the results of the questionnaires. Given that risk preferences depend on the domain, I only use the pension-domain SOEP question for my comparisons (and refer to it simply as the SOEP question), since it is the most relevant measures for the pension context and significantly different from the others.[36] The DOSPERT scale and the SOEP question are significantly different from each other at a 1% significance level following a Wilcoxon signed-rank test. The DOSPERT scale produces a mean risk preference of 2.45 on a scale from 1 to 10 (from risk averse to risk seeking). The SOEP question produces a mean risk preference of 4.16 on a similar scale. Compared to the risk preferences of the elicitation methods (between 3.39 and 3.76), the former is relatively low and the latter relatively high. (They are all different at a 1% significance level according to a Wilcoxon signed-rank test.) However, the questionnaires provide relative rather than absolute risk preferences, which makes dissimilar levels less relevant.

TABLE 5.4 Overview of the questionnaires

	DOSPERT	SOEP (pension)	SOEP (financial)	SOEP (general)
Category				
Mean	2.45	4.16	4.68	4.76
Median	2	4	4	4
St.Dev.	1.43	1.99	2.06	2.04
N	426	426	426	426

Relevant for the comparison with the questionnaires is the broader similarity of risk preferences. Hypothesis 3 states that each of the three elicitation methods is perfectly correlated with the DOSPERT scale and the SOEP question. I look at the Spearman correlations between the two for this. These are

[36] The three SOEP questions (general, financial, and pension) have a Cronbach's alpha that is high (α = 0.93). I test if the answers to the SOEP questions are significantly different with a Friedman test. They are different at a 1% significance level. Post-hoc pairwise comparisons with the Wilcoxon signed-rank test and Bonferroni adjustments show all risk preferences are different from each other, the pension domain is different from the other domains at a1% significance level. The general and financial domain are different from each other at a 5% significance level. This suggests subjects' risk preferences are different in general and in the financial and pension domains and confirms earlier findings of domain dependency in risk preferences (Weber, Blais, and Betz 2002; Blais and Weber 2006).

more suitable than normal (Pearson) correlation for comparisons between questionnaires and elicitation methods (Reynaud and Couture 2012; Crosetto and Filippin 2016). Table 5.5 shows the Spearman correlations between the EG, HL, and CS methods and the DOSPERT scale and the SOEP question. The DOSPERT scale's correlations with the elicitation methods are significantly different from zero at a 10% significance interval, while the SOEP question's are significantly different from zero at a 1% significance level. Since the third hypothesis involves perfect correlation (rather than no correlation), I use a Fisher *Z* transformation to determine if the correlations are also significantly different from ρ = 0.9. All correlations are also significantly different from ρ = 0.9 at a 1% significance level. Therefore, I reject the third hypothesis.

To rank the methods' correlations, I use Hotelling's *t* tests and Steiger's *Z* tests to determine if they are significantly different from each other. Again, subjects with inconsistent risk preferences are omitted from these tests, because they require an observation for every variable in the comparison. These omissions decrease the sample sizes (to between *N* = 224 and *N* = 321) and generally increase the correlations (to between ρ = 0.08 and ρ = 0.12 for the DOSPERT scales and between ρ = 0.26 and ρ = 0.41 for the SOEP questions). For the correlations with the DOSPERT scales, there are no significant differences between elicitation methods. For the correlations with the SOEP questions there are: the EG method is significantly different from the CS method at 10% and 5% significance levels for the *t* and *Z* statistic, respectively (*N* = 224); and the HL method is significantly different from the CS method at a 10% significance level for the *Z* statistic only (*N* = 294). Since not all authors use the Spearman correlations, e.g. Crosetto and Filippin (2016), I also look at the normal (Pearson) correlations as a robustness check (see Table A.7 in Appendix D). These are similarly significant and compared to the Spearman correlations slightly higher for the DOSPERT scales (between ρ = 0.10 and ρ = 0.15) and slightly lower for the SOEP questions (between ρ = 0.23 and ρ = 0.36).

TABLE 5.5 Spearman correlations of the risk preference elicitation methods and questionnaires

	EG	HL	CS
DOSPERT	0.0806*	0.1037*	0.1082*
SOEP (pension)	0.2669***	0.3044***	0.3931***
N	426	321	294

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level) for the correlation's difference from zero.

The DOSPERT-scale's correlations are quite low (from ρ = 0.08 to ρ = 0.11). In the literature, comparisons find stronger correlation for both the general and financial-domain DOSPERT scales (from ρ = 0.25 to ρ = 0.40) with the EG and HL methods (Reynaud and Couture 2012; Crosetto and Filippin 2016). Although another study does not find a significant correlation between the financial-domain DOSPERT scale and the HL method (Crosetto and Filippin 2016). Nevertheless, these results are disappointing. A possible reason for them could be that the Dutch translation of the DOSPERT scale is not as extensively validated and hence may not be as suitable as one would hope (Weber, Blais, and Betz 2002; Blais and Weber 2006). In any case, because there are nog significant differences between the elicitation methods' correlations, all methods are ranked equally for the DOSPERT-part of the third hypothesis.

The SOEP question's correlations are moderate (from ρ = 0.27 to ρ = 0.39). But they are comparable if not better than the literature. One study only finds a significant correlation for the EG method and financial-domain SOEP question, albeit a high one (ρ = 0.42), and none for the HL method or general SOEP question (Reynaud and Couture 2012). Others do find significant correlations, although lower (from ρ = 0.24 to ρ = 0.30), between all three methods and the general SOEP question (Van Rooij, Kool, and Prast 2007; Crosetto and Filippin 2016). For the SOEP-part of the third hypothesis, I do rank the methods based on their approximation of perfect correlation: the CS method is first, the HL method second, and the EG method third.

Hypothesis 4 states that each of the three elicitation methods is explained by the DOSPERT scale and the SOEP questions. The explanatory power is measured in the same way as Crosetto and Filippin (2016). They regress the questionnaires and several demographic variables on the elicitation methods' risk preferences and look at the increases in adjusted R^2 from inclusion of each questionnaire. It involves two regressions, one with and one without the respective questionnaire, to determine the increase in adjusted R^2 from the questionnaire. The base regression only involves the risk preference and controls, i.e. the demographic variables – age, gender, education, income, and wealth – and two order dummies for the second and third position of the method. The follow-up regression is the same, but for the inclusion of the questionnaire.

Table 5.6 shows the increases in adjusted R^2 from inclusion of the DOSPERT scale and the SOEP question for the three elicitation methods. The explanatory power of the questionnaires is pretty low, i.e. the increase in adjusted R^2 does not exceed 15% (expressed in extra percentage point of explanatory power). Nevertheless, they are considerably larger than those of Crosetto and Filippin (2016). They only find small increases for the EG method (2.7% for the SOEP question and 1.6% and 1.9% for the general and financial-domain DOSPERT scales, respectively) and none for the HL method. My findings also show clear differences between the DOSPERT scale and SOEP question. The former increases the adjusted R^2 much less than the latter and the differences between elicitation methods is also smaller. Furthermore, the differences across methods are inverted between the two questionnaires. I rank the methods based on their increases in adjusted R^2, which means: the EG method first, the HL method second, and the CS method third for the DOSPERT scale; and the exact opposite for the SOEP question: the CS method first, the HL method second, and the EG method third. Given that the changes in adjusted R^2 decrease or increase with sample size for the DOSPERT scale and SOEP question, respectively, I run the same regressions with only the consistent subjects ($N = 214$) as a robustness check (see Table A.9 in Appendix D). But this limitation on the regressions only emphasises the above relationships and I hence discard it as a possible explanation.

TABLE 5.6 Increases in adjusted R^2 due to the inclusion of the questionnaires

	EG	HL	CS
DOSPERT	2.67	2.51	1.74
SOEP (pension)	5.92	8.04	13.80
N†	403	306	279

† The sample sizes differ per elicitation method because of omitted data for the inconsistent risk preferences and, to a lesser extent, missing demographic variables.

5.2.3. *Ranking of the different risk preference elicitation methods*

Every hypothesis provides a score for each risk preference elicitation method. I aggregate these scores at the two comparisons levels (i.e. elicitation methods and questionnaires). Table 5.7 shows the ranking of the EG, HL, and CS methods based on the hypothesis and comparison level scores. For the methods-comparison level, the HL method ranks first, the CS method second, and the EG method third. For the questionnaires-comparison level, the CS method ranks first, the HL method second, and the EG method again third. Overall, the HL and CS methods jointly perform best and the EG method performs worst. As a robustness check, I also compute the final ranking with equal weight for every hypothesis. The HL and CS method still jointly perform best and the EG method's performance remains the worst. Given the similar results, I see no reason to deviate from the initial ranking methodology. The HL and CS methods are thus the best-performing risk preference elicitation methods in the pension domain and the EG method performs worst of the three.

TABLE 5.7 Ranking of the risk preference elicitation methods based on the hypothesis scores

	Hypothesis						Total
	1	2	3a	3b	4a	4b	
EG	3	2	2	3	1	3	3
HL	1	2	2	2	2	2	1/2
CS	2	2	2	1	3	1	1/2
	Total						
	Methods		Questionnaires				
EG	3		3				3
HL	1		2				1/2
CS	2		1				1/2

5.3. Additional analyses: comparison with the conventional methods and possible explanations for inconsistency and instability in elicited risk preferences

Besides ranking the EG, HL, and CS methods, the experiment also provides data that allow several additional analyses. Subjects also complete two conventional methods of risk preference measurement for pension, which can be similarly compared to the above methods and questionnaires. It is also useful to find the causes of inconsistent and unstable risk preference elicitation. If these findings can be explained by subjects' demographic information or numerical and financial literacy, inconsistency and instability may be decreased or even avoided in future risk preference elicitation methods.

5.3.1. *A comparison the conventional methods' risk preferences*

In addition to the comparison between the EG, HL, and CS risk preference elicitation methods, I also make a comparison between those methods and the conventional methods found around the financial sector. For this reason, subjects also complete two conventional methods during the experiment. It is interesting to investigate how these compare to the more scientifically-based methods, because this could prescribe changing methods in practice. Although the two conventional methods are not significantly different from each other following a Wilcoxon signed-rank test, they have a low internal consistency according to Cronbach's alpha (α = 0.59). Furthermore, the first and second conventional method have the same medians ($\tilde{x}$ = 2) and similar means and standard deviations (μ = 2.26 with σ = 0.79 and μ = 2.29 with σ = 1.14), but different choice ranges (from 1 to 4 and from 1 to 6), which is suprising.

To compare them to the other risk preference elicitation methods, I again use the Spearman correlations, which look at rankings. Table 5.8 shows the Spearman correlations between the EG, HL, and CS methods, the questionnaires, and the two conventional methods. (Subsection 4.3.2 discusses the correlations between the EG, HL, and CS methods and the questionnaires.). Most correlations are different from zero at a 1% significance level, apart from those between the EG, HL, and CS methods and the DOSPERT scale and the conventional methods. The former are all different from zero at a 10% significance level. For the EG, HL, and CS and conventional methods, all but one are insignificantly different from zero. The exception being that between the CS method and the first conventional method, which is different from zero at a 10% significance level. Using the Fisher *Z* transformation, I find that all reported correlations are different from $\rho = 0.9$ at a 1% significance level. As a robustness check, I also look at normal (Pearson) correlations, but these are only slightly lower with comparable significance levels (see Table A.10 in Appendix D).

