How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods

Csermely, Tamás; Rabas, Alexander

DOI:
10.57938/a6fa10ac-1597-4f68-9cc9-94f451eb9a77

Published: 01/10/2014

Document Version
Publisher's PDF, also known as Version of record

Department of Economics
Working Paper No. 185

How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods

Tamás Csermely
Alexander Rabas

October 2014

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods

Tamás Csermely[a] Alexander Rabas[b]

[a] Vienna University of Economics and Business, Department of Economics
[b] University of Vienna, Doctoral School of Economics

Abstract

The question of how to measure and classify people's risk preferences is of substantial importance in the field of Economics. Inspired by the multitude of ways used to elicit risk preferences, we conduct a holistic investigation of the most prevalent method, the multiple price list (MPL) and its derivations. In accordance with previous literature, we find that revealed preferences differ under various and even the same versions of the MPL. Thus, an arbitrary selection of a particular risk assessment method can lead to biased results especially if researchers investigate its connection to other phenomena. In order to resolve this issue, we determine the most stable version of the MPL by using multiple measures of within-method consistency, and the version with the highest forecast accuracy by using behavior in two economically relevant games as benchmarks. A derivation of the well-known method by Holt and Laury (2002), where the highest payoff is varied instead of probabilities, emerges as the best MPL method in both dimensions.

Keywords: Risk, MPL, Experiment, Revealed Preferences

JEL Classification: C91 · D81

[a]Corresponding Author,
Welthandelsplatz 1, D4, 1020 Vienna, Austria
tamas.csermely@wu.ac.at
Tel.: +43 1 31336 4264

[b]Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
alexander.rabas@univie.ac.at

1 Introduction

Risk is a fundamental concept that affects human behavior and decisions in many real-life situations. Whether a person wants to invest in the stock market, tries to select the best health insurance or just wants to cross the street, he/she will face risky decisions every day. Therefore, risk attitudes are of high importance for decisions in many economics-related contexts. A multitude of studies elicit risk preferences in order to control for risk attitudes, as it is clear that they might play a relevant role in explaining results - e.g. de Véricourt et al. (2013) in the newsvendor setting, Murnighan et al. (1988) in bargaining, Beck (1994) in redistribution or Tanaka et al. (2010) in linking experimental data to household income, to name just a few. Moreover, several papers try to shed light on the causes of risk-seeking and risk-averse behavior in the general population with laboratory (Harrison and Rutström, 2008), internet (von Gaudecker et al., 2011) and field experiments (Andersson et al., 2013; Harrison et al., 2007). Since the seminal papers by Holt and Laury (2002, 2005), approximately 20 methods have been published which provide methods to elicit risk preferences. They differ from each other in terms of the varied parameters, representation and framing. Many of these different risk elicitation methods have the same theoretical foundation under expected utility theory (EUT) and therefore claim to measure the same parameter - a subject's "true" risk preference. However, there are significant differences in results depending on the method used, as an increasing number of evidence suggests.

It follows that if someone's revealed preference is dependent on the measurement method used, scientific results and real-world conclusions might be biased and misleading. In the end, our goal is to provide a blueprint on how to measure risk - and potentially other - preferences so that one's actual choice in the elicitation tasks clearly reflect his/her real preferences and are stable over time.

As far as existing comparison studies are concerned, they usually compare two to three methods with each other and often use different stakes, parameters, framing, representation, etc., which makes their results hardly comparable. Our paper complements existing experimental literature by making the following contribution: Taking the method by Holt and Laury (2002) as a basis, we conduct a comprehensive comparison of the multiple price list (MPL) versions of risk elicitation methods by classifying all methods into nine categories. To the best of our knowledge, no investigation - including various measures of between- and within-method consistency - has ever been conducted in the literature that incorporates such a high number of methods. To isolate the effect of different methods, we consistently use the MPL representation and calibrate the risk intervals to be the same for each method assuming expected utility theory (EUT) and constant relative risk aversion (CRRA), while also keeping the risk-neutral expected payoff of each method constant and employing a within-subject design.

We investigate the within-method consistency of each method by comparing the differences in subjects' initial and repeated decisions within the same MPL method. We assess the methods' self-perceived complexity and shed light on whether between-method consistency exists by comparing the distributions of risk attitudes among elicitation methods. In the end, we provide suggestions which specific MPL representation to use by comparing our results to decisions in two benchmark games that resemble real-life settings: investments in capital markets and auctions. Therefore, we analyze the methods along two dimensions: robustness and predictive power. We solidify our results by various robustness and consistency checks e.g. by relaxing our assumptions on the functional form of subjects' risk attitudes.

We find that a particular modification of the method by Holt and Laury (2002) derived by Drichoutis and Lusk (2012) has the highest predictive power in investment settings both according to the OLS regression and Spearman rank correlation. In addition, specific certainty equivalent methods derived by Bruner (2009) perform also relatively well in these analyses. Moreover, the method by Drichoutis and Lusk (2012) clearly outperforms the other methods in terms of within-method consistency and it is perceived as relatively simple - in the end, our study provides the recommendation for researchers to implement this method when measuring risk attitudes.

1.1 Multiple Price Lists Explained

Incentivized risk preference elicitation methods aim to quantify subjects' risk perceptions based on their revealed preferences. We present nine methods in a unified structure - the commonly used MPL format - to our subjects, taking one of the most cited methods as a basis: Holt and Laury

Table 1: *Risk Parameter Intervals (Holt/Laury)*

Interpretation by Holt/Laury (2002)	Switching Point	Risk Parameter Interval
highly risk loving	1	$\rho \leq -0.95$
very risk loving	2	$-0.95 < \rho \leq -0.49$
risk loving	3	$-0.49 < \rho \leq -0.15$
risk neutral	4	$-0.15 < \rho \leq 0.15$
slightly risk averse	5	$0.15 < \rho \leq 0.41$
risk averse	6	$0.41 < \rho \leq 0.68$
very risk averse	7	$0.68 < \rho \leq 0.97$
highly risk averse	8	$0.97 < \rho \leq 1.37$
stay in bed	Never	$\rho > 1.37$

Notes: This table indicates the mapping from a subject's chosen switching point into the resulting risk parameter intervals in each method; the leftmost column contains the interpretation of the risk intervals; "Never" means a subject prefers the option "Left" in each row

(2002). The MPL table structure is as follows: Each table has multiple rows, and in each row all subjects face a lottery (two columns) on one side of the table, and a lottery or a certain payoff (one or two columns) on the other side, depending on the particular method. Then, from row to row, one or more of the parameters change. The methods differ from each other by the parameter which is changing. As the options on the right side become strictly more attractive from row to row, a subject indicates the row where he/she wants to switch from the left option to the right option. This switching point then gives us an interval for a subject's risk preference parameter according to Table 1[1], assuming EUT and CRRA[2].

Andersen et al. (2006) consider that the main advantage of the MPL format is that it is transparent to subjects and it provides simple incentives for truthful preference revelation. They additionally list its common usage,

[1] To ease comparison to existing studies, we used exactly the same coefficient intervals as Holt and Laury (2002).

[2] $u(c) = \frac{c^{1-\rho}}{1-\rho}$

its simplicity and the little time it takes as further benefits. As far as the specific risk elicitation method in the MPL framework designed by Holt and Laury (2002) is concerned, it has proven itself numerous times in providing explanations for several phenomena such as behavior in 2x2 games (Goeree et al., 2003), market settings (Fellner and Maciejovsky, 2007), smoking, heavy drinking, being overweight or obese (Anderson and Mellor, 2008), consumption practices (Lusk and Coble, 2005) and many others.

We acknowledge that recent studies (Tanaka et al., 2010; Bocqueho et al., 2011) document potential empirical support for prospect theory (PT) when it comes to risk attitudes: Wakker (2010) provides an extensive review on risk under PT. However, our choice for EUT is justified by Harrison et al. (2010), who provide evidence that the majority of subjects behave according to EUT. Moreover, we justify using CRRA as Wakker (2008) claims that it is the most commonly postulated assumption among economists. Most recently, Chiappori and Paiella (2011) provide evidence on the validity of this assumption in economic-financial decisions.[3]

We group our aforementioned nine risk elicitation methods into two categories:

1. The certainty equivalent methods (CE methods), where on one side of the table there is always a 100% chance of getting a particular certain payoff and on the other side there is a lottery.

2. The Holt/Laury methods (HL methods), where you face lotteries on both sides.

We therefore primarily conduct a clean comparison of different MPL risk elicitation methods. What we do not claim, however, is that the method devised by Holt/Laury (2002) (or MPL in general) is the most fitting to measure people's risk preferences - we strive to give a recommendation to researchers who already intend to use Holt/Laury (2002) in their studies, and provide a better alternative that shares its attributes with the original Holt/Laury design.

It should be mentioned that there is an alternative interpretation of our study, to be taken with a grain of salt: The different MPL methods can

[3] Note that this approach is also popular among economists due to its computational ease: The vast majority of economic experiments assumes CRRA as well, which makes our results comparable to theirs.

Table 2: *Method Overview*

	What is changing?			
Method	Probability	Highest Payoff	Lowest Payoff	Sure Payoff
CEp	**yes**	no	no	no
CEhigh	no	**yes**	no	no
CElow	no	no	**yes**	no
CEsure	no	no	no	**yes**
CEall	no	**yes**	**yes**	**yes**
HLp	**yes**	no	no	NA
HLhigh	no	**yes**	no	NA
HLlow	no	no	**yes**	NA
HLall	**yes**	**yes**	**yes**	NA

Notes: This table indicates which parameters change from row to row in each method.

also be conceived as a mapping of existing risk elicitation methods (from other frameworks) to the MPL space.

Up to now, different risk elicitation methods were compared by keeping the original designs and comparing the differences in elicited risk parameters. But the approach to keep the original designs comes at a price: The methods differ in many dimensions, so any differences found can be attributed to any of those particular characteristics. Our approach can be understood as a way to make all risk elicitation methods as similar as possible, with the drawback of losing the direct connection to the original representation. This paper should therefore primarily be seen as a comparison of different MPL risk elicitation methods, and the resulting comparison of existing risk elicitation methods by mapping them into the same space is only reported for the sake of completeness.

1.2 Literature Review

We will now discuss the different methods in greater detail and how they are embedded in the literature, if at all. Table 2 provides a summary of the exact parameter that is changing across methods. For a complete list of all methods with the corresponding parameter values (as presented to subjects), please refer to the Appendix.

1.2.1 CE Methods

Among the CE methods, there are four parameters that can be changed: The sure payoff (*sure*), the high payoff of the lottery (*high*), the low payoff of the lottery (*low*) or the probability of getting the high payoff (p) (or the probability of getting the low payoff ($1 - p$), respectively). The parameters must of course be chosen in such a way that $high > sure > low$ always holds. For instance, we denote the CE method where the low payoff is changing by "CElow", the CE method with the varying certainty equivalent by "CEsure" or the Certainty Equivalent method where the probabilities are changing as "CEp". For example, Cohen et al. (1987) used risk elicitation tasks in which probabilities and lottery outcomes were held constant and only the certainty equivalent was varied.

A recent investigation by Abdellaoui et al. (2011) presents a similar method (CEsure method) in an MPL format with 50-50 probabilities. Bruner (2009) presents a particular certainty equivalent method, where the certainty equivalent and the lottery outcomes are held constant, but the corresponding probabilities of the lotteries are changing (CEp method). Additionally, Bruner (2009) introduces another method where only the potential high outcomes of lotteries vary (CEhigh method). Binswanger (1980) introduced a method (CEall) where only one of the options has a certainty equivalent. The other options consist of lotteries where the probabilities are fixed at 50-50, but both the high and the low payoff are changing. These methods have later been redesigned and presented in a more sophisticated format as a single choice task by Eckel and Grossman (2002, 2008). Although not present in the literature, we chose to include a method where the potential low outcome varies for reasons of completeness (CElow method). For examples, see Tables 9-13 in the Appendix, which correspond to the CE methods.

1.2.2 HL Methods

Holt and Laury (2002, 2005) introduced one of the most cited elicitation method under EUT up to now, where potential outcomes are held constant and the respective probabilities change (HLp). Drichoutis and Lusk (2012) suggest a similar framework where the outcomes of different lotteries change while the probabilities are held constant. We differentiate these methods further into HLhigh and HLlow depending on whether the high or the low outcome is varied in the MPL. Additionally, the HLall method varies both the probabilities and the potential earnings at the same time.

Many risk elicitation tasks used in the literature fit into the framework of choosing between different lotteries. Sabater and Grande (2002) provide ten discrete options with different probabilities and outcomes to choose from. Lejuez et al. (2002) introduce the Balloon Analogue Risk Task (BART) where subjects could pump up a balloon, and their earnings depend on the final size of the balloon. The larger the balloon gets, the more likely it will explode, in which case the subject earns nothing. Visschers et al. (2005) and Andreoni and Harbaugh (2010) use a pie chart for probabilities and a slider for outcomes to visualize a similar trade-off effect in their experiment. Crosetto and Filippin (2013) present their Bomb Risk Elicitation Task with an interesting framing which quantifies the aforementioned tradeoff between probability and potential earnings with the help of a bomb explosion. For examples, see Tables 14-17 in the Appendix, which correspond to the HL methods.

1.2.3 Questionnaire Methods

In addition to the MPL methods, we chose to also incorporate questionnaire risk elicitation methods into our study. Several methods have been introduced that evaluate risk preferences with non-incentivized survey-based methods, and the questions and the methodology they use are very similar. The most recently published ones include the question from the German Socio-Economic Panel Study (Dohmen et al., 2011) or the Domain-Specific Risk-Taking Scale (DOSPERT) by Blais and Weber (2006). For a more detailed description, see the last paragraph of Section 2.

1.2.4 Comparison Studies

The question arises which method to use if there are so many different methods for risk elicitation and whether they lead to the same results. There are comparison studies, but the majority of them compares only two methods with each other, thus their scope is limited. The question of within-method consistency has been addressed by several papers: Harrison et al. (2005) document high re-test stability of the method introduced by Holt and Laury (2002). Andersen et al. (2008b) test consistency of the HLp (2002) method within a 17-month time frame. They find some variation in risk attitudes over time, but do not detect a general tendency for risk attitudes to increase or decrease. This result was confirmed in Andersen et al. (2008a). Recently, Straznicka (2012) records high temporal stability of five known methods on the aggregate level and a relatively small change in individual perceptions.

Interestingly, more work has been done on the field of between-method consistency. Fausti and Gillespie (2000) compare risk preference elicitation methods with hypothetical questions using results from a mail survey. Isaac and James (2000) conclude that risk attitudes and relative ranking of subjects is different in the Becker-DeGroot-Marschak (BDM) procedure and in the first-price sealed-bid auction setting. Berg et al. (2005) confirm that assessment of risk preferences varies generally across institutions in auction settings. In another comparison study, Bruner (2009) shows that changing the probabilities versus varying the payoffs leads to different levels of risk aversion in the HL tasks. Moreover, Dave et al. (2010) conclude that subjects show different degrees of risk aversion in the Holt and Laury (2002) and in the Eckel and Grossman (2008) task. Their results were confirmed by Reynaud and Couture (2012) who used farmers as the subject pool in a field experiment. Bleichrodt (2002) argues that a potential reason for these differences might be attributed to the fact that the original method by Eckel and Grossman (2008) does not cover the risk seeking domain, which can be included with the slight modification we made. Dulleck et al. (2014) test the method devised by Andreoni and Harbaugh (2010) using a graphical representation against the HLp and describe both a surprisingly high level of within- and inter-method inconsistency. Drichoutis and Lusk (2012) compare the HLp method to a modified version of it where probabilities are held constant. Their analysis reveals that the elicited risk preferences differ from each other both at the individual and at the aggregate level. Most recently, Crosetto and Filippin (2013) compare five different risk preference elicitation methods and confirm the relatively high instability across methods.

In parallel, a debate among survey-based and incentivized preference elicitation methods emerged. Eckel and Grossman (2002) conclude that non-incentivized survey-based methods provide misleading conclusions for incentivized real-world settings. In line with this finding, Anderson and Mellor (2009) claim that non-salient survey-based elicitation methods and the HLp method yield different results. On the contrary, Lönnqvist et al. (2011) provide evidence that the survey-based measure, which Dohmen et al. (2011) had implemented, explains decisions in the trust game better than the HLp task. Charness and Viceisza (2012) provide evidence from developing countries that hypothetical willingness-to-risk questions and the HLp task deliver deviating results.

1.2.5 Theoretical Considerations

A recent stream of literature broadens the horizon of investigation to theoretical aspects of different elicitation methods: Weber et al. (2002) show that people have different risk attitudes in various fields of life, thus risk preferences seem to be domain-specific. Hey et al. (2009) investigate noise and bias under four different elicitation procedures and emphasize that elicitation methods should be regarded as strongly context specific measures. Harrison and Rutström (2008) provide an overview and a broader summary of elicitation methods under laboratory conditions, whereas Charness et al. (2013) evaluate several risk preference elicitation methods based on their advantages and disadvantages and give suggestions which ones to use.

In addition, there is evidence that framing and representation matters. Wilkinson (2005) advised against using pie charts showing probabilities and payoffs as human beings are not good at estimating angles. Hershey et al. (1982) identify important sources of bias to be taken into account and pitfalls to avoid when designing elicitation tasks. Most importantly, these include task framing, differences between the gain and loss domains and the variation of outcome and probability levels. Von Gaudecker et al. (2008) show that the same risk elicitation methods for the same subjects deliver different results when using different frameworks - e.g. multiple price list, trade-off method, ordered lotteries, graphical chart representation, etc. This procedural indifference was confirmed by Attema and Brouwer (2012) as well, which implies that risk preferences on an individual level are susceptible to the representation and framing used.

All previous paragraphs therefore lead us to the conclusion that methods should be compared to each other by using the same representation and format. This justifies our decision to compare our methods using the standard MPL framework which guarantees that the differences cannot be attributed to the different framing and representation of elicitation tasks. However, this comes at the price that we had to change some of the methods slightly, which implies that they are not exactly the same as they were published originally. We certainly do not claim that the MPL is the only valid framework, but our choice for it seems justified by its common usage

Table 3: *Link between MPL Representation and Literature*

Method	Corresponding Literature
CEp	Bruner (2009)
CEhigh	Bruner (2009)
CElow	
CEsure	Cohen, Jaffray and Said (1987),
	Abdellaoui, Driouchi and L'Haridon (2011)
CEall	Binswanger (1980), Eckel and Grossman (2008)
HLp	Holt and Laury (2002), Holt and Laury (2005)
HLhigh	Drichoutis and Lusk (2012)
HLlow	Drichoutis and Lusk (2012)
HLall	Sabater and Grande (2002), Lejuez et al. (2002),
	Andreoni and Harbaugh (2010), Crosetto and Filippin (2013)
Questionnaire	Weber, Blais and Betz (2002), Dohmen et al. (2011)

Notes: In the first column, this table lists all MPL and questionnaire methods, and in the second column the corresponding literature.

and relative simplicity. We consider a future investigation using a different representation technique as a potentially interesting addition. Also, we emphasize that the differences in our results exist among the MPL representations of the methods and they can only be generalized to the original methods to a limited extent; see Table 3 for an overview of the link between the MPL representation and the particular method that was published originally.