TABLE 5.8 Spearman correlations of the risk preference elicitation methods and questionnaires

	EG	HL	CS	Conventional I	Conventional II
DOSPERT	0.0806*	0.1037*	0.1082*	0.2887***	0.3254***
SOEP (pension)	0.2669***	0.3044***	0.3931***	0.2954***	0.2968***
SOEP (financial)	0.1626***	0.2261***	0.3326***	0.3475***	0.3717***
SOEP (general)	0.1984***	0.2521***	0.3222***	0.3037***	0.3040***
Conventional I	0.0414	0.0725	0.1007*	1	
Conventional II	0.0313	0.0607	0.0710	0.4728***	1
N	426	321	294	426	426

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level) for the correlation's difference from zero.

There are several comparisons to be made between these correlations. First, consider the correlations within each method sort. Recall that the EG, HL, and CS methods are moderately to highly correlated with each other (between $\rho = 0.62$ and $\rho = 0.67$). The conventional methods are considerably less correlated internally ($\rho = 0.47$). The two sorts of methods are, however, compared with different correlations, i.e. Spearman and Pearson correlations, respectively. But for either correlation, there remains a difference, albeit a smaller one. The EG, HL, and CS methods have lower Spearman than Pearson correlations (between $\rho = 0.47$ and $\rho = 0.61$), and the conventional methods have higher Pearson than Spearman correlations, but less so ($\rho = 0.45$). The Pearson correlations are all significantly different from each other at a 1% significance level, according to Fisher *Z* transformations. The Spearman correlations are insignificantly different, apart from that between the EG and CS methods on the one hand and conventional methods on the other, which are different at a 1% significance level. This suggests the EG, HL, and CS methods produce more internally coherent risk preferences than the conventional methods. Especially for absolute risk preference (based on Pearson correlations), but even for relative risk preferences (based on Spearman correlations). This difference can be explained by the real pension incomes and concrete (CRRA) risk preferences of the scientific methods, rather than the more subjective questions of the conventional methods, which are much harder to calibrate. As such, in case of widespread use of one of these sorts of methods, the former produce more consistent risk preferences across variations of the method.

Second, like the EG, HL, and CS methods for Hypothesis 3, the Spearman correlations between the conventional methods and the questionnaires are investigated. The EG, HL, and CS methods have small and moderate correlation with respectively the DOSPERT scale and the SOEP question (on average $\rho = 0.10$ and $\rho = 0.32$). All correlations are different from zero and $\rho = 0.9$ and most at a 1%

significance level.[37] The conventional methods have moderate correlations with both questionnaires (on average ρ = 0.31 and ρ = 0.30, respectively), which are all different from zero and ρ = 0.9 at a 1% significance level. The differences in correlations between the scientific and conventional methods for the DOSPERT scale are significant at a 1% significance level, according to a Fisher *Z* transformation. But they are not significantly different for the SOEP question, apart from the correlation between the CS method and conventional methods, at a 10% significance level. The conventional methods thus have higher correlations with the DOSPERT scale and slightly lower correlations with the SOEP question. So far, I have only looked at the pension-domain SOEP question, since it is most relevant for pension-context risk preferences and had the highest correlations with the EG, HL, and CS methods out of the three variants. The conventional methods, however, have higher correlations with the general and, especially, financial-domain SOEP questions. The differences between the SOEP questions are only significant for the EG method, at a 5% to 10% level with the Hotelling *t* tests and Steiger *Z* tests, respectively. But the pattern is noteworthy nonetheless. The conventional methods may be more suited for financial risk preferences, rather than those for pensions specifically. This may also explain their better performance with the DOSPERT scale, which measures financial and not pension risk preferences after all.

Finally, the correlations between the EG, HL, CS, and conventional methods themselves are also of some interest. These correlations, or lack thereof, are low and not significantly different from zero, apart from the CS and first conventional methods, which are different from zero at a 10% significance level. It suggests that the two sorts of methods measure completely different "risk preferences", despite shared moderate correlations with third risk preference indicators, i.e. the questionnaires. This is important because it suggests that the choice between the scientific EG, HL, and CS methods and the current conventional methods results in very different elicited risk preferences.

5.3.2. Inconsistent answers are (partly) caused by exhaustion and numerical literacy

Inconsistent risk preferences are an important problem in risk preference elicitation methods. They do not produce risk preference and hence cannot be interpreted Given the proposed link between complexity and inconsistent answers (Charness, Gneezey, and Imas 2013; Holt and Laury 2014), I take a closer look at this explanation. Table 5.9 shows, per elicitation method, the number of consistent answers, the percentage of inconsistency, and the perceived easiness. One straightforward indicator of complexity is the self-reported easiness of methods. If methods' perceived easiness is related to inconsistent answers, complexity may be the explanation. To determine whether the reported easiness is significantly different across method, I use a Friedman test, which finds differences at a 1% significance level. This difference is further specified with post-hoc pairwise comparisons with Wilcoxon signed-rank tests and Bonferroni adjustments. They find significant differences in the easiness of the EG and HL methods at a 5% significance level and the EG and CS methods at a 1% significance level, but no significant difference between the easiness of the HL and CS methods. Since, only the EG method is significantly different from the other methods, the possibility of inconsistent answers may make a method feel more complex, but there is no clear relationship between the percentages of inconsistent answers and the self-reported easiness. As such, like Crosetto and Filippin (2016), I find no relationship between inconsistent risk preferences and perceived complexity of the method.

[37] The exception are the correlations with the DOSPERT scale, which are only significantly different from zero at a 10% significance level (see also Subsection 5.2.2 and Table 5.5).

TABLE 5.9 Overview of inconsistency in the risk preference elicitation methods			
	EG	HL	CS
N consistent	426	321	294
% inconsistent	-	24.65	30.99
Easiness	6.32	6.52	6.60

Given these inconclusive results, I also examine other variables that may cause inconsistent answers. In order to investigate possible other causes of inconsistent risk preference, I run four logistic regressions of inconsistency on numerical and financial literacy and several demographic variables. The dummy variable for inconsistency has a value of one for any subject that has inconsistent risk preferences in either the HL, the CS, or both methods and a value of zero for all other subjects. First, I regress inconsistency on several demographic variables and I subsequently include numerical literacy (for the second logit regressions), financial literacy (third), and both literacies (fourth). (Table A.11 in Appendix D shows the results for the logistic regressions of inconsistency in the HL and CS methods.) There are several significant relationships that emerge from these logistic regressions. The second (of four) age group is more likely to give inconsistent answers. The finding is persistent over the four logit regressions at a 10% significance level. It is hard to explain why subjects between the age of 50 and 60 years old are more likely to answer inconsistently than younger or older subjects. But there are other significant explanators that do make sense, namely: educational level and numerical literacy (at 1% to 5% significance levels). Higher educational levels and higher numerical literacy decrease the chance of inconsistent answering, suggesting the complexity of elicitation methods can be a challenge for subjects.

To determine if this relationship holds for both methods, I run the same logistic regressions on the inconsistent answers of the HL and CS methods separately. Because the consistency is now measured per method, I include a dummy variable for the order in which the method was presented to subjects. The relationship to complexity (i.e. educational levels and numerical literacy) disappears for the CS method. (Table A.12 in Appendix D shows the results for the logistic regressions of inconsistency in the CS methods.) The fourth income group is considerably more likely to answer consistently at a 10% significance level. Furthermore, the second order – with the CS method second, after the EG method and before the HL methods – decrease the changes of an inconsistent answer compared to the first order – when the CS method is last, after the HL method. This indicates subjects get exhausted when they complete two long methods after each other.

For the HL method, however, the relationship to complexity (i.e. educational levels and numerical literacy) holds better. Table 5.10 shows the results of the logistic regressions of inconsistency for the HL method. Subjects seem more inconsistent as they get older, but only at 5% and 10% significance levels. Educational level, again, decreases inconsistencies as it gets higher. Although the relationship is a bit more sporadic and weaker than before (either a 5% or 10% significance levels). Numerical literacy has an even stronger positive effect on subjects' consistency and at a 1% significance level. Furthermore, financial literacy, when numerical literacy is not considered, also improves consistency at a 5% significance level. Finally, the order of the methods also has an impact: the second and fifth order – when the HL method is last – increase inconstant answers. This corresponds with the theory of fatigued subjects underperforming on consistency. Together with the role of educational levels and numerical literacy, consistency in the HL method seems closely related to intellectual capacity and exhaustion of subjects. An additional problem may be the relatively high educational level and numerical literacy of my sample, which means population-wide application will see an even larger share of inconsistent answers. On the other hand, when subjects face only one elicitation method that is incentivised, i.e. it affects their pension risk, they may be less susceptible to fatigue. Nevertheless, complexity seems an important challenge for the HL method and fatigue is for both methods.

Returning to educational level, it has an exceptionally strong positive effect, especially in relation to the first level. Since this represents only 22 subjects with the lowest educational level, which have a much higher degree of inconsistent answers (82%) than the others, I look at the relationship between the intermediate and high educational levels more closely. They differ much less from each other in the first HL regression, and the difference actually disappears and subsequently reverses in the second to fourth regression which include numerical and financial literacy. As a robustness check, I therefore also run the logistic regressions without the first educational level (see Table A.13 in Appendix D). The significant effect of educational level now totally disappears, while numerical literacy's remains roughly the same. Thus, the subjects with lowest educational level have exceptional difficulties with the HL method, but at higher educational levels it is especially numerical literacy that affects inconsistency.

TABLE 5.10 Logistic regressions for inconsistency in the HL method

	Logit I	Logit II	Logit III	Logit IV
Female	0.0780	-0.1350	-0.1366	-0.2804
Age group				
2	0.5704	0.5796	0.6688	0.7331*
3	0.6301	0.5950	0.5070	0.5086
4	0.7856	0.6328*	0.8311**	0.7176*
Education				
2	-0.6248	-0.5355	-0.9796*	-0.8222
3	-1.0521**	-0.8299	-1.2477**	-1.0427*
Income				
2	0.0782	-0.0158	0.2510	0.1979
3	-0.4321	-0.4578	-0.6446	-0.6088
4	-0.3242	-0.2898	-0.1805	-0.1589
5	-0.6603	-0.6018	-0.6254	-0.5141
Wealth				
2	-0.0044	0.0507	-0.4252	-0.4113
3	-0.4555	-0.3920	-0.5992	-0.5879
4	-0.4998	-0.4273	-0.1539	-0.2067
5	-0.6297	-0.5825	-0.7022	-0.6709
6	-0.4766	-0.3918	-0.6049	-0.6062
Order				
2	0.9442**	0.9427**	1.0451**	1.0363**
3	0.2428	0.1920	0.4267	0.3442
4	0.6433	0.5681	0.6523	0.5491
5	1.1167**	1.1191**	1.1777**	1.1422**
6	0.3159	0.2494	0.2162	0.1543
Numerical literacy		-0.5207***		-0.5269***
Financial literacy			-0.2189**	-.1579
Constant	-0.7241	0.3304	0.5247	1.2732
Pseudo R^2	8.37%	12.61%	11.38%	15.33%
N	403	403	347	347

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

5.3.3. *A closer look at stability of risk preferences and an attempt to explain it*

The primary comparison uses stability of risk preferences as an indicator of accuracy (e.g. through correlations), but they do not investigate the degree of stability itself. Knowing the causes of (in)stable risk preference, can improve future risk preference elicitation methods. I therefore examine the risk preference stability more closely. To investigate risk preference stability across elicitation methods, I construct a new variable which measures stability as the inverse of the difference in risk preferences. It is the average absolute difference between the EG, HL, and CS methods' risk preferences. Total stability thus means no (zero) difference, while higher differences mean less stability. For the subjects with inconsistent risk preferences in either the HL or CS method, the stability is determined by the difference between the EG method and the CS or HL method, respectively. Table 5.11 shows the average risk preference stability across the three elicitation methods. Out of 401 subjects, 81 subjects (or 21%) have totally stable risk preferences and 239 subjects (61%) approach stability, with maximally $1^1/_3$ category difference in risk preferences. These results are similar to those in the literature (Anderson and Mellor 2009; Reynaud and Couture 2012). Four out of five subjects on average differ two or less choices between the methods. This means that although the implied precision of specific risk preference categories may be unjustified, overall the elicitation methods provide a pretty constant measure of a subject's risk preference.