2 Design

We provide a laboratory experiment to compare different MPL risk elicitation methods. Subjects answered the risk elicitation questions first. Then, benchmark games were presented to them to gauge predictive power, which was followed by a non-incentivized questionnaire. We will provide a detailed description on the exact procedures of each part in the later paragraphs.

We conducted ten sessions at the Vienna Center for Experimental Economics (VCEE) with 97 subjects using a within-subject design.[4] Sessions lasted 1 hour and 55 minutes on average. The range of earnings was between 3.00€ and 50.00€, amounting to a payment of 20.78€ on average. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). The software ORSEE (Greiner, 2004) was used for recruiting subjects.

After receiving instructions on screen and in written form, subjects went through the nine incentivized risk elicitation methods in random order. In order to avoid potential incentive effects mentioned by Holt and Laury (2002), the expected earnings for a risk-neutral individual were equal in each and every method. Furthermore, to avoid potential biases due to the difference between gains and losses (Hershey et al., 1982), each of our lotteries is set in the gains domain. Andersen et al. (2006) found evidence that there is a slight tendency of anchoring and choosing a switching point around the middle for risk elicitation tasks. In order to counteract anchoring and one-directional distortion of preferences as a consequence of this unaviodable pull-to-center effect, each risk elicitation task appeared randomly either top-down or bottom-up. Depending on randomization, out of nine potential switching opportunities the fourth or the sixth option were the risk-neutral switching points.

Subjects also had the opportunity to look at their given answer and modify it right after each decision if they wished to do so. After making a decision in each method, we asked subjects the following question: "On a scale from 1 to 10, how difficult was it for you to make a decision in the previous setting?". With this question we assessed self-perceived complexity of the tasks, since there is evidence in the literature (Mador et al., 2000) that subjects make noisier decisions if the complexity of a lottery increases, and therefore a less complex method is preferred. Moreover, Dave et al. (2010) outline the trade-offs between noise, accuracy and subjects' mathematics skills. They suggest that it is a good strategy to make MPL tasks simpler for subjects. In this spirit, we asked our subjects to indicate the row in which they switched from the 'LEFT' column to the 'RIGHT' column, thereby enforcing a single switching point (SSP). Using this framework, subjects were not required to make a decision for each and every row in each method, which would have meant more than 100 monotonous, repeti-

[4] One subject has been excluded after repeatedly being unable to answer the control questions correctly.

tive binary choices throughout the experiment. Additionally, this approach ensures that the subjects were guaranteed to give answers without preference reversals. We consider this option more viable than accepting multiple switching points - thus allowing inconsistent choices - and using the total number of “safe” choices to determine a subject’s risk coefficient interval, in line with the findings of Gonzalez and Wu (1999) or Jacobson and Petrie (2009).

In order to test within-method consistency, three of the nine methods were randomly chosen and presented to subjects again - without telling them that they had already encountered that particular method. This approach allows us to test both within-method and inter-method consistency. The modification of subjects’ answers was allowed here once as well. The perceived complexity of tasks was also elicited again.

Control questions were used for the preference elicitation methods and for every benchmark game in order to verify that subjects understood the task they were about to perform. We incorporated the random lottery incentive system emphasized by Cubitt et al. (1998). Thus, the computer chose one of the twelve risk preference methods and one of the eight rows within that particular method on a random basis to be payoff-relevant. Additionally, one of the four benchmark games was chosen to be payoff-relevant as well. Blanco et al. (2010) provided evidence that hedging behavior and the corresponding biased beliefs and actions can only be problematic if the hedging opportunities are highly transparent. Taking this consideration into account, we provided feedback on the outcome of the risk elicitation tasks only at the end of the experiment. Moreover, the risk elicitation tasks were very heterogeneous in terms of potential earnings and their respective payoffs. In addition, the subjects did not know which elicitation task and which row within that specific task would be payoff-relevant to them. Thus, it was not possible for subjects to create a portfolio and use hedging behavior over different parts of the experiment.

On top of the risk elicitation tasks, we used three different benchmark games resembling real-life situations as well as situations relevant to economists. As behavior in these settings should only depend on risk attitudes, they will serve as benchmarks to contribute to the debate which risk elicitation methods are appropriate to predict behavior in these games. The benchmark games appeared in a randomized order. First, we used the same investment task as Charness and Gneezy (2010). Here, subjects could decide how much they wanted to invest in stocks and bonds out of an endowment of 10€. Subjects know that any investment in bonds is a safe investment, and therefore they received the same amount they had invested in bonds as income. Additionally, the amount they invested in stocks was to be be multiplied by 2.5 or lost completely with equal chance. Under EUT, this setting implies that both risk neutral and risk seeking decision makers should invest the entire amount. Thus, in order to be able to differentiate between them, we introduced another investment setting where the potential payment for stocks was 1.5 times the invested amount.

The other benchmark game was a first-price sealed-bid auction against a computerized opponent in line with Walker, Smith and Cox (1987). Subjects could bid between 0.00€ and 20.00€ of their endowment, and the subjects knew that the computer bid any amount between 0.00€ and 20.00€ with equal chance. The potential earnings (E_1 for subject 1) according to the bids (x_1; x_2) are:

$$E_1 = \begin{cases} 20 - x_1 & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 < x_2 \\ 20 - x_i \text{ or } E_1 = 0 \text{ (with 50\% chance)} & \text{if } x_1 = x_2 \end{cases}$$

Our benchmark games are deliberately chosen in such a way that risk is clearly relevant in the games, while being one step away from the artificial risk elicitation mechanisms. Therefore, all benchmark games are framed heavily, while still ensuring that risk attitudes should be the only factor driving a subject's decisions. The investment settings are very similar to the risk elicitation mechanisms described above in the sense that they resemble a CE method (with the difference that you choose your sure payoff and your lottery at the same time). The auction is more complex, as the optimal risk-neutral solution is harder to compute, but here you basically choose your own lottery, too. We therefore expect stronger correlation for the investment games. Our aim is to let the data speak and to see which method explains behavior best in each case.

The experiment concluded with an extensive questionnaire. Subjects received an additional 3.00€ for filling it out. Harrison et al. (2009) provide evidence that the existence of a show-up fee could lead to an elevated level of risk aversion in the subject pool. In our experiment, this moderate show-up fee was only pointed out to the subjects after making their decisions in the risk elicitation methods and the benchmark games. Thus, the show-up fee could not distort their preferences or their answers. In order to incorporate survey-based measures, we asked subjects to provide an answer on a ten-point Likert-scale to the following two questions in line with Dohmen et al. (2011): "In GENERAL, are you a person who is fully prepared to take risks or do you try to avoid taking risks?" and "In FINANCIAL SITUATIONS, are you a person who is fully prepared to take risks or do you try to avoid taking risks?". The perceived complexity of these questions was elicited as well. In the questionnaire, we elicited the following socioeconomic factors: Age, gender, field of study, years of university education, nationality, high school grades in mathematics, monthly income and monthly expenditure. Furthermore, we elicit cognitive ability by conducting a cognitive reflection test (Frederick, 2005). Lastly, we assessed subjects' personalities in line with Rammstedt and John (2007) who provide a short measure of personality traits according to the BIG5[5] methodology introduced by Costa and McCrae (1992).

3 Results

We will first establish in sections 3.1 and 3.2 that the elicited risk parameter is highly dependent on the particular variant of MPL used because the overall distributions of switching points are very diverse and the rank correlations between the different methods are low in most circumstances. In section 3.3 we apply multiple measures of a quality of a method. First we use the benchmark games to let the data speak which risk elicitation methods predict behavior in these games best. In Section 3.3.2, we will show which method produces the most stable results overall. Section 3.4 concludes with the result that the HLhigh method is the most stable method and the has the highest predictive power.

3.1 Overall Distributions are Different

According to expected utility theory, a subject's behavior does not depend on which parameters are changed from row to row, as his underlying risk parameter value is constant. As the different versions of the MPL are calculated in such a way that the same switching point implies the same risk

[5] In the BIG5, personality is measured along five dimensions: Agreeableness, Conscientousness, Extraversion, Neuroticism and Openness.

parameter interval, a consistent individual should have the same switching point in all versions of the MPL. This implies that the distributions of switching points should be the same across methods, barring some noise.

First, see Figure 1 for a graphical representation of the distributions. It is clearly visible at first glance that the distributions are not the same across all methods. For example, in the CEp method, most subjects would be classified as highly risk loving, whereas in the HLhigh method the majority of subjects would be classified as risk averse.
To verify whether distributions across methods are the same, we conduct two tests: a Friedman test, which shows that the means are not the same across methods ($p < 0.0001$), and a Kruskal-Wallis test, which shows that the distribution of answers is not the same across methods ($p < 0.0001$). We conclude that the switching points are, contrary to standard theory but in accordance with the literature, dependent upon the version of the particular MPL variation used.

To see which specific versions are significantly different from each other, we conduct a series of Wilcoxon tests, the natural pairwise analogue to the Kruskal-Wallis test. We use the Wilcoxon test to give a comparison of the distributions, as a difference in distributions is a more meaningful statistic here than a comparison of means. The p-values of the pairwise tests can be found in Table 4. Out of 55 pairwise comparisons, 28 comparisons indicate that methods are different at $p < 0.001$. 34 (43) instances suggest that methods are different at $p < 0.01$ (0.05) significance levels.[6]

We conclude that different methods deliver significantly different results, and that the different versions of the MPL cannot be used interchangeably, as the estimated risk preference parameter depends heavily on the version used. A subject might easily be classified as risk loving in one version and as risk averse in another. Of course we do not know a subject's true risk preferences, and therefore any of the methods might be able to classify a subject correctly. To provide an answer to this puzzle, see Section 3.3.2, where we conduct a quality assessment of the different methods.

[6] Note that one should be careful while reading this table and the ones following because of the presence of a multiple testing problem; therefore, we introduce a new notation in the tables: p-values lower than 0.001 are denoted by three stars, p-values lower than 0.01 are denoted by two stars and p-values lower than 0.05 are denoted by one star. $p < 0.001$ can be interpreted as significant, even when using the conservative Bonferroni correction (see Abdi, 2007).

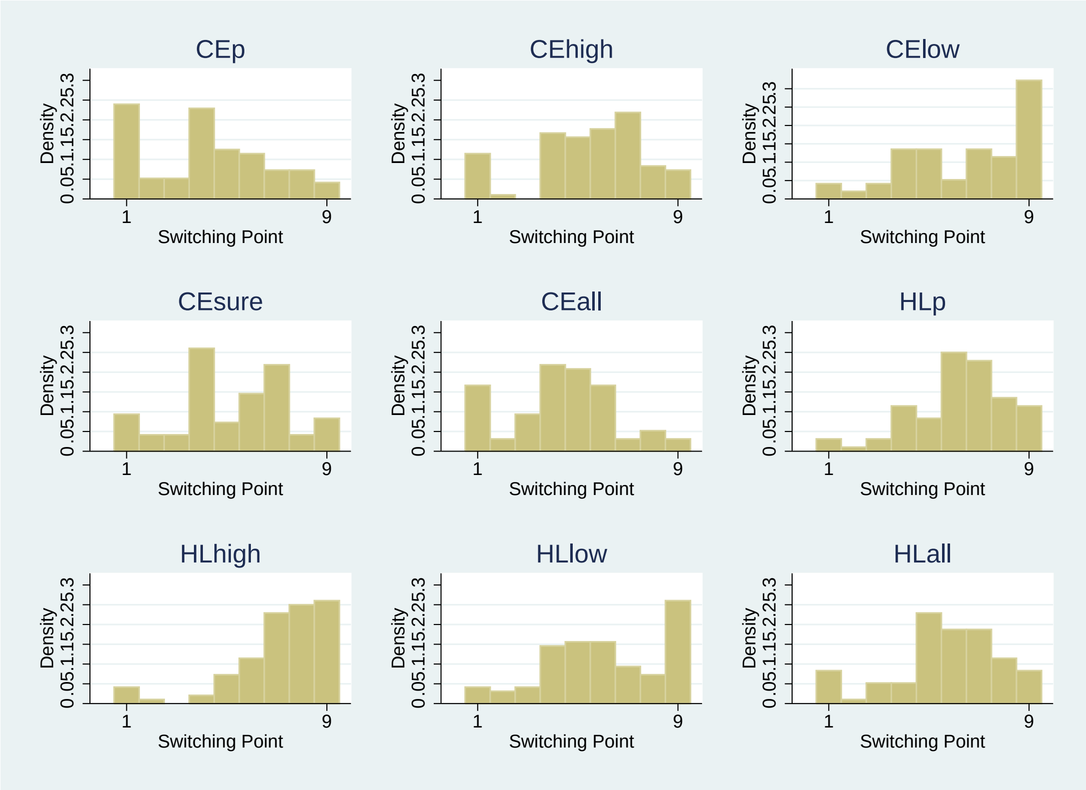


Figure 1 *Distributions of risk preferences; a low value indicates risk loving and a high value indicates risk averse behavior; x-axis: switching points (e.g. risk preferences) of subjects, where 1 means a subject switches from left to right in the first row and 9 means a subject never switches; y-axis: frequency of switching point*

Table 4: *Pairwise Wilcoxon test for equality of distribution*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall	GQ
CEhigh	.00***									
CElow	.00***	.00***								
CEsure	.00***	.37	.00***							
CEall	.79	.00***	.00***	.00**						
HLp	.00***	.00**	.28	.00***	.00***					
HLhigh	.00***	.00***	.02*	.00***	.00***	.00***				
HLlow	.00***	.02*	.23	.01**	.00***	.68	.00***			
HLall	.00***	.31	.02*	.04*	.00***	.08	.00***	.39		
GQ	.02*	.03*	.00***	.29	.04*	.00***	.00***	.00***	.00**	
FQ	.00***	.64	.00**	.29	.00***	.02*	.00***	.04*	.36	.00***

*Notes: p-values of pairwise Wilcoxon tests are displayed; GQ: general question; FQ:financial question; stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

Table 5: *Spearman Rank Correlation Coefficients*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall	GQ
CEhigh	.46***									
CElow	.33***	.44***								
CEsure	.05	.22*	.26*							
CEall	.03	.18	-.03	.19						
HLp	.17	.15	.17	.21*	-.04					
HLhigh	.20	.39***	.21*	.03**	.21*	.25*				
HLlow	.31**	.28**	.25*	.19	-.02	.13	.21*			
HLall	.24	.21*	-.01	.08	.19	.04	-.01	.08		
GQ	.15	.13*	.06	-.12	.11	.02	.14	.04	.06	
FQ	.26*	.23*	.29*	.18	.10	-.04	.04	.24*	.13	.46***

*Notes: Table includes the nine different methods and the questionnaires (GQ: general questionnaire, FQ: financial questionnaire); stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

3.2 Rank Correlations are Low

In this section we look at the rank correlation coefficients between the different methods and the questionnaire answers. If there are high rank correlations between the risk elicitation methods, one might argue that it is irrelevant which one is used if one intends to control for risk attitudes under any given circumstance. Rank correlations between the MPL methods and the questionnaire measures can be found in Table 5. We see that some of the correlations are significant, but only 11% of all pairwise comparisons in total if we test conservatively at $p < 0.001$ because of the multiple testing problem. Pay special attention to the fact that HLp, the most widely used method today, has no significant rank correlations with any of the other methods.[7] See also Table 20 in the Appendix for standard correlations, which basically give the same results as Table 5.

These findings provide further evidence that the elicitation procedure should be chosen with care as the elicited risk aversion coefficient and also the relative ranking of subjects according to each method varies within broad boundaries.

[7] Also, the Financial questionnaire (FQ) results have much higher correlations with the other methods than the general questionnaire (GQ) results, strengthening the argument that risk attitudes are domain specific.

3.3 Method Quality Indicators

We use two avenues to measure a method's quality: its predictive power (section 3.3.1) and its stability (section 3.3.2).

3.3.1 Predictive Power

In order to see which method predicts behavior best in our benchmark games, we look at three statistics: predictive power by simple OLS regression, predictive power by Spearman rank correlation, and absolute average deviation from the prediction.

In Table 6, we see in the upper part the outcome of an OLS regression where we control for personality measures (BIG5 test, see Section 2) and socioeconomic variables (self-reported income, gender, age, the number of correct answers on a cognitive reflection test (CRT) and years of university education). In the lower part of Table 6 you see Spearman rank correlation coefficients, which we include because one of the most important factors in a measure of a subject's risk attitudes is the correct rank ordering of subjects. The OLS regression can be understood as follows: The dependent variable is the outcome of a particular benchmark game, and the independent variables include the outcome of one of the risk elicitation methods plus all controls metioned above.[8] The corresponding adjusted R^2 values can be found in parentheses below the coefficients. The OLS regression equation is then given by

$$BG_{i,j} = \beta_0 + \beta_1 * MPL_j + \sum_{k=2}^{6} \beta_k * BIG5_k + \sum_{l=7}^{11} \beta_l * SE_l + \beta_{12} * CRT + \epsilon_i,$$

where i denotes the index of benchmark games (BG), j denotes the index of risk measures, MPL denotes the outcome of a risk elicitation method, $BIG5$ denotes personality measures according to the BIG5, SE denotes the socioeconomic variables and CRT denotes the number of correct answers in the cognitive reflection test.