TABLE 5.11 Risk preference stability by average difference between choices across the methods

	Difference in choices								
	0	$^2/_3$	1	$1^1/_3$	2	$2^2/_3$	3	$3^1/_3$	4 ≤
Frequency	81	55	44	59	82	24	17	7	32
Cumulative frequency	81	136	180	239	321	345	362	369	401
Percentage	21%	14%	11%	15%	18%	6%	4%	2%	8%
Cumulative percentage	21%	35%	46%	61%	80%	86%	90%	92%	100%

To explain stability, I run a linear regression of stability on the demographic variables and numerical and financial literacy. But none of the variables in this regression significantly affect stability. As an additional test, I also run logistic regressions for the chance of totally stable risk preferences. For this reason, I create a dummy variable for stable risk preferences, which is zero if there is a difference in risk preference categories and one when there is not (i.e. zero difference). The latter is the case for 81 subjects (or 21%), as indicate above. Like the previous regression, however, the logistic regression of stable risk preferences provides no explanation for stability from the demographic variables or numerical and financial literacy, i.e. all regression coefficients are insignificant. (Tables A.14 and A.15 in Appendix D show the results of the linear regression of risk preference stability and the logistic regressions of stable risk preferences, respectively.) Because neither educational levels or numerical and financial literacy affect stability in any of the regressions, it seems complexity is not the cause of instability of risk preference, which is the same as Reynaud and Couture (2012) conclude based on their similar analysis. This suggest instable risk preference across elicitation methods are not the result of incomprehension for certain methods. Since other demographic variables cannot explain instability either, it remains difficult to identify the determinants for instable risk preferences.

6. Conclusion

There is no consensus, in either academia or practice, on which risk preference elicitation method is best for the pension domain. This is problematic because Dutch pension funds have a legal obligation to elicit clients' risk preferences. I address this problem by examining which method performs best in the pension domain. A literature-based comparison of methods narrows down the most viable risk preference elicitation methods for the pension domain. First, methods should meet four criteria: produce CRRA parameters, use realistic pension scenarios, test rational risk preferences, and be easy and short. This leaves three viable risk preference elicitation methods for the pension context: the multiple lottery choice or EG method from Eckel and Grossman (2002), the choice list with two gambles or HL method from Holt and Laury (2002), and the choice sequence or CS method. The other risk preference elicitation methods are not viable for the pension domain because they do not meet one or more criteria: the investment game, the certainty equivalent, the choice list and sequence with certain payoffs, and the dynamic methods are all too prone to irrationalities; the certainty equivalent is also too complex; and the dynamic method does not use realistic pension scenarios.

An experiment-based comparison subsequently determines which of the EG, HL, and CS methods performs best in the pension domain. Subjects drawn from the *AFM Consumentenpanel* complete the three methods in random order, as well as several other questions. The methods are adapted to the pension context in four ways: a pension-specific presentation, an equal number risk preference categories, uniform CRRA ranges, and customised pension scenarios. I determine the methods' performance based on two comparison levels: the similarity between the methods' risk preferences and the similarity between each method's risk preferences and those of two questionnaires, the DOSPERT scale and SOEP question. My results show risk preferences that are quite similar for the different methods and correlate moderately to strongly. The HL method performs best on similarity, followed by the CS method, and then the EG method. The methods' correlations with and explanatory power from the questionnaires is weak (for the DOSPERT scale) to moderate (for the SOEP questions). For this comparison level, the CS method performs best, followed by the HL method. The EG method, again, performs worst. Summing the methods' positions, I conclude that the HL and CS risk preference elicitation methods together perform best in the pension domain, followed by the EG method.

Independent of this comparison, I make three additional analyses. Firstly, I examine the conventional methods. These do not significantly correlate with the EG, HL, and CS methods and only moderately with the two questionnaires. Interestingly, the conventional methods also correlate less with each other than the EG, HL, and CS methods. Secondly, I examine what may cause inconsistent risk preferences in the HL and CS methods. For the HL method, this is partly explained by educational level and numerical literacy. But I find no explanation for inconsistency in the CS method. Thirdly, I look at the stability of risk preference across elicitation methods. Only one fifth is completely stable, but the majority is relatively stable, changing one to two risk preference categories between methods. I try to explain stability, but I do not find any significant relationships.

These results have implications for risk preference elicitation in the pension domain and elsewhere. Although elicitation methods abound, my four viability criteria show some are more viable for the pension domain than others, specifically, the EG, HL, CS methods. Therefore, it narrows down the field of risk preference elicitation methods for both future research and practice. I also show how to successfully adapt these methods to the pension domain, with respect to their presentation, risk preference categories, CRRA ranges, and realistic pension scenarios. Adapting methods to the relevant task and domain allows for more accurate risk preference elicitation. Furthermore, my experiment with the EG, HL, and CS methods shows relatively similar and strongly correlated risk preferences. These results are better than those in much of the literature and imply greater accuracy than one would initially expect. The stability of risk preference across elicitation method, although relatively good, remains imperfect. This is problematic for presumably accurate risk preference elicitation. The results are, however, better than those for the conventional methods, suggesting wider application could improve risk preference elicitation for pensions nevertheless. Furthermore, inconsistency in the HL and CS methods is considerable. This is beneficial for the identification of subjects who do not understand the elicitation method, but is also costly because these subjects do not produce a usable risk preferences. Inconsistency in the HL method is inversely explained by educational levels and numerical literacy, which implies improvements in its ease of use may produce more consistent risk preferences. This is especially important since the current abundance of inconstant subjects is problematic for widespread application. An additional concern, however, is the relatively high educational level and numerical literacy of my sample, which may suggest even higher shares of inconsistent answers for these methods with the population-wide application that is necessary for pensions. Finally, these results also extrapolate to other domains where risk preference are elicited. For instance, Dutch law also requires risk preferences elicitation in other industries, such as investing and insurance. These industries are similar to pensions and hence face many of the same problems in risk preference elicitation.

My research faces several limitations. Firstly, my experiment is not incentivised, Whether the elicitation methods use real or hypothetical incentives as well as their size affects risk preferences according to the literature.[38] But only pension funds can avoid the hypotheticality, since they are the ones that determine subjects' pension risk. Secondly, elicitation methods should use as realistic pension scenarios as possible, due the above incentive effects as well as the domain dependency of risk preferences. Since I only have five broad income categories based on dated information and I use rudimentary multiplication factors for the risk-return trade-off, incorporation of more (detailed) information can further improve accuracy. Thirdly, better risk preference modelling may improve elicitation. Instead of the EUT model with CRRA, more complicated modelling may account for e.g. time preferences, probability weighting, and loss aversion.

Future research could build on my findings and address some of their limitations. Pension funds could experiment with actual risk preference elicitation to make incentives real. They and academics could also test further customisation of the pension scenarios, with information such as subjects' current income, wealth, and time to retirement as well as expected investment risk and returns and inflation rates. Experiments may also include more extended modelling. Another interesting area for future research involves the consistency and stability of risk preferences. For inconsistent risk preferences, the relation with educational level and numerical literacy that I find for the HL method may pose an interesting lead. Unstable risk preferences, however, remain totally unexplained and hence also provide numerous roads of inquiry.

References

Abdellaoui, M., Driouchi, A., & l'Haridon, O. (2011). Risk aversion elicitation: reconciling tractability and bias minimization. *Theory and Decision 71*(1), 63-80.

Akay, A., Martinsson, P., Medhin, H., & Trautmann, S. (2012). Attitudes Toward Uncertainty Among the Poor: Evidence from Rural Ethiopia. Environment for Development Discussion Paper No. 1004.

Albert, S.M., & Duffy, J., (2012). Differences in risk aversion between young and older adults. *Neuroscience and Neuroeconomics 1*, 3–9.

Alserda, G., Dellaert, B., Swinkels, L., & van der Lecq, F. (2016). Pension risk preferences. Netspar Industry Series Design Paper No. 62

Andersen, S., Harrison, G.W., Lau, M. I., & Rutstrom, E.E. (2006). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383-405.

Andersen, S., Cox, J., Harrison, G., Lau, M., Rutstroem, E., & Sadiraj, V. (2011). Asset integration and attitudes to risk: Theory and evidence. Durham University Working Paper.

Anderson, L.R., & Mellor, J.M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measue. *Journal of Risk and Uncertainty 39*, 137-160.

Arrow, K.J., (1965). Aspects of the Theory of Risk Bearing. Yrjö Jahnssonin Säätiö, Helsinki.

Arrondel, L., & Calvo Pardo, H. F. (2002). Portfolio choice with a correlated background risk: Theory and evidence. DELTA Working Paper No. 200216.

Baltussen, G., Post, T., & Van den Assem, M.J. (2008). Risky choice and the relative size of stakes. National Bureau of Economic Research Working Paper No. 989242.

[38] Risk aversion increases when real incentives increase (Binswanger 1981; Kachelmeier and Shehata 1992; Holt and Laury 2002).[38] This relationship holds for small (Holt and Laury 2002) and larger incentives, of up to three times monthly income (Kachelmeier and Shehata 1992). With hypothetical incentives, however, incentive size does not affect risk preferences (Kachelmeier and Shehata 1992; Holt and Laury 2002; Reynaud and Couture 2010). Generally, hypothetical incentives generate similar risk preferences as low real incentives, irrespective of the size of the hypothetical incentives (Holt and Laury 2002). Although, under specific circumstances, risk aversion does increase with increasing hypothetical incentives, for instance, after a series of increasing real incentives (Binswanger 1980) or for the EG method (Reynaud and Couture 2010).