Additionally, we can calculate for each risk elicitation method (but not for the questionnaires) a point prediction in each of the benchmark games. In Table 7 we report the absolute average deviations from these predictions,

8 It is not possible to add controls in the Spearman rank correlation.

averaged over all three benchmark games according to the formula

$$AAD = (\sum_{i=1}^{n} |H_i - H_i^*| + \sum_{i=1}^{n} |L_i - L_i^*| + \sum_{i=1}^{n} \frac{|A_i - A_i^*|}{2})/(3n),$$

where H_i denotes high investment game outcomes and H_i^* high investment game predictions (L stands for investment low and A for auction).[9]

In the auction, none of the methods produce statistically significant results in the OLS regression. This is puzzling, as the auction can in itself be seen as a risk elicitation procedure, albeit with heavy framing. Recent literature, however, provides evidence that not only risk attitudes but also other factors like regret aversion (Engelbrecht-Wiggans and Katok, 2008) could drive behavior in auctions. As far as the Spearman rank correlation is concerned, the CEp method is the only one that is rank correlated ($p < 0.05$) with auction behavior.

In the investment games, the methods produce much better results. In the low investment setting, HLhigh has the biggest explanatory power, with CEhigh also having significant explanatory power. Note that it is surprising that HLhigh is the best predictor both in the regression and the rank correlation, as the investment games in themselves can be interpreted as certainty-equivalent methods, so one would expect one of these methods to perform best.

In the high investment setting, many methods (HLhigh, HLlow, CEhigh, CElow, CEp, and the questionnaires) are able to explain a part of the variance, with CEp being the one giving the best results ($p < 0.01$). Note that in this setting, survey-based measures perform very well, so questionnaire measures seem to serve as good proxies for subjects' risk preferences in some circumstances.[10]

As far as the deviations from the predictions are concerned, HLhigh performs best with an average deviation of 1.75 across all benchmark games

[9] Note that we divide the deviation in the auction game by 2 because the choice range in the auction game is twice as high.

[10] Note that in all regressions, none of the controls were significant at $p < 0.05$. This implies that behavior seems to primarily be driven by risk attitudes.

with CEp and CElow also having low deviations.[11]

In conclusion, HLhigh and CEp yield the best results in explaining behavior, with HLhigh having the lowest deviation from the prediction of behavior in the benchmark games. We conclude that HLhigh has the highest predictive power with CEp being a close runner-up.[12]

Table 6: *Explanatory Power*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall	GQ	FQ
	OLS coefficients										
Auction	.21	.14	.15	.01	.14	.01	0	.18	-.16	0.13	-0.08
	(.05)	(.03)	(.04)	(.02)	(.02)	(.02)	(.02)	(.04)	(.03)	(.03)	(.03)
Inv. Low	.04	.27***	.13	0	.21	.11	.43***	.07	.13	0.3*	0.09
	(.00)	(.03)	(.00)	(.00)	(.01)	(.00)	(.08)	(.00)	(.00)	(.02)	(.00)
Inv. High	.27***	.21*	.08	-.07	-.17	.13	.19	.19*	-.06	0.28**	0.25**
	(.16)	(.12)	(.09)	(.09)	(.11)	(.09)	(.11)	(.12)	(.08)	(.13)	(.13)
	Spearman rank correlation coefficients										
Auct.	.23*	.09	.14	.17	.07	.11	.06	.16	-.13	.10	.11
Inv. L	.02	.19	.17	.06	.06	.11	.36***	.12	.04	.19	.11
Inv. H	.28**	.28**	.05	.00	.03	.13	.26*	.23*	.09	.31**	.28**

*Notes: In the OLS regression, the dependent variable is the outcome in one of the four benchmark games, the independent variables are the outcome from one method plus controls (age, gender, BIG5, CRT test, income, years of university education); the adjusted R^2 value for the regression can be found below a coefficient; Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

Table 7: *Deviations from Predictions*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall
Deviation	1.91	2.41	2.19	2.03	2.11	2.27	**1.75**	2.17	2.11

Notes: Absolute average deviations from the predictions in the benchmark games

[11] Note that one might be concerned with this analysis because if a method generally classifies subjects as risk-averse, it is not surprising that it explains behavior well in the low investment setting, as subjects naturally behave risk-averse in this setting due to the parameters. However, this critique is not valid for any method that provides good predictions across multiple benchmark games (e.g. HLhigh).

[12] For a robustness check of our results, refer to Table 19 in the Appendix, which confirms our results if people's risk attitudes follow constant absolute risk aversion instead of CRRA.

3.3.2 Stability Measures

In this section, we evaluate the stability of the different MPL representations. Remember that after our subjects had gone through all nine MPL methods, three of them were randomly chosen and presented to them again. A method can be described as stable if the given answers between the first and the second time a method was encountered are very similar. To analyze this similarity, we use three criteria: equality of overall distribution, equality of rank ordering and absolute average deviation between the first and second answers. For reasons of completeness, we also report the perceived complexity of each method.[13]

Table 8 reports these measures. In the first column we give p-values from a Kolmogorov-Smirnov test that tests whether the distributions of the first and second time a method is encountered are the same. A significant p-value means that the distributions are significantly different from each other, indicating a low stability of overall distribution across a 30 minute time period.
The second column gives the rank correlation between the first and second time a method was encountered. This measure is important because if a method's overall distribution merely shifted up or down without changing the rank ordering of subjects, this method can also be described as stable. The third column reports the absolute average deviation (AAD) of subject's answers when a particular method is presented to them again, compared to the first time. A lower value is therefore better. The last column gives the means of the perceived complexity of a method on a 1 to 10 Likert scale.
To visualize these results, we also report the distributions of the differences in switching points between the first and second time a method is encountered in Figure 2.

Any method that does not yield stable results over a 30 minute time period cannot be described as stable, and stability is a highly preferable characteristic in a risk measure. For the KS-test (column 1 in Table 8), stability means a nonsignificant result, indicating that the overall distributions of answers are not too different between the first and second time a method

[13] We do not use Complexity as a stability measure, as the impact of a higher perceived complexity is not clear. On the one hand, one might argue that a higher measure in these categories implies noisier behavior, on the other hand one might argue that a subject takes more time thinking about the problem at hand.

Table 8: *Stability Measures*

Method	KS-Test	Rank Corr.	AAD	Complexity
CEp	**.453**	.51***	1.60	3.42
CEsure	.003	.51**	**1.37**	3.92
CEhigh	**.644**	.39*	1.48	3.97
CElow	.007	.35	1.96	3.20
CEall	.005	.16	1.8	4.81
HLp	**.240**	.23	**1.33**	4.21
HLhigh	**.879**	.45**	**1.24**	3.78
HLlow	.006	.25	2.04	4.29
HLall	.000	.19	1.85	5.75

*Notes: First column: p-values for a Kolmogorov-Smirnov-Test of equality of distribution; Second column: rank correlation between the distributions of first and second answers (stars indicate significant rank correlation); AAD means Absolute Average Deviation between the first and second decision in the same method; Complexity indicates a subject's perceived complexity of a method; stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

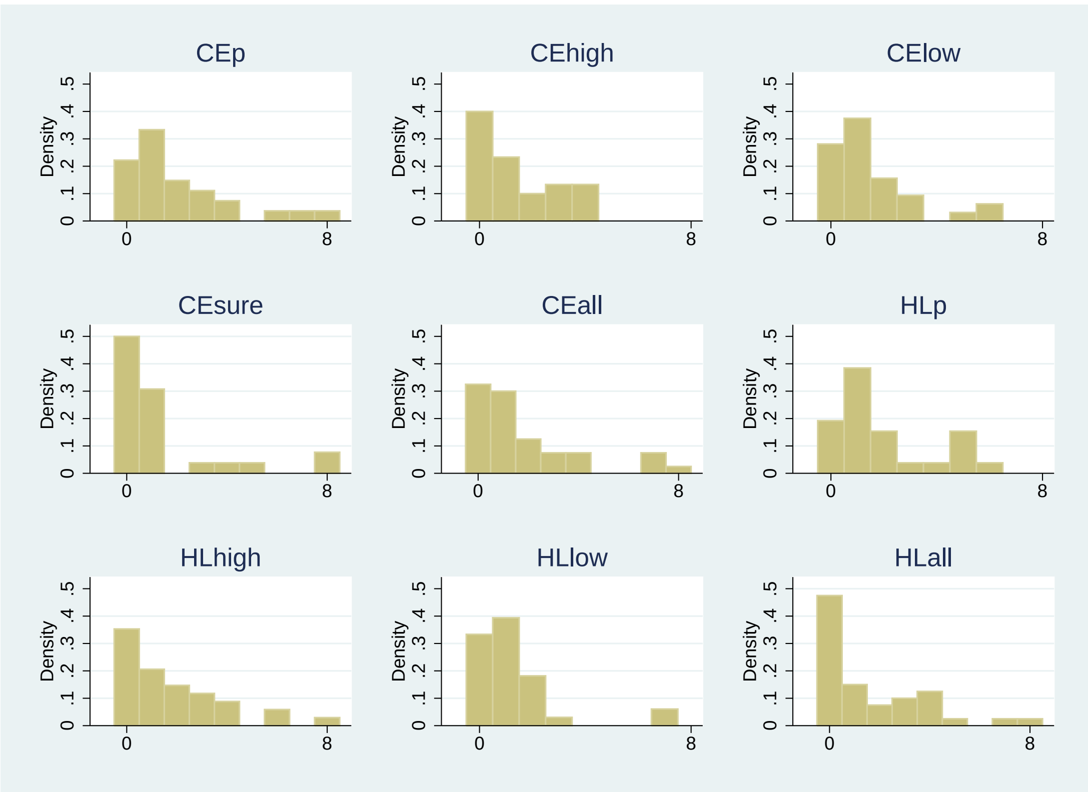


Figure 2 *Distributions of absolute differences in switching points between the first and second time a method is encountered*

was encountered. Four of the methods have a nonsignificant p-value: CEp, CEhigh, HLp, HLhigh.
A significant rank correlation (column 2 in Table 8) also indicates a stable risk measure, indicating a shift in the distribution, but no change in rank ordering. We see that three of those four methods have significant rank correlations with $p < 0.01$: CEp, CEsure and HLhigh.
A low absolute average deviation in answers is also an indicator of a stable risk measure, and the method with the lowest deviation is HLhigh, followed by HLp and CEsure.
Concerning the complexity, we see that a method that is perceived as less complex does not necessarily imply more stability in answers, as CElow has the lowest complexity rating, yet it is classified as unstable in all three categories. However, a general tendency of low complexity indicating more stability can be observed.

We conclude that HLhigh is the most stable method, as it is the only method that performs well in all three categories, with the overall distributions of switching points not significantly different, high rank correlations and low average deviation. CEp, CEsure and HLp perform well in two of the three categories.

3.4 Results Conclusion

In the benchmark games, as far as predictive power is concerned, we conclude that HLhigh has the highest predictive power with CEp being a close second.[14]
We also found a connection between stability and self-reported complexity of methods. We conclude that only the HLhigh (Drichoutis and Lusk, 2012), HLp (Holt and Laury, 2002), CEsure (Cohen et al., 1987; Abdellaoui et al., 2011) and CEp (Bruner, 2009) methods lead to consistent results within a 30-minute time frame, with the HLhigh method being by far the most consistent: The HLhigh method's performance is superior to the other methods in terms of deviations from normative predictions, overall and relative stability across time, etc. Our findings are further supported by the fact that we controlled for personality traits, various socioeconomic factors and cognitive reflection in our analyses.

[14] For non-incentivized surveys, our data shows that eliciting preferences with general and financial questions is a relatively good predictor compared to several incentivized elicitation methods.

Therefore, we conclude that while CEp also has high predictive power and good stability in answers, the most stable MPL method with the highest predictive power is HLhigh, which corresponds to a method derived by Drichoutis and Lusk (2012) in our alternative interpretation.

4 Conclusion

We conducted a holistic assessment and analysis of MPL risk elicitation methods that are present in the economics literature with a sophisticated experimental design using a unified framework and representation method. Previous findings in the literature (Dave et al, 2010; Crosetto and Filippin, 2013; etc.) indicate that between-method consistency of particular methods is low. We confirm this finding by extending our analysis to all popular methods. Furthermore, we show that distributional differences among methods are far from negligible. All this implies that an arbitrary selection of a particular risk assessment method can lead to differing results and misleading revealed preferences. Thus, it matters which elicitation method is used by researchers in order to control for risk and other preferences.

Our main takeaway is that we provide a suggestion which elicitation method to use based on objective criteria that assess within-method as well as between-method consistency and validity in real-world settings, and that suggestion is to use the HLhigh method by Drichoutis and Lusk (2012).

In a broader context, we researchers should take care when choosing which risk elicitation method they use. To be taken into consideration are the nature of the task they intend to control for, trade-off effects between noise, exactness and simplicity. Moreover, we find that changing both the potential rewards and probabilities is perceived as relatively complex by subjects and yields inconsistent results. The debate between changing the probabilities or rewards (Bruner, 2009) seems to be far from settled as one of the methods in each context (HLhigh and CEp) delivers promising results. In addition, our findings might provide guidance in implementing other elicitation methods in the MPL format - e.g. loss aversion (Gächter et al., 2010), willingness to pay (Kahneman et al., 1990), individual discount rates (Harrison et al., 2002) - in terms of whether to vary probabilities, rewards or using a certainty equivalent.

Acknowledgements

We thank the Vienna Centre of Experimental Economics (VCEE), University of Vienna, for allowing us to run our experiments in their laboratory, and the Austrian Science Fund (FWF) under project S10307-G14 for their grateful support. We thank Jean-Robert Tyran, Wieland Müller, Erik Wengström, Karl Schlag, Owen Powell, James Tremewan, Rupert Sausgruber, Thomas Stephens and Stefan Minner for reading our paper, improving it with their comments and suggestions and guiding us in the right direction over the course of the project several times.

References

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

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In: N.J. Salkind (Ed.), Encyclopedia of Measurement and Statistics (pp 103-107). Thousand Oaks, CA: Sage.

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

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E. (2008a). Lost in state space: Are preferences stable? International Economic Review 49(3), 1091-1112.

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E. (2008b). Eliciting risk and time preferences. Econometrica 76(3), 583-618.

Anderson, L. R., Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. Journal of Health Economics 27(5), 1260-1274.

Anderson, L. R., Mellor, J. M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. Journal of Risk and Uncertainty 39(2), 137-160.

Andersson, O., Holm, H. J., Tyran, J. R., Wengström, E. (2013). Deciding for others reduces loss aversion. Lund University Working paper 2013:30.

Andreoni, J., Harbaugh, W. T. (2010). Unexpected utility: Experimental tests of five key questions about preferences over risk. University of Oregon Economics Department Working Papers 2010-14.

Attema, A., Brouwer, W. (2012). In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks. MPRA Paper No. 36100.

Beck, H. B. (1994). An experimental test of preferences for the distribution of income and individual risk aversion. Eastern Economic Journal 20(2), 131-145.

Berg, J., Dickhaut, J., McCabe, K. (2005). Risk preference instability across institutions: A dilemma. Proceedings of the National Academy of Sciences of the United States of America 102(11), 4209-4214.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. American Journal of Agricultural Economics 62(3), 395-407.

Blais, A. R. Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. Judgment and Decision Making 1(1), 33-47.

Blanco, M., Engelmann, D., Koch, A. K., Normann, H. (2010). Belief elicitation in experiments: is there a hedging problem?. Experimental Economics 13(4), 412-438.

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities, Health Economics 11(5), 447-456.

Bocqueho, G., Jacquet, F., Reynaud, A. (2011). Expected utility of prospect theory maximizers - Results from a structural model based field experiment. Paper prepared for presentation at the EAAE 2011 Congress, Zürich, Switzerland

Bruner, D. M. (2009). Changing the probability versus changing the reward. Experimental Economics 12(4), 367-385.

Charness, G., Gneezy, U. (2010). Portfolio choice and risk attitudes - An experiment. Economic Inquiry 48(1), 133-146.

Charness, G., Gneezy, U., Imas, A. (2013). Experiential methods: Eliciting risk preferences. Journal of Economic Behavior and Organization 87(1), 43-51.

Charness, G., Viceisza, A. (2012). Comprehension and risk elicitation in the field: Evidence from Rural Senegal. University of California at Santa Barbara, Economics Working Paper Series qt5512d150.

Chiappori, P., Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. Journal of the European Economic Association 9(6), 1021-1052.

Cohen, M., Jaffray, J.-Y., Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. Organizational Behavior and Human Decision Processes 39(1), 1-22.

Cokely, E. T., Galestic, M., Schulz, E., Ghazal, S., Garcia-Retamero, R. (2012). Measuring risk literacy - The Berlin Numeracy Test. Judgment and Decision Making 7(1), 25-47.

Costa, P. T., McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, Inc.

Crosetto, C., Filippin, A. (2013). A theoretical and experimental appraisal of five risk elicitation methods. Jena Economic Research Papers, 2013-009.

Crosetto, C., Filippin, A. (2013). The "Bomb" risk elicitation task. Journal of Risk and Uncertainty 47(1), 31-65.

Cubitt, R. P., Starmer, C., Sugden, R. (1998). On the validity of the random lottery incentive system. Experimental Economics 1(2), 115-131.

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.

De Véricourt, F., Jain, K., Bearden, J. N., Filipowicz, A. (2013). Sex, risk and the newsvendor. Journal of Operations Management 31(1-2), 86-92.

Dohmen, T., 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.

Drichoutis, A., Lusk, J. (2012). Risk preference elicitation without the confounding effect of probability weighting. MPRA Paper No. 37776.

Dulleck, U., Fell, J., Fooken, J. (2014). Within-subject intra- and inter-method consistency of two experimental risk attitude elicitation methods. German Economic Review, forthcoming.

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

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.

Engelbrecht-Wiggans, R., Katok, E. (2008). Regret and feedback information in first price sealed-bid auctions. Management Science 54(4), 808-819.

Fausti, S., Gillespie, J. (2000). A comparative analysis of risk preference elicitation procedures using mail survey results. Vancouver, BC: 2000 Annual Meeting of Western Agricultural Economics Association.

Fellner, G., Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. Journal of Economic Psychology 28(3), 338-350.

Fischbacher, U. (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10(2), 171-178.

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives 19(4), 25-42.

Gächter, S., Johnson, E. J., Herrmann, A. (2010). Individual-level loss aversion in riskless and risky choices. IZA Discussion Papers 2961, Institute for the Study of Labor (IZA).

Goeree, J. K., Holt, C. A., Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. Games and Economic Behavior 45(1), 97-113.

Gonzalez, R., Wu, G. (1999). On the shape of the probability weighting function. Cognitive Psychology 38(1), 129-166.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In: K. Kremer, and V. Macho (Eds.), Forschung und wissenschaftliches Rechnen 2003 (pp. 79-93). GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung.

Harrison, G. W., Humphrey, S. J., Verschoor, A. (2010). Choice under Uncertainty: Evidence from Ethiopia, India and Uganda. The Economic Journal 120(543), 80-104.