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics 11*(2), 122–133.
Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, *112*(2), 537-579.
Benartzi, S. & Thaler, R.H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*(1), 78-98.
Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiatatum Imperalas Petropolitanae 5*, 175-192.
Bernoulli, D. (1954). "Exposition of a new theory on the measurement of risk." Edited by L. Sommer. *Econometrica 22*(1), 23-36.
Binswanger, H.P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics 62*(3), 395-407.
Binswanger, H.P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural India. *The Economic Journal*, *91*(364), 867-890.
Blais, A.R., & Weber, E.U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making 1*(1).
Bosch-Domenech, A., & Silvestre, J. (2006). Reflections on gains and losses: A 2 × 2 × 7 experiment. *Journal of Risk and Uncertainty 33*, 217–235.
Bosch-Domenech, A., & Silvestre, J., (2010). Averting risk in the face of large losses: Bernoulli vs. Tversky and Kahneman. *Economics Letters 107*, 180–182.
Bruner, D. M. (2009). Changing the probability versus changing the reward. *Experimental Economics 12*(4), 367-385.
Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. *Neuropsychology Review 19*(1), 102-114.
Charness, G., & Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry 48*(1), 133-146.
Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization 83*(1), 50-58.
Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization 87*, 43-51.
Chakravarty, S., Harrison, G.W., Haruvy, E.E., & Rutstrom, E.E. (2011). Are you risk averse over other people's money? *Southern Economic Journal 77*, 901–913.
Cokely, E.T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012) Measuring risk literacy: The Berlin Numeracy Test. *Judgement and Decision Making 7*(1), 24-47.
Comeig, I., Holt, C. A., & Jaramillo-Gutiérrez, A. (2015). Dealing with risk: Gender, stakes, and probability effects. University of Valencia, ERI-CES, DPEB Discussion Papers in Economic Behaviour No. 0215.
Cox, J. C., & Sadiraj, V. (2006). Small-and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior 56*(1), 45-60.
Cox, R., Kamolsareeratana, A., & Kouwenberg, R. (2017). Gambling addiction in the stock market. Working paper.
Crosetto, P., & Filippin, A. (2013a). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty 47*(1), 31-65.
Crosetto, P., & Filippin, A. (2013b). A theoretical and experimental appraisal of five risk elicitation methods. Jena Economic Research Paper No. 2013009.
Crosetto, P., & Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics 19*, 613-641
Croson, R. & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448-474
Dave, D., & Saffer, H. (2007). Risk tolerance and alcohol demand among adults and older adults. National Bureau of Economic Research Working Paper No. 13482.
Dave, C., Eckel, C.C., Johnson, C.A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better?. *Journal of Risk and Uncertainty 41*(3), 219-243.
Deck, C., Lee, J., Reyes, J., & Rosen, C. (2010). Measuring risk aversion on multiple tasks: Can domain specific risk attitudes explain apparently inconsistent behavior?
Dellaert, B., Donkers, B., Turlings, M., Steenkamp, T., & Vermeulen, E. (2016). Naar een nieuwe aanpak voor risicoprofielmeting voor deelnemers in pensioenregelingen. Netspar Industry Series Design Paper No. 49.
Dreber, A., Rand, D.G., Wernerfelt, N., Garcia, J.R., Vilar, M.G., Lum, J.K., & Zeckhauser, R. (2011). Dopamine and risk choices in different domains: findings among serious tournament bridge players. *Journal of Risk and Uncertainty 43*(1), 19–38.

Dohmen, T.J., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G.G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. DIW Discussion Papers No. 511.

Dohmen, T.J., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G.G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522-550.

Doyle, J.R. (2013). Survey of time preference, delay discounting models. *Judgement and Decision Making 8*(2), 116-135.

Eckel, C.C., & Grossman, P.J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior 23*(4), 281-295.

Eckel, C.C., & Grossman, P.J. (2008a). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1-17.

Eckel, C.C., & Grossman, P.J. (2008b). Men, women and risk aversion: Experimental evidence. In *Handbook of Experimental Economics Results, Volume 1*. By Plott, C.R., & Smith, V.L. (Eds.). Elsevier, 1061-1073.

Eckel, C.C., Grossman, P.J., Johnson, C.A., de Oliveira, A.C.M., Rojas, C., & Wilson, R.K. School environment and risk preferences: Experimental evidence. *Journal of Risk and Uncertainty 45*(3), 265-292.

Edwards, A., Elwyn, G., & Mulley, A. (2002). Explaining risks: turning numerical data into meaningful pictures. *BMJ: British Medical Journal 324*(7341), 827.

Elston, J.A., Harrison, G.W. & Rutstrom, E.E. (2005). Characterizing the entrepreneur using field experiments. Working Paper, Max Planck Institute of Economics.

Filippin, A., & Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science 62*(11), 3138-3160.

Friedman, M. & Savage, L.J. (1948). The utility of choices involving risk. *Journal of Political Economy 56*(4), 279-304.

Fehr-Duda, H., De Gennaro, M., & Schubert, R. (2006). Gender, financial risk, and probability weights. *Theory and decision 60*(2-3), 283-313.

Galarza, F.B. (2009). Choices Under Risk in Rural Peru. Department of Agricultural and Applied Economics, University of Wisconsin, Staff Paper No. 542.

von Gaudecker, H.M., van Soest, A., & Wengstrom, E., (2011). Heterogeneity in risky choice behavior in a broad population. *The American Economic Review 101*(2), 664–694.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *BMJ: British Medical Journal 327*(7417), 741.

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts?. *Risk analysis 25*(3), 623-629.

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

Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica 77*, 1637–1664.

Goldstein, D.G., Johnson, E.J., & Sharpe, W.F. (2008). Choosing outcomes versus choosing products: Consumer-focused retirement investment advice. *Journal of Consumer Research 35*(3), 440-456.

Haidous, No.N., & Sawilowsky, S.S. (2013). Robustness and power of the Kornbrot rank difference, signed ranks, and dependent samples T-test. *American Journal of Applied Mathematics and Statistics 1*(5), 99-102.

Hammond, P.J., & Zank, H. (2014). "Chapter 2: Rationality and Dynamic Consistency Under Risk and Uncertainty." In *Handbook of the Economics of Risk and Uncertainity, Volume 1*. By M.J. Machine & W.K. Viscusi. Oxford: North-Holland, 99-133.

Harbough, W.T., Krause, K., & Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics* (5), 53-84.

Harrison, G.W., (1986). An experimental test for risk aversion. *Economics Letters 21*, 7–11.

Harrison, G. W., Lau, M.I., & Rutstrom, E.E. (2007). Estimating risk attitudes in Denmark: A field experiment. *The Scandinavian Journal of Economics 109*(2), 341-368.

He, H., Martinsson, P., Sutter, M., (2012). Group decision making under risk: an experiment with student couples. *Economics Letters 117*, 691–693.

Hess, R., Visschers, V.H., & Siegrist, M. (2011). Risk communication with pictographs: The role of numeracy and graph processing. *Judgment and Decision Making 6*(3), 263.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science 290*(5500), 2261-2262.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644-1655.

Holt, C. A., & Laury, S. K. (2014). “Chapter 4: Assessment and Estimation of Risk Preferences.” In *Handbook of the Economics of Risk and Uncertainity, Volume 1*. By M.J. Machine & W.K. Viscusi. Oxford: North-Holland, 99-133.
Holt, C.A., Smith, A., (2009). An update on Bayesian updating. *Journal of Economic Behavior and Organization* (69), 125–134.
Isaac, R. M., & James, D. (2000). Just who are you calling risk averse?. *Journal of Risk and Uncertainty 20*(2), 177-187.
Jacobson, S., Petrie, R., (2009). Learning from mistakes: what do inconsistent choices over risk tell us? *Journal of Risk and Uncertainty* 38 (2), 143–158.
Kachelmeier, S. J., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *The American Economic Review*, 1120-1141.
Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives 5*(1), 193-206.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.
Kapteyn, A., & Teppa, F. (2011). Subjective measures of risk aversion, fixed costs, and portfolio choice. *Journal of Economic Psychology*, *32*(4), 564-580.
Kornbrot, D.E. (1990). The rank difference test: A new and meaningful alternative to the Wilcoxon signed ranks test for ordinal data. *British Journal of Mathematical and Statistical Psychology 43*(2), 241-264.
Lahiri, K., & Song, J.G. (2000). The effect of smoking on health using a sequential self-selection model. *Health Economics 9*, 491–511.
Laury, S. (2005). Pay one or pay all: Random selection of one choice for payment. Andrew Young School of Policy Studies Research Paper Series No. 0613.
Laury, S.K., & Holt, C.A. (2008). Further reflections on the reflection effect. In *Risk Aversion in Experiments*. By Cox, J.C. & Harrison, G.W. (Eds.). Emerald Group Publishing Limited, 405-440.
Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., Strong, D.R., & Brown, R.A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied 8*(2), 75.
Lejuez, C.W., Aklin, W.M., Zvolensky, M.J., & Pedulla, C.M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence 26*(4), 475-479.
Lipkus, I.M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making 27*(5), 696-713.
Lipkus, I.M., & Hollands, J.G. (1999). The visual communication of risk. *JNCI monographs 25*, 149-163.
List, J.A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy 115*, 482–493.
Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review 90*(2), 426-432.
Loewenstein, G.F., Weber, E.U., Hsee, C.K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin 127*(2), 267.
Lusardi, A. (1998). On the importance of the precautionary saving motive. *The American Economic Review 88*(2), 449–453.
Lusardi, A., & Mitchell, O.S. (2007). Baby boomers retirement security: the role of planning, financial literacy and housing wealth. Journal of Monetary Economics 54, 205–224.
Lusardi, A., & Mitchell, O.S. (2011). Financial literacy and planning: Implications for retirement well-being. In *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*. By A. Lusardi and O.S. Mitchell (Eds.). Oxford University Press, 17-39.
Markowitz, H. (1952a). Portfolio selection. *The Journal of Finance 7*(1), 77-91.
Markowitz, H. (1952b). The utility of wealth. *Journal of Political Economy 60*(2), 151-158.
Meyer, J. (2014). “Chapter 3: The Theory of Risk and Risk Aversion.” In *Handbook of the Economics of Risk and Uncertainity, Volume 1*. By M.J. Machine & W.K. Viscusi. Oxford: North-Holland, 99-133.
Millner, E.L. & Pratt, M.D. (1991). Risk aversion and rent-seeking: An extension and some experimental evidence. *Public Choice 69*(1), 81-92.
Murningham, J.K., Roth, A.E., & Schoumaker, F. (1988). Risk aversion in bargaining: An experiment study. *Journal of Risk and Uncertainty 1*(1), 101-124.
Nelson, J.A. (2014). Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. *Journal of Economic Surveys* 29(3), 566–585.
Nelson, J.A. (2016). Not-so-strong evidence for gender differences in risk taking. *Feminist Economics 22*(2), 114-142.
von Neumann, J., & Morgenstern, O., 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

Porter, T.K., (2011). Stuffing the Mattress: An Experiment on Risk and Bank Runs. Senior Thesis, College of William and Mary.
Post, T., Van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? decision making under risk in a large-payoff game show. *The American Economic Review 98*(1), 38-71.
Pratt, J.W. (1964). Risk aversion in the small and in the large. *Econometrica 32*(1-2), 122-136.
Prelec, D. (1998). The probability weighting function. *Econometrica* 497-527.
Rosen, H.S., & Wu, S. (2004). Portfolio choice and health status. *Journal of Financial Economics 72(*3), 457–484.
Quiggin, J. (2014). "Chapter 12: Non-Expected Utility Models Under Objective Uncertainty" In *Handbook of the Economics of Risk and Uncertainity, Volume 1*. By M.J. Machine & W.K. Viscusi. Oxford: North-Holland, 701-728.
Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica 68*(5), 1281-1292.
Rabin, M., & Thaler, R. H. (2001). Anomalies: Risk aversion. *The Journal of Economic Perspectives 15*(1), 219-232.
Reynaud, A., & Couture, S. (2012). Stability of risk preference measures: Results from a field experiment on French farmers. *Theory and Decision 73*(2), 203-221.
van Rooij, M. C., Kool, C. J., & Prast, H. M. (2007). Risk-return preferences in the pension domain: Are people able to choose?. *Journal of Public Economics 91*(3), 701-722.
van Rooij, M. C., Lusardi, A., & Alessie, R. J. (2011a). Financial literacy and retirement planning in the Netherlands. *Journal of Economic Psychology 32*(4), 593-608.
van Rooij, M., Lusardi, A., & Alessie, R. (2011b). Financial literacy and stock market participation. *Journal of Financial Economics 101*, 449-472.
Skillings, J.H., & Mack, G.A. (1981). On the use of a Friedman-type statistic in balanced and unbalanced block designs. *Technometrics 23*, 171–177
Sprenger, C. (2010). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy 1*(2), 17.
Tanaka, T., Camerer, C., & Nguyen, Q., 2010. Risk and time preferences: Experimental and household data from vietnam. *The American Economic Review 100*(1), 557–571.
Tversky, A., & Kahneman, D. (1981). The fFraming of decisions and the psychology of choice. *Science 211*(4481), 453-458.
Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics 106*(4), 1039-1061.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297-323.
Weber, E.U., Blais, A.R., & Betz, N.E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making 15*(4), 263-290.

Appendix

A. Experiment texts in English and Dutch

A.1. Translation of the experiment text in English

----------- NEW PAGE -----------

----------- INTRODUCTION -----------

Some people like to take risks with their pension income, while other people like to avoid risks. Imagine that your pension fund asks you the following questions. Your answers will be subsequently used to determine the risk that is taken while investing your pension savings.

You are asked to choose between two pension products. Each pension product is summarised in a circle (see below). Each circle contains two pension incomes.

The pension incomes are based on a net monthly income of €[AMOUNT]

The pension income is the net amount you will receive monthly (including the public pension) when you are 75 years old. This is independent of the age at which you receive your first pension.

You will receive the high pension income of the circle in the positive scenario and the low pension income in the circle in the negative scenario. The black part of the circle is the positive scenario and the white part of the circle is the negative scenario. The chances of the positive and negative scenarios are indicated in percentages and by the size of the black and white parts of the circle. A 10%chance means that the scenario occurs 1e out of 10 times. A 50% chance means that the scenario occurs 5 out of 10 times.

----------- NEW PAGE -----------

----------- TEST QUESTION -----------

Below you see two example pension products.

The pension income is the net amount you will receive monthly (including the public pension) when you are 75 years old.

[TWO EXAMPLARY PENSION PRODUCTS IN PIE CHARTS]

Imagine you have chosen the second pension product. In case of the positive scenario, how much monthly pension income will you receive when you are 75 years old?

1. €[AMOUNT]
2. €[AMOUNT]
3. €[AMOUNT]
4. €[AMOUNT]

----------- NEW PAGE -----------

You answered the test question [CORRECTLY/INCORRECTLY]. In case you choose the second pension product – when you are 75 years old – you will have a 50% chance to receive €[AMOUNT] (in case of the positive scenario) and a 50% chance to receive €[AMOUNT] (in case of the negative scenario).

If you choose the first pension product – when you are 75 years old – you will have a 50% chance to receive €[AMOUNT] (in case of the positive scenario) and a 50% chance to receive €[AMOUNT] (in case of the negative scenario).

[TWO EXAMPLARY PENSION PRODUCTS IN PIE CHARTS]

----------- NEW PAGE -----------

----------- THE EG METOD -----------

Below you are asked to choose between ten different pension products. You can only choose one of the ten pension products.

All pension products have different pension incomes in both scenarios. The chances of the positive and negative scenario are always equal (both 50%).

The pension income is the net amount you will receive monthly (including the public pension) when you are 75 years old.

There are no right or wrong choices. It is solely about your own preferences.

Which of the ten pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART]
B. [PENSION PRODUCT IN PIE CHART]
C. [PENSION PRODUCT IN PIE CHART]
D. [PENSION PRODUCT IN PIE CHART]
E. [PENSION PRODUCT IN PIE CHART]
F. [PENSION PRODUCT IN PIE CHART]
G. [PENSION PRODUCT IN PIE CHART]
H. [PENSION PRODUCT IN PIE CHART]
I. [PENSION PRODUCT IN PIE CHART]
J. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

----------- THE HL METOD -----------

You are now asked to choose between two different pension products several times. You can only choose one of the two pension products, pension product A or B.

The pension incomes remain the same in every question. The chance of the positive scenario and the chance of the negative scenario change every question; the chance of the positive scenario gradually increases.

The pension income is the net amount you will receive monthly (including the public pension) when you are 75 years old.

There are no right or wrong choices. It is solely about your own preferences.

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?
A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

----------- THE CS METOD -----------

You are now asked to choose between two different pension products several times. You can only choose one of the two pension products, pension product A or B.

All pension products have different pension incomes in both scenarios and these change for every question. The chances of the positive and negative scenario are always equal (both 50%).

The pension income is the net amount you will receive monthly (including the public pension) when you are 75 years old.

There are no right or wrong choices. It is solely about your own preferences.

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

Which of the two pension products below do you prefer?

A. [PENSION PRODUCT IN PIE CHART] B. [PENSION PRODUCT IN PIE CHART]

----------- NEW PAGE -----------

----------- COMPLEXITY OF THE METODS -----------

You have chosen between pension products in three different ways. Below you can indicate how difficult or easy you found the different methods.

Can you indicate how you found method 1 on a scale from 1 (very difficult) to 10 (very easy)?

Method 1: you choose once between ten pension incomes
[ILLUSTRATION OF THE FIRST METHOD]

1	2	3	4	5	6	7	8	9	10

Very difficult Very easy

And how did you find method 2 on a scale from 1 (very difficult) to 10 (very easy)?

Method 2: you choose several times between two pension incomes
[ILLUSTRATION OF THE SECOND METHOD]

1	2	3	4	5	6	7	8	9	10

Very difficult Very easy

And how did you find method 3 on a scale from 1 (very difficult) to 10 (very easy)?

Method 1: you choose several times between two pension incomes, one after another
[ILLUSTRATION OF THE THIRD METHOD]

1	2	3	4	5	6	7	8	9	10

Very difficult — Very easy

----------- NEW PAGE -----------

----------- THE SOEP QUESTIONS -----------

How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

Please tick a box on the scale, where the value 0 means: 'unwilling to take risks' and the value 10 means: 'fully prepared to take risk'.

1	2	3	4	5	6	7	8	9	10

Unwilling to take risks — Fully prepared to take risks

In different domains, you can behave differently. How do you see your risk preference in the following domains?

How is it in financial matters?

1	2	3	4	5	6	7	8	9	10

Unwilling to take risks — Fully prepared to take risks

How is it for you pension?

1	2	3	4	5	6	7	8	9	10

Unwilling to take risks — Fully prepared to take risks

----------- NEW PAGE -----------

----------- THE DOSPERT SCALE -----------

For each statement asked, you must indicate the likelihood that you take part in the activity specified or you adopt the behaviour specified if you were in the situation described. You must then choose one of the 10 options ranging from "Extremely unlikely" to "Extremely likely" by using the scale.

1. Betting a day's income at the horse races.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

2. Investing 10% of your annual income in a moderate growth diversified fund.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

3. Betting a day's income at a high-stake poker game.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

4. Investing 5% of your annual income in a very speculative stock.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

5. Betting a day's income on the outcome of a sporting event.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

6. Investing 10% of your annual income in a new business venture.

1	2	3	4	5	6	7	8	9	10

Extremely unlikely — Extremely likely

----------- NEW PAGE -----------

----------- THE CONVENTIONAL METHODS -----------

De value of your pension income depends on the investing results of your pension fund. These can vary and your pension income can thus decrease.

How would you react when your expected monthly pension income would see a sudden significant decrease?
1. I would lose sleep over it
2. I would dislike it a lot
3. I would be disappointed, but I know it is a possibility
4. I would not lose any sleep over it

Which decrease of your expected monthly pension income would you find acceptable?
1. No decrease
2. A decrease of 3%
3. A decrease of 10%
4. A decrease of 15%
5. A decrease of 25%
6. A decrease of more than 25%

----------- NEW PAGE -----------

----------- NUMERICAL LITERACY -----------

Please answer the questions below. Do not use a calculator but feel free to use the space available for notes (i.e., scratch paper).

Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? ________ out of 50 throws.

----------- NEW PAGE -----------

Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? (please indicate the probability in percent). ___________ %

----------- NEW PAGE -----------

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, how many times would the die show the number 6? ___________ out of 70 throws.

----------- NEW PAGE -----------

In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red? _____________%

----------- NEW PAGE -----------

----------- FINANCIAL LITERACY -----------

The coming questions are about knowledge of financial matters. Do not look anything up and do not use a calculator for these questions. They are about your own knowledge.

Imagine you have 100 euro on a savings account and 20% interest annually and you never withdraw any money or interest. How much money would be on the savings account after five years?
1. More than 200 euro
2. Exactly 200 euro
3. Less than 200 euro
4. I do not know

Imagine that the interest on your savings account is 1% annually and inflation is 2% annually. After 1 year, would you be able to buy more, exactly the same, or less than today with the money on your savings account?
1. More than today
2. Exactly the same as today
3. Less than today
4. I do not know

Please indicate fort the following 5 questions wat you think.

1. A share in a company normally gives a less risky return than an investment fund that invests in shares.

1	2	3
True	Not true	I do not know

2. If you invest for the long term, the annual management costs of investment funds are not important.

1	2	3
True	Not true	I do not know

3. To make a return in the stock market, you must not buy and sell share too often.

1	2	3
True	Not true	I do not know

4. When you buy an investment product with leverage (for example, a turbo, sprinter, or speeder), you invest with borrowed money and effectively do not pay interest.

1	2	3
True	Not true	I do not know

5. When you buy a leveraged investment product for 1,000 euro (for example, a turbo, sprinter, or speeder) of which the value is derived from the AEX Index, you have a smaller chance to lose your entire investment than when you invest 1,000 euro in an AEX Index fund.

1	2	3
True	Not true	I do not know

How many of the previous 7 knowledge question do you think you have answered correctly? ....

----------- END -----------

A.2. Original experiment text in Dutch

----------- NIEUWE PAGINA -----------

----------- INTRODUCTIE -----------

Sommige mensen nemen graag veel risico met hun pensioeninkomen terwijl andere mensen graag risico mijden. Stelt u zich voor dat uw pensioenfonds u de volgende vragen voorlegt. Uw antwoorden worden vervolgens gebruikt om het risico waarmee uw pensioenvermogen wordt belegd te bepalen.