Harrison, G. W., Johnson, E., McInnes, M. M., Rutström, E. E. (2005). Temporal stability of estimates of risk aversion. Applied Financial Economics Letters 1(1), 31-35.

Harrison, G. W., Lau, M. I., Rutström, E. E. (2009). Risk attitudes, randomization to treatment, and self-selection to experiments. Journal of Economic Behavior and Organization 70(3), 498-507.

Harrison, G. W., Lau, M. I., Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. American Economic Review 92(5), 1606-1617.

Harrison, G. W., List, J. A., Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. Econometrica 75(2), 433-458.

Harrison, G. W., Rutström, E. E. (2008). Risk aversion in the laboratory. In: J.C. Cox and G.W. Harrison (Eds.), Risk Aversion in Experiments (pp. 41-196). Research in Experimental Economics 12. Bingley, UK: Emerald.

Harrison, G. W., Rutström, E. E. (2009). Expected utility theory and prospect theory: one wedding and a decent funeral. Experimental Economics 12(2), 133-158.

Hershey, J. C., Kunreuther, H. C., Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. Management Science 28(8), 936-954.

Hey, J. D., Morone, A., Schmidt, U. (2009). Noise and bias in eliciting preferences. Journal of Risk and Uncertainty 39(3), 213-235.

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

Holt, A. C., Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. American Economic Review 95(3), 902-912.

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.

Kahneman, D., Knetsch, J. L., Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. Journal of Political Economy 98(6), 1325-1348.

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 - BART. Journal of Experimental Psychology: Applied 8(2), 75-84.

Lönnqvist, J-E., Verkasalo, M., Walkowitz, G., Wichardt, P. C. (2011). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. Cologne Graduate School Working Paper Series 02-03.

Lusk, J. L., Coble, K. H. (2005). Risk perceptions, Risk preference, and acceptance of risky food. American Journal of Agricultural Economics 87(2), 393-405.

Mador, G., Sonsino, D., Benzion, U. (2000). On complexity and lotteries' evaluation - three experimental observations. Journal of Economic Psychology 21(6), 625-637.

Murnighan, J. K., Roth, A. E., Schoumaker, F. (1988). Risk aversion in bargaining: An experimental study. Journal of Risk and Uncertainty 1(1), 101-124.

Rammstedt, B., John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the BIG Five Inventory in English and German. Journal of Research in Personality 41(1), 203-212.

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.

Sabater-Grande, G., Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games - An experimental test. Journal of Economic Behavior and Organization 48(1), 37-50.

Straznicka, K. (2012). Temporal stability of risk preference measures. GATE - Groupe d'Analyse et de Théorie Économique Lyon - St Étienne Working Paper No. 1236.

Tanaka, T., Camerer, C. F., Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. American Economic Review 100(1), 557-571.

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., De Vries, N. N. K. (2009). Probability information in risk communication: A review of the research literature. Risk Analysis 29(2), 267-287.

Von Gaudecker, H. M., van Soest, A., Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. American Economic Review 101(2), 664-694.

Von Gaudecker, H. M., van Soest, A., Wengström, E. (2008). Selection and mode effects in risk preference elicitation experiments. IZA Discussion Paper No. 3321.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. Health Economics 17(12), 1329-1344.

Wakker, P. P. (2010). Prospect theory: For risk and ambiguity. Cambridge, UK: Cambridge University Press.

Walker, J. M., Smith, V. L., Cox, J. C. (1987). Bidding be havior in first-price sealed bid auctions: Use of computerized Nash competitors. Economics Letters 23(3), 239-244.

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.

Wilkinson, L., Wills G. (2005). The grammar of graphics. Berlin, Germany: Springer

5 Appendix

Risk Preference Elicitation Methods

Note that all the preference elicitation methods described here are represented top-down for simplicity reasons. An example of a bottom-up representation of a particular method is provided in later in the Appendix.

Table 9: *CEp method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
1	8	-	-	0.62	4	0.38	12
1	8	-	-	0.564	4	0.4364	12
1	8	-	-	0.52	4	0.48	12
1	8	-	-	0.481	4	0.5192	12
1	8	-	-	0.447	4	0.553	12
1	8	-	-	0.411	4	0.589	12
1	8	-	-	0.374	4	0.626	12
1	8	-	-	0.323	4	0.677	12

Table 10: *CEhigh method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
1	8	-	-	0.5	4	0.5	10.63
1	8	-	-	0.5	4	0.5	11.16
1	8	-	-	0.5	4	0.5	11.7
1	8	-	-	0.5	4	0.5	12.32
1	8	-	-	0.5	4	0.5	13.04
1	8	-	-	0.5	4	0.5	14.07
1	8	-	-	0.5	4	0.5	15.75
1	8	-	-	0.5	4	0.5	20.31

Table 11: *CElow method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
1	9.5	-	-	0.5	6.1	0.5	12
1	9.5	-	-	0.5	6.61	0.5	12
1	9.5	-	-	0.5	6.89	0.5	12
1	9.5	-	-	0.5	7.09	0.5	12
1	9.5	-	-	0.5	7.24	0.5	12
1	9.5	-	-	0.5	7.38	0.5	12
1	9.5	-	-	0.5	7.51	0.5	12
1	9.5	-	-	0.5	7.66	0.5	12

Table 12: *CEsure method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
1	8.91	-	-	0.5	4	0.5	12
1	8.5	-	-	0.5	4	0.5	12
1	8.16	-	-	0.5	4	0.5	12
1	7.85	-	-	0.5	4	0.5	12
1	7.57	-	-	0.5	4	0.5	12
1	7.27	-	-	0.5	4	0.5	12
1	6.96	-	-	0.5	4	0.5	12
1	6.56	-	-	0.5	4	0.5	12

Table 13: CEall method

Left Option				**Right Option**			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.5	8	0.5	8	0.5	10	0.5	5.34
0.5	10	0.5	5.34	0.5	11	0.5	3.83
0.5	11	0.5	3.83	0.5	12	0.5	2.61
0.5	12	0.5	2.61	0.5	13	0.5	1.83
0.5	13	0.5	1.83	0.5	14	0.5	1.41
0.5	14	0.5	1.41	0.5	15	0.5	1.21
0.5	15	0.5	1.21	0.5	16.5	0.5	1.09
0.5	16.5	0.5	1.09	0.5	20.5	0.5	1.01

Table 14: HLp method

Left Option				**Right Option**			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.2	9	0.8	7.2	0.2	17.2	0.8	0.45
0.3	9	0.7	7.2	0.3	17.2	0.7	0.45
0.4	9	0.6	7.2	0.4	17.2	0.6	0.45
0.5	9	0.5	7.2	0.5	17.2	0.5	0.45
0.6	9	0.4	7.2	0.6	17.2	0.4	0.45
0.7	9	0.3	7.2	0.7	17.2	0.3	0.45
0.8	9	0.2	7.2	0.8	17.2	0.2	0.45
0.9	9	0.1	7.2	0.9	17.2	0.1	0.45

Table 15: *HLhigh method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.5	9	0.5	7.2	0.5	10.96	0.5	3.7
0.5	9	0.5	7.2	0.5	11.55	0.5	3.7
0.5	9	0.5	7.2	0.5	12.15	0.5	3.7
0.5	9	0.5	7.2	0.5	12.87	0.5	3.7
0.5	9	0.5	7.2	0.5	13.75	0.5	3.7
0.5	9	0.5	7.2	0.5	15.01	0.5	3.7
0.5	9	0.5	7.2	0.5	17.21	0.5	3.7
0.5	9	0.5	7.2	0.5	23.83	0.5	3.7

Table 16: *HLlow method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.5	16.09	0.5	7	0.5	3.7	0.5	17.2
0.5	15.3	0.5	7	0.5	3.7	0.5	17.2
0.5	14.41	0.5	7	0.5	3.7	0.5	17.2
0.5	13.35	0.5	7	0.5	3.7	0.5	17.2
0.5	12.18	0.5	7	0.5	3.7	0.5	17.2
0.5	10.85	0.5	7	0.5	3.7	0.5	17.2
0.5	9.29	0.5	7	0.5	3.7	0.5	17.2
0.5	7.35	0.5	7	0.5	3.7	0.5	17.2

Table 17: *HLall method*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.99	7.55	0.01	0	0.81	8.08	0.19	0
0.94	7.93	0.06	0	0.78	8.73	0.22	0
0.89	8.28	0.11	0	0.75	9.28	0.25	0
0.84	8.60	0.16	0	0.72	9.83	0.28	0
0.79	8.98	0.21	0	0.69	10.53	0.31	0
0.74	9.33	0.26	0	0.66	11.33	0.34	0
0.69	9.70	0.31	0	0.63	12.90	0.37	0
0.64	10.05	0.36	0	0.62	28.95	0.38	0

Example of Bottom-Up Representation

Table 18: *HLhigh method; Bottom-Up Appearance*

Left Option				Right Option			
p_1^L	π_1^L	p_2^L	π_2^L	p_1^R	π_1^R	p_2^R	π_2^R
0.5	23.83	0.5	3.7	0.5	9	0.5	7.2
0.5	17.21	0.5	3.7	0.5	9	0.5	7.2
0.5	15.01	0.5	3.7	0.5	9	0.5	7.2
0.5	13.75	0.5	3.7	0.5	9	0.5	7.2
0.5	12.87	0.5	3.7	0.5	9	0.5	7.2
0.5	12.15	0.5	3.7	0.5	9	0.5	7.2
0.5	11.55	0.5	3.7	0.5	9	0.5	7.2
0.5	10.96	0.5	3.7	0.5	9	0.5	7.2