U wordt steeds gevraagd een keuze te maken tussen verschillende pensioenenproducten.
Elk pensioenproduct is samengevat in een cirkel (zie hieronder). In iedere cirkel staan twee pensioeninkomens.

De pensioeninkomens zijn gebaseerd op een netto maandinkomen van €[BEDRAG]

Het pensioeninkomen is het bedrag dat u netto maandelijks ontvangt (inclusief AOW) als u 75 jaar oud bent. Dit is ongeacht op welke leeftijd u voor het eerst pensioen ontvangt.

Het hoge pensioeninkomen bij de cirkel ontvangt u in het positieve scenario en het lage pensioeninkomen bij de cirkel ontvangt u in het negatieve scenario. Het zwarte deel van de cirkel is het positieve scenario en het witte deel van de cirkel is het negatieve scenario. Hoe groot de kans is op het positieve of negatieve scenario, is aangegeven in procenten en de grootte van het vlak in de cirkel. In de afbeelding hieronder zit u een voorbeeld van twee pensioenproducten samengevat in cirkels. Een kans van 10% betekent dat het scenario in 1 van de 10 gevallen plaatsvindt. Een kans van 50% betekent dat het scenario in 5 van de 10 gevallen plaatsvindt.

----------- NIEUWE PAGINA -----------

----------- TOETSVRAAG -----------

Hieronder ziet u een voorbeeld van twee pensioenproducten.

Het pensioeninkomen is het bedrag dat u netto maandelijks ontvangt (inclusief AOW) als u 75 jaar oud bent.

[TWEE VOORBEELD PENSIONPRODUCTEN IN SCHIJFDIAGRAMMEN]

Stelt u voor dat u het tweede pensioenproduct hebt gekozen. Als het positieve scenario zich voordoet, hoeveel maandelijks pensioeninkomen ontvangt u dan als u 75 jaar oud bent?

1. [BEDRAG]
2. [BEDRAG]
3. [BEDRAG]
4. [BEDRAG]

----------- NIEUWE PAGINA -----------

U heeft de testvraag [GOED/FOUT] beantwoord. Als u voor pensioen B kiest, heeft u – wanneer u 75 jaar oud bent – 50% kans op €[BEDRAG] (positief scenario) en 50% kans op €[BEDRAG] (negatief scenario).

Kiest u voor pensioen A, dan heeft u – wanneer u 75 jaar oud bent – 50% kans op €[BEDRAG] (positief scenario) en 50% kans op €[BEDRAG] (negatief scenario).

[TWEE VOORBEELD PENSIONPRODUCTEN IN SCHIJFDIAGRAMMEN]

----------- NIEUWE PAGINA -----------

----------- DE "EG" METHODE -----------

Hieronder vragen wij u te kiezen tussen tien verschillende pensioenproducten. U kunt slechts één van de tien pensioenproducten kiezen.

Alle pensioenproducten hebben verschillende pensioeninkomens in beide scenario's. De kans op het positieve of negatieve scenario is steeds gelijk (beide 50%).

Het pensioeninkomen is het bedrag dat u netto maandelijks ontvangt (inclusief AOW) als u 75 jaar oud bent.

Er zijn geen goede of foute keuzen. Het gaat uitsluitend om uw eigen voorkeur.

Welk van de tien onderstaande pensioenproducten heeft uw voorkeur?

A. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
C. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
D. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
E. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
F. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
G. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
H. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
I. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
J. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

----------- DE "HL" METHODE -----------

U wordt nu een aantal keer gevraagd te kiezen tussen twee verschillende pensioenproducten. U kunt per vraag steeds één van de twee pensioenproducten kiezen, pensioenproduct A of B.

De pensioeninkomens blijven iedere vraag gelijk. De kans op het positieve scenario en de kans op het negatieve scenario veranderen bij iedere vraag; de kans op het positieve scenario wordt alsmaar groter.

Het pensioeninkomen is het bedrag dat u netto maandelijks ontvangt (inclusief AOW) als u 75 jaar oud bent.

Er zijn geen goede of foute keuzen. Het gaat uitsluitend om uw eigen voorkeur.

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?

A. [PENSIONPRODUCT IN SCHIJFDIAGRAM]
B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

----------- DE “CS” METHODE -----------

U wordt nu een aantal keer gevraagd te kiezen tussen twee verschillende pensioenproducten. U kunt per vraag steeds één van de twee pensioenproduct kiezen, pensioenproduct A of B.

Alle pensioenproducten hebben verschillende pensioeninkomens in beide scenario’s en deze veranderen bij iedere vraag. De kans op het positieve of negatieve scenario is steeds gelijk (beide 50%).

Het pensioeninkomen is het bedrag dat u netto maandelijks ontvangt (inclusief AOW) als u 75 jaar oud bent.

Er zijn geen goede of foute keuzen. Het gaat uitsluitend om uw eigen voorkeur.

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

Welk van de twee onderstaande pensioenproducten heeft uw voorkeur?
A. [PENSIONPRODUCT IN SCHIJFDIAGRAM] B. [PENSIONPRODUCT IN SCHIJFDIAGRAM]

----------- NIEUWE PAGINA -----------

----------- MOEILIJKHEID VAN DE METHODES -----------

U heeft op drie verschillende manieren keuzes gemaakt tussen pensioeninkomens. Geef hieronder aan hoe ingewikkeld of simpel u de verschillende methoden vond.

Kunt u aangeven hoe u methode 1 heeft ervaren op een schaal van 1 (heel ingewikkeld) tot en met 10 (heel simpel)?

Methode 1: U kiest eenmalig tussen 10 pensioeninkomens
[ILLUSTRATIE VAN DE EERSTE METHODE]

1	2	3	4	5	6	7	8	9	10

Heel ingewikkeld Heel simpel

En hoe heeft u methode 2 ervaren op een schaal van 1 (heel ingewikkeld) tot en met 10 (heel simpel)?

Methode 2: U kiest een aantal keer tussen 2 pensioeninkomens
[ILLUSTRATIE VAN DE TWEEDE METHODE]

1	2	3	4	5	6	7	8	9	10

Heel ingewikkeld Heel simpel

En hoe heeft u methode 3 ervaren op een schaal van 1 (heel ingewikkeld) tot en met 10 (heel simpel)?

Methode 3 : U kiest een aantal keer tussen 2 pensioeninkomens achter elkaar
[ILLUSTRATIE VAN DE DERDE METHODE]

1	2	3	4	5	6	7	8	9	10

Heel ingewikkeld Heel simpel

----------- NIEUWE PAGINA -----------

----------- DE SOEP VRAGEN -----------

Personen verschillen in hun risicobereidheid. Sommige personen zijn zeer risicobereid en andere zijn zeer risicomijdend.

Kunt u aangeven hoe risicobereid u in het algemeen bent op een schaal van 1 (zeer risicomijdend) tot en met 10 (zeer risicobereid)?

1	2	3	4	5	6	7	8	9	10

Zeer risicomijdend Zeer risiobereid

Mensen kunnen meer of minder risicobereid zijn afhankelijk van het onderwerp.

Kunt u aangeven hoe risicobereid u bent op een schaal van 1 (zeer risicomijdend) tot en met 10 (zeer risicobereid) op de volgende gebieden?

1. Financiële zaken

1	2	3	4	5	6	7	8	9	10

Zeer risicomijdend Zeer risiobereid

2. Pensioenzaken

1	2	3	4	5	6	7	8	9	10

Zeer risicomijdend Zeer risiobereid

----------- NIEUWE PAGINA -----------

----------- DE DOSPERT SCHAAL -----------

Hoe waarschijnlijk is het dat u onderstaande activiteit zou uitvoeren wanneer de kans zich zou voordoen?

Geef op een schaal van 1 (zeer onwaarschijnlijk) tot en met 10 (zeer waarschijnlijk) antwoord op de volgende vragen.

1. Een inkomen van een dag inzetten bij paardenraces

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

2. 10% van uw jaarinkomen investeren in een middelmatig groeiende beleggingsmaatschappij

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

3. Een inkomen van een dag inzetten bij een pokerspel (kaartspel)

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

4. 5% van uw jaarinkomen investeren in zeer speculatieve aandelen

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

5. Een inkomen van een dag inzetten op de uitslag van een sportevenement (zoals bijvoorbeeld voetbal of basketbal)

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

6. 10% van uw jaarinkomen investeren in een nieuw bedrijf

1	2	3	4	5	6	7	8	9	10

Zeer onwaarschijnlijk Zeer waarschijnlijk

----------- NIEUWE PAGINA -----------

----------- DE “CONVENTIONELE” METHODEN -----------

De waarde van uw pensioeninkomen hangt af van de beleggingsresultaten van uw pensioenfonds. Deze kunnen schommelen en uw pensioeninkomen kan dan ook minder worden.

Hoe zou u reageren wanneer uw verwachte maandelijkse pensioeninkomen te maken zou krijgen met plotselinge forse dalingen?

1. Ik zou er slapeloze nachten van krijgen
2. Ik vind het heel vervelend
3. Ik vind het jammer, maar ik weet dat dit kan gebeuren
4. Ik slaap net zo lekker als anders

Welke daling van uw verwachte maandelijks pensioeninkomen vindt u acceptabel?

1. Geen daling
2. Een daling van 3%
3. Een daling van 10%
4. Een daling van 15%
5. Een daling van 25%
6. Een daling van meer dan 25%

----------- NIEUWE PAGINA -----------

----------- NUMMERIEKE GELETTERDHEID -----------

Tot slot vragen we u om enkele raadsels op te lossen. Gebruik alstublieft geen rekenmachine, maar gebruik pen en papier als u wilt.

In een klein dorp met 1000 inwoners zijn er 500 inwoners lid van een koor. Van deze 500 leden zijn er 100 mannelijk. Van de 500 inwoners die geen lid zijn van het koor, zijn er 300 mannelijk. Als ik willekeurig een man kies uit dit dorp, wat is dan de kans dat hij lid is van het koor? Geef deze kans aan in procenten. .... %

----------- NIEUWE PAGINA -----------

Beeldt u in dat we een vervalste dobbelsteen gooien (6 zijden). De kans dat u met deze dobbelsteen een zes gooit is twee keer zo hoog als de kans op elk ander getal op zich. Gemiddeld gezien, hoe vaak verwacht u dat u met deze dobbelsteen een zes gooit als u 70 keer gooit? .... van de 70 worpen

----------- OF -----------

Beeldt u in dat we 50 keer een vijfzijdige dobbelsteen werpen. Hoe vaak verwacht u dat we een oneven getal (1, 3, of 5) krijgen als resultaat van deze 50 worpen? .... van de 50 worpen

----------- EN EVENTUEEL -----------

In een bos zijn 20% van de paddenstoelen rood, 50% bruin, en 30% wit. Een rode paddenstoel is giftig met een kans van 20%. Een paddenstoel die niet rood is, is giftig met een kans van 5%. Wat is de kans dat een giftige paddenstoel in het bos rood is? .... %

----------- NIEUWE PAGINA -----------

----------- FINANCIELE GELETTERDHEID -----------

De volgende vragen gaan over kennis van geldzaken en planning van bestedingen. Zoekt u bij deze vragen niets op en gebruikt u geen rekenmachine. Het gaat steeds om uw eerste idee.

Veronderstel dat u 100 euro op een spaarrekening hebt en de rente is 20% per jaar en u neemt nooit geld of rente op. Hoeveel zou u dan na vijf jaar in het totaal op de rekening hebben?

1. Meer dan 200 euro
2. Precies 200 euro
3. Minder dan 200 euro
4. Ik weet het niet

Veronderstel dat de rente op uw spaarrekening 1% per jaar is en de inflatie is gelijk aan 2% per jaar. Zou u dan na 1 jaar meer, precies hetzelfde of minder kunnen kopen dan vandaag met het geld op de rekening?

1. Meer dan vandaag
2. Precies hetzelfde als vandaag
3. Minder dan vandaag
4. Ik weet het niet

Kunt u voor de volgende 5 vragen aangeven wat u hiervan vindt?

1. Een aandeel van een bedrijf geeft normaal gesproken een minder risicovol rendement dan een beleggingsfonds dat in aandelen belegt.

1	2	3
Waar	Niet waar	Ik weet het niet

2. Als je voor de lange termijn belegt, dan zijn de jaarlijkse beheerkosten van beleggingsfondsen niet belangrijk.

1	2	3
Waar	Niet waar	Ik weet het niet

3. Om rendement te maken in de aandelenmarkt, moet je aandelen niet te vaak aan- en verkopen.

1	2	3
Waar	Niet waar	Ik weet het niet

4. Wanneer u een beleggingsproduct met een hefboom koopt (bijvoorbeeld een turbo, sprinter of speeder), dan belegt u met geleend geld en betaalt u effectief rente.

1	2	3
Waar	Niet waar	Ik weet het niet

5. Wanneer u voor 1.000 euro een hefboomproduct koopt (bijvoorbeeld een turbo, sprinter of speeder) waarvan de waarde afhangt van de stand van de AEX Index, dan heeft u een kleinere kans om uw gehele inleg kwijt te raken dan wanneer u 1.000 euro belegt in een AEX Indexfonds.

1	2	3
Waar	Niet waar	Ik weet het niet

Hoeveel van de voorgaande 7 kennisvragen denkt u correct te hebben beantwoord? ....

----------- EINDE -----------

A.3. The SOEP questions in the original German version

Wie schätzen Sie sich persönlich ein: Sind Sie im allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?

Bitte kreuzen Sie ein Kästchen auf der Skala an, wobei der Wert 0 bedeutet: "gar nicht risikobereit" und der Wert 10: "sehr risikobereit". Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Man kann sich in verschiedenen Bereichen ja auch unterschiedlich verhalten. Wie wurden Sie Ihre Risikobereitschaft in Bezug auf die folgenden Bereiche einschätzen?

Bitte kreuzen Sie in jeder Zeile ein Kästchen auf der Skala an!

Wie ist das beim Autofahren?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Wie ist das bei Geldanlangen?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Wie ist das bei Freizeit und Sport?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Wie ist das bei Ihrer beruflichen Karriere?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Wie ist das bei Ihrer Gesundheit?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

Wie ist das beim Vertrauen in fremde Menschen?

0	1	2	3	4	5	6	7	8	9	10

Gar nicht risikobereit — Sehr risikobereit

B. Overview of the pension incomes per income category

TABLE A.1 Multiplication factors for incomes and incomes of the first income category

Method	Pension	Multiplication factors		Income category 1 (€1,960)	
		Positive scenario	Negative scenario	Positive scenario	Negative scenario
The EG method					
	A	0.65	0.65	€1,270	€1,270
	B	0.6878	0.62	€1,350	€1,210
	C	0.7354	0.59	€1,440	€1,160
	D	0.7844	0.56	€1,540	€1,100
	E	0.8262	0.53	€1,620	€1,040
	F	0.8562	0.50	€1,680	€980
	G	0.8744	0.47	€1,710	€920
	H	0.8835	0.44	€1,730	€860
	I	0.8868	0.41	€1,740	€800
	J	0.8874	0.38	€1,740	€740
The HL method					
	A	0.7	0.6	€1,370	€1,170
	B	0.9	0.4	€1,760	€780
The CS method					
		Uses the EG method's pensions		*Uses the EG method's pensions*	
Introduction					
		0.75	0.55	€1,470	€1,080
Test question					
	A	0.75	0.55	€1,470	€1,080
	B	0.85	0.45	€1,660	€880

The multiplication factors are rounded to four decimals whenever necessary. The amounts are rounded to tens.

TABLE A.2 Incomes of the second and third income categories					
Method	Pension	Income category 2 (€2,240)		Income category 3 (€2,940)	
		Positive scenario	Negative scenario	Positive scenario	Negative scenario
The EG method					
	A	€1,460	€1,460	€1,910	€1,910
	B	€1,540	€1,390	€2,020	€1,820
	C	€1,650	€1,320	€2,160	€1,730
	D	€1,760	€1,250	€2,310	€1,650
	E	€1,850	€1,190	€2,430	€1,560
	F	€1,920	€1,120	€2,520	€1,470
	G	€1,960	€1,050	€2,570	€1,380
	H	€1,980	€990	€2,600	€1,290
	I	€1,990	€920	€2,610	€1,210
	J	€1,990	€850	€2,610	€1,120
The HL method					
	A	€1,570	€1,340	€2,060	€1,760
	B	€2,020	€900	€2,650	€1,180
The CS method					
		Uses the EG method's pensions		*Uses the EG method's pensions*	
Introduction					
		€1,680	€1,230	€2,210	€1,620
Test question					
	A	€1,680	€1,230	€2,210	€1,620
	B	€1,900	€1,010	€2,500	€1,320

All amounts are rounded to tens.

TABLE A.3 Incomes of the fourth and fifth income categories					
Method	Pension	Income category 4 (€3,740)		Income category 5 (€4,830)	
		Positive scenario	Negative scenario	Positive scenario	Negative scenario
The EG method					
	A	€2,430	€2,430	€3,140	€3,140
	B	€2,570	€2,320	€3,320	€2,990
	C	€2,750	€2,200	€3,550	€2,850
	D	€2,930	€2,090	€3,790	€2,700
	E	€3,090	€1,980	€3,990	€2,560
	F	€3,200	€1,870	€4,130	€2,410
	G	€3,270	€1,760	€4,220	€2,270
	H	€3,300	€1,640	€4,270	€2,120
	I	€3,310	€1,530	€4,280	€1,980
	J	€3,310	€1,420	€4,280	€1,830
The HL method					
	A	€2,620	€2,240	€3,380	€2,900
	B	€3,360	€1,490	€4,350	€1,930
The CS method					
		Uses the EG method's pensions		*Uses the EG method's pensions*	
Introduction					
		€2,800	€2,050	€3,620	€2,660
Test question					
	A	€2,800	€2,050	€3,620	€2,660
	B	€3,180	€1,680	€4,100	€2,170

All amounts are rounded to tens.

C. Original designs of the risk preference elicitation methods and numeracy test

TABLE A.4 Overview of the original EG method design

Option	Probability	Payoff	Prob.	Payoff	EV	CRRA range
A	50%	€16	50%	€16	€16	2.00 < *r* < ∞
B	50%	€24	50%	€12	€18	0.67 < *r* > 2.00
C	50%	€32	50%	€8	€20	0.38 < *r* > 0.67
D	50%	€40	50%	€4	€22	0.20 < *r* > 0.38
J	50%	€448	50%	€0	€24	-∞ < *r* < 0.20

The original EG method comes from Eckel and Grossman (2002). Payoffs were originally in dollars, but are one-on-one transferred to euros for uniformity. All CRRA ranges are based on Holt and Laury (2014), because Eckel and Grossman (2002) do not classify them.

TABLE A.5 Overview of the original HL method design

	Option A				Option B				EV(A)-	CRRA range
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	EV(B)	
1	10%	€2.00	90%	€1.60	10%	€3.85	90%	€0.10	€1.17	-∞ < *r* < -1.71
2	20%	€2.00	80%	€1.60	20%	€3.85	80%	€0.10	€0.83	-∞ < *r* < -0.95
3	30%	€2.00	70%	€1.60	30%	€3.85	70%	€0.10	€0.50	-0.95 < *r* < -0.49
4	40%	€2.00	60%	€1.60	40%	€3.85	60%	€0.10	€0.16	-0.49 < *r* < -0.15
5	50%	€2.00	50%	€1.60	50%	€3.85	50%	€0.10	-€0.18	-0.15 < *r* < 0.15
6	60%	€2.00	40%	€1.60	60%	€3.85	40%	€0.10	-€0.51	0.15 < *r* < 0.41
7	70%	€2.00	30%	€1.60	70%	€3.85	30%	€0.10	-€0.85	0.41 < *r* < 0.68
8	80%	€2.00	20%	€1.60	80%	€3.85	20%	€0.10	-€1.18	0.68 < *r* < 0.97
9	90%	€2.00	10%	€1.60	90%	€3.85	10%	€0.10	-€1.52	0.94 < *r* < 1.37
10	100%	€2.00	0%	€1.60	100%	€3.85	0%	€0.10	-€1.85	1.37 < *r* < ∞

The original HL method comes from Holt and Laury (2002). Payoffs were originally in dollars, but are one-on-one transferred to euros for uniformity. The first CRRA range is based on Dave et al. (2010), because Holt and Laury (2002) do not classify it.

TABLE A.6 Overview of the original CS method design

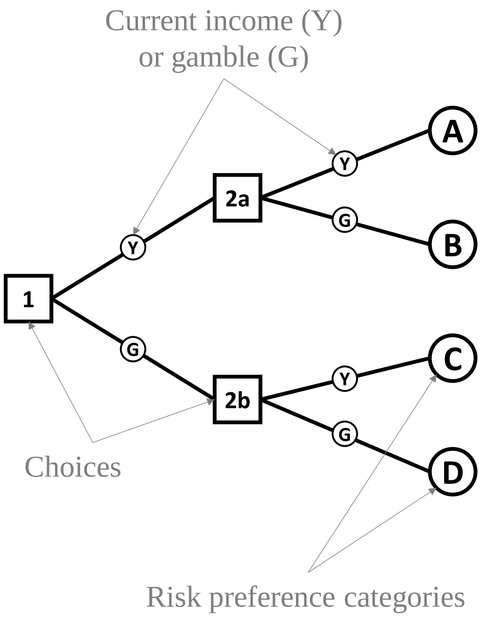
	Option A		Option B				Next step‡	CRRA range
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff		
1	100%	1 * Y†	50%	2 * Y	50%	$^{2}/_{3}$ * Y	2a / 2b	n/a
2a	100%	1 * Y	50%	2 * Y	50%	$^{4}/_{5}$ * Y	A	3.76 < *r* < ∞
							B	2.00 < *r* < 3.76
2b	100%	1 * Y	50%	2 * Y	50%	$^{1}/_{2}$ * Y	C	1.00 < *r* < 2.00
							D	-∞ < *r* < 1.00

† Subjects are asked to imagine different multiplication of their own current income (Y) to represent the payoffs.

‡ The next step is either the follow-up question or the final risk preference category (with the letters corresponding to very risk averse (A), moderately risk averse (B), slightly risk averse (C), and risk neutral to risk loving (D)). Left of the slash-sign is the next step for subjects who choose option A and right of the slash-sign for option B.

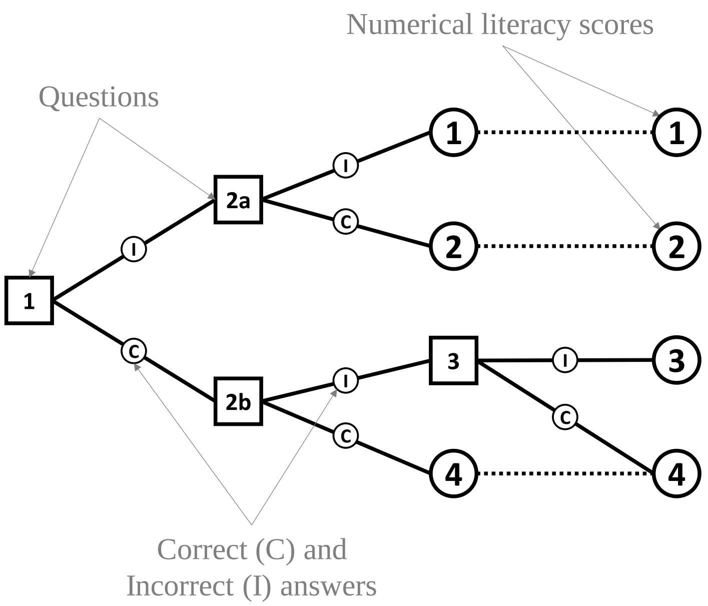
The original CS method comes from Barsky et al. (1997).

FIGURE A.1 Decision tree of the original CS method design



The original CS method comes from Barsky et al. (1997). Subjects are asked to between their own current income (Y) or a gamble (G) with multiplication of this income. Based on question 1 and follow-up question 2a or 2b, they are categorised according to risk preference in very risk averse (A), moderately risk averse (B), slightly risk averse (C), and risk neutral to risk loving (D).

FIGURE A.2 Decision tree for the Berlin Numeracy Test



The Berlin Numeracy test is created by Cokely et al. (2012). Subjects are asked a number of numerical questions. Based on their performance they are categorised according to numerical literacy in in four quartiles (where category 1 is the least numerically literate and category 4 the most).

D. Additional experimental results and robustness checks

TABLE A.7 Spearman correlations of the risk preference elicitation methods

	EG	HL	CS
EG	1		
	(426)		
HL	0.4686***	1	
	(321)	(321)	
CS	0.6055***	0.5256***	1
	(294)	(224)	(294)

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level) for the correlation's difference from zero. Sample size is indicated in brackets below the correlations

TABLE A.8 Pearson correlations of the risk preference elicitation methods and questionnaires

	EG	HL	CS
DOSPERT	0.1361***	0.1464***	0.1031*
SOEP (pension)	0.2309***	0.2875***	0.3617***
N	426	321	294

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level) for the correlation's difference from zero.

TABLE A.9 Increases in adjusted R^2 due to the inclusion of the questionnaires: consistent sample†

	EG	HL	CS
DOSPERT	4.66	3.57	2.68
SOEP (pension)	6.29	6.88	13.98
N	214	214	214

† The sample consists of consistent-only subjects, i.e. subjects who are consistent in both the HL and CS methods.

TABLE A.10 Pearson correlations of the risk preference elicitation methods and questionnaires

	EG	HL	CS	Conventional I	Conventional II
DOSPERT	0.1361***	0.1464***	0.1031*	0.2500***	0.2880***
SOEP (pension)	0.2309***	0.2875***	0.3617***	0.2988***	0.3181***
SOEP (financial)	0.1373***	0.2223***	0.3112***	0.3386***	0.3890***
SOEP (general)	0.1642***	0.2278***	0.2843***	0.2890***	0.3130***
Conventional I	0.0226	0.0304	0.0892	1	
Conventional II	0.0015	0.0070	0.0397	0.4523***	1
N	426	321	294	426	426

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

TABLE A.11 Logistic regressions for inconsistency in the HL and CS methods

	Logit I	Logit II	Logit III	Logit IV
Female	0.1331	0.0035	0.0033	-0.0803
Age group				
2	0.5699*	0.5697*	0.5972*	0.6231*
3	0.3539	0.3101	0.3197	0.2999
4	0.3211	0.2263	0.3413	0.2730
Education				
2	-1.4137***	-1.4874**	-1.8119***	-1.7568***
3	-1.5946***	-1.5004**	-1.8075***	-1.7230***
Income				
2	0.1337	0.0713	0.1636	0.1334
3	-0.2440	-0.2581	-0.3698	-0.3479
4	-0.5207	-0.5319	-0.5368	-0.5450
5	-0.2727	-0.2486	-0.3742	-0.3340
Wealth				
2	-0.1791	-0.1293	-0.4348	-0.4028
3	-0.1269	-0.0591	-0.1308	-0.0952
4	-0.5884	-0.5158	-0.3836	-0.3758
5	-0.3937	-0.3244	-0.3308	-0.2651
6	-0.2178	-0.1666	-0.2997	-0.2945
Numerical literacy		-0.2313**		-0.2197**
Financial literacy			-0.0243	0.0069
Constant	1.4906**	1.9931***	1.9463**	2.2484***
Pseudo R^2	4.02%	2.91%	4.53%	5.44%
N	403	403	347	347

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

TABLE A.12 Logistic regressions for inconsistency in the CS method

	Logit I	Logit II	Logit III	Logit IV
Female	-0.1438	-0.1755	-0.3309	-0.3494
Age group				
2	0.2624	0.2593	0.1820	0.1834
3	-0.1535	-0.1684	-0.1858	-0.1948
4	-0.3572	-0.3819	-0.3285	-0.3457
Education				
2	-0.2691	-0.2546	-0.1129	-0.0910
3	-0.1610	-0.1295	-0.1457	-0.1162
Income				
2	-0.1564	-0.1701	-0.3606	-0.3368
3	-0.6611	-0.6618	-0.7849*	-0.7779
4	-1.2127**	-1.2159**	-1.1967**	-1.1974**
5	-0.6217	-0.6171	-0.7904	-0.7821
Wealth				
2	0.0988	0.1142	-0.0190	-0.0075
3	0.4303	0.4518	0.4765	0.5896
4	-0.0007	0.0286	0.1297	0.1400
5	0.2633	0.2884	0.3618	0.3871
6	0.3383	0.3598	0.3472	0.3588
Order				
2	-1.9092***	-1.9163***	-1.7284***	-1.7281***
3	-0.3795	-0.3810	-0.1204	-0.1247
4	0.2953	0.2902	0.5277	0.5239
5	-0.0172	-0.0219	0.3349	0.3292
6	-0.6211	-0.6260	-0.3595	-0.3631
Numerical literacy		-0.0583		-0.0515
Financial literacy			0.0598	0.0665
Constant	0.1480	0.2658	-0.2451	-0.1828
Pseudo R^2	8.50%	8.56%	8.67%	8.71%
N	403	403	347	347

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

TABLE A.13 Logistic regressions for inconsistency in the HL method without one educational level				
	Logit I	Logit II	Logit III	Logit IV
Female	0.2211	-0.0600	-0.0307	-0.2095
Age group				
2	0.7755*	0.8459**	0.8636*	1.0469**
3	0.8765**	0.8618*	0.7011	0.7560
4	1.1707***	1.0342**	1.2019***	1.1353**
Education				
3	-0.4597	-0.3150	-0.2695	-0.2138
Income				
2	0.4215	0.2782	0.5798	0.5392
3	-0.1202	-0.1877	-0.3776	-0.3460
4	-0.1376	-0.1500	-0.0471	-0.0278
5	-0.4110	-0.3572	-0.4071	-0.2338
Wealth				
2	0.0246	0.1625	-0.3389	-0.2275
3	-0.6346	-0.5130	-0.7709	-0.7074
4	-0.5528	-0.4754	-0.2389	-0.2882
5	-0.6961	-0.6721	-0.7154	-0.7223
6	-0.5461	-0.4812	-0.6856	-0.7410*
Order				
2	0.9593**	1.0299**	1.1400**	1.1921**
3	0.2538	0.2312	0.4761	0.3889
4	0.6993	0.6781	0.6880	0.6177
5	1.0824**	1.1144**	1.1619**	1.1249**
6	0.2429	0.2034	0.2053	0.1259
Numerical literacy		-0.5513***		-0.6131***
Financial literacy			-0.2129*	-0.1448
Constant	-1.7705***	-0.6051	-0.9150	0.0530
Pseudo R^2	8.40%	13.02%	10.60%	15.76%
N	381	381	327	327

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

TABLE A.14 Linear regressions for stability of risk preferences

	Reg. I	Reg. II	Reg. III	Reg. IV
Female	0.1102	0.0554	-0.0333	-0.0681
Age group				
2	0.3306	0.3276	0.3111	0.3185
3	0.0483	0.0264	-0.0101	-0.0208
4	0.0014	-0.0392	-0.0265	-0.0523
Education				
2	-0.2441	-0.2204	-0.2927	-0.2591
3	-0.3345	-0.2846	-0.4059	-0.3629
Income				
2	-0.4007	-0.4307	-0.4538	-0.4704
3	-0.3687	-0.3817	-0.3032	-0.3027
4	-0.4707	-0.4811	-0.3791	-0.3898
5	-0.0038	-0.0050	-0.0459	-0.0426
Wealth				
2	-0.0680	-0.0495	-0.2779	-0.2648
3	-0.411	-0.0146	-0.0941	-0.0849
4	-0.3350	-0.2904	-0.2220	-0.2075
5	-0.3123	-0.2821	-0.3047	-0.2805
6	-0.0162	0.0069	0.0647	0.0672
Numerical literacy		-0.0906		-0.0759
Financial literacy			-0.0359	-0.0246
Constant	1.9929***	2.1904***	2.2472***	2.3492***
R^2	6.01%	6.61%	5.80%	6.23%
N	371	371	320	320

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).

TABLE A.15 Logistic regressions for stable risk preferences				
	Logit I	Logit II	Logit III	Logit IV
Female	-0.2004	-0.2303	-0.0115	-0.0558
Age group				
2	0.1862	0.1837	0.0541	0.0632
3	-0.0975	-0.1088	-0.4465	-0.4598
4	0.1433	0.1234	0.1271	0.0994
Education				
2	-0.4333	-0.4211	-0.2536	-0.2105
3	-0.1313	-0.1055	-0.0386	0.0162
Income				
2	0.6254	0.6112	0.4375	0.4196
3	0.3808	0.3757	0.1551	0.1561
4	0.4716	0.4665	0.3239	0.3100
5	0.3505	0.3520	0.2595	0.2624
Wealth				
2	0.1663	0.1735	0.2045	0.2177
3	-0.1589	-0.1468	-0.2676	-0.2580
4	0.4017	0.4235	0.0769	0.0897
5	0.2178	0.2310	0.0304	0.556
6	-0.2850	-0.2761	-0.5613	-0.5625
Numerical literacy		-0.0458		-0.0852
Financial literacy			-0.0009	0.0110
Constant	-1.5791*	-1.4796	-1.3758	-1.2631
Pseudo R^2	1.70%	1.74%	2.02%	2.15%
N	371	371	320	320

Statistical significance is indicated by * (10% level), ** (5% level), and *** (1% level).