Sample Screenshots

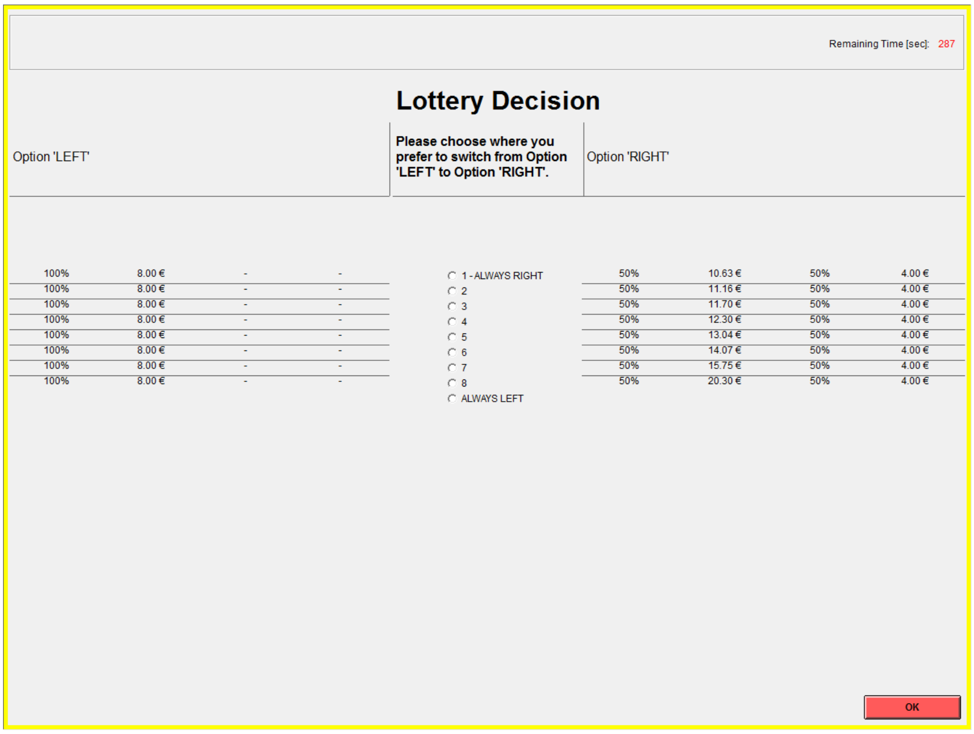


Figure 3 *Decision Making Screen for lotteries; subjects indicate which row they wanted to switch from the left to the right option by clicking one of the radio buttons in the middle*

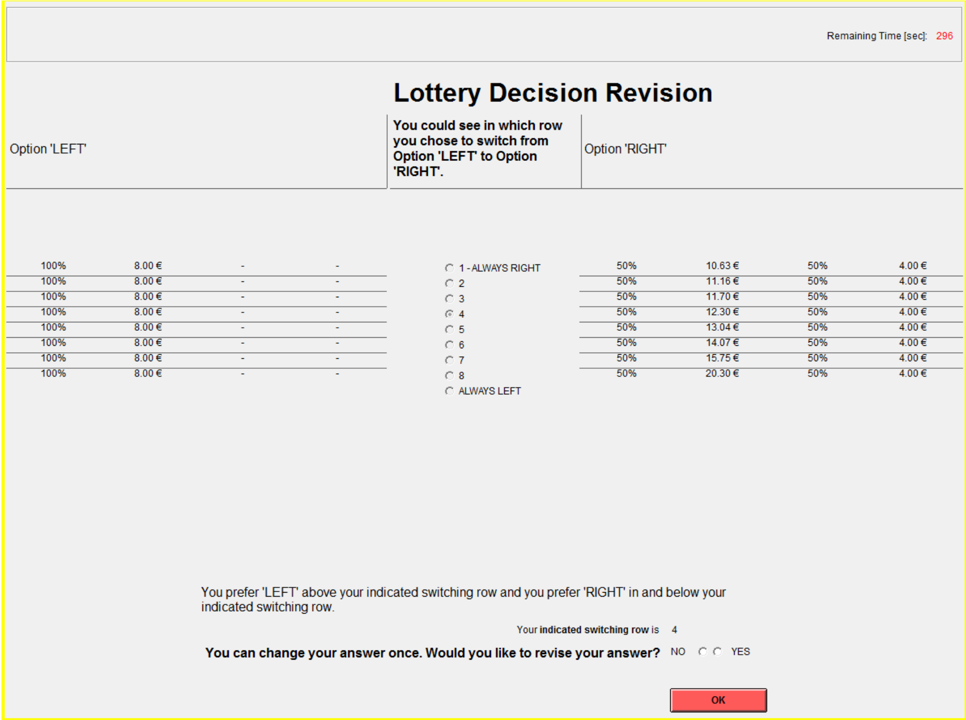


Figure 4 *Revision Screen for Lotteries; subjects indicated whether they want to revise their first decision*

Robustness Check (CARA)

In this section we report the regressions from Section 3.3.1 with the different assumption that a subject's utility function follows not CRRA but CARA, that is, it follows: $u(c) = -\frac{e^{-ac}}{a}$. We chose CARA for our robustness check as it is the most commonly assumed utility function after CRRA. In nearly all instances, the results do not change at all, significancewise. We conclude that the results are indifferent to the assumption of the underlying utility function. Our hypothesis that CRRA is a valid assumption is further supported by the fact that the adjusted R^2 coefficients of the regressions assuming CRRA are higher

Table 19: *Explanatory Power of the methods with CARA (OLS)*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall
	OLS coefficients (with controls)								
Auction	-5.23	6.83	5.47	1.29	3.67	-1.24	-0.01	5.03	4.81
	(.05)	(.05)	(.04)	(.02)	(.02)	(.02)	(.02)	(.04)	(.04)
Inventment Low	0.99	6.83*	4.57	1.65	9.08	2.37	10.62***	1.85	3.05
	(.00)	(.02)	(.00)	(.00)	(.01)	(.00)	(.08)	(.00)	(.00)
Investment High	6.90***	3.66	2.63	1.57	7.66	2.88	4.72	5.63*	0.18
	(.16)	(.10)	(.09)	(.08)	(.11)	(.09)	(.11)	(.12)	(.08)
Controls:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: Regression results with CARA instead of CRRA; the dependent variable is the outcome in one of the benchmark games, the independent variables are the outcome from one method and controls; the adjusted R^2 value for the regression can be found below the coefficients; Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

Robustness Check (Standard Correlations)

Table 20: *Correlation Coefficients*

	CEp	CEhigh	CElow	CEsure	CEall	HLp	HLhigh	HLlow	HLall	GQ
CEhigh	.46***									
CElow	.37***	.46***								
CEsure	.01	.18	.25*							
CEall	.02	.12	0	.16						
HLp	.13	.12	.15	.23**	-.08					
HLhigh	.07	.27**	.10	.26*	.12	.26*				
HLlow	.27**	.21*	.20	.17	-.04	.12	.10			
HLall	.17	.16	-.06	.04	.17	-.01	-.08	.03		
GQ	.12	.16	0	-.14	.09	-.07	.12	.02	.01	
FQ	.25**	.17	.27**	.19	.07	-.05	-.02	.20	.03	.46**

*Notes: Included are the nine methods and the questionnaires (GQ: general question; FQ:financial question); stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001*

How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods

Electronic Supplementary Material

Tamás Csermely[a] Alexander Rabas[b]

[a] Vienna University of Economics and Business, Department of Economics
[b] University of Vienna, Doctoral School of Economics

[a]Corresponding Author,
Welthandelsplatz 1, D4, 1020 Vienna, Austria
tamas.csermely@wu.ac.at
Tel.: +43 1 31336 4264

[b]Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
alexander.rabas@univie.ac.at

Instructions

General Instructions

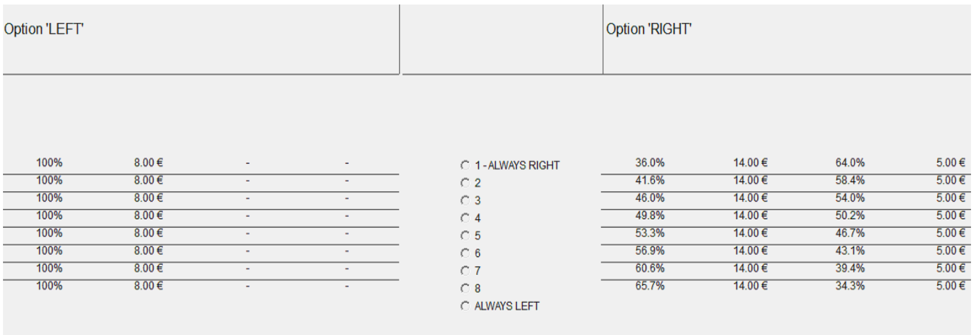
Welcome to this experiment in decision making.

You will be asked to make a series of choices that will affect your payment at the end of the experiment. Please pay close attention to the instructions, and if you have any questions raise your hand and an employee of the lab will help you with any questions you might have.

Also, you will be asked to fill out a short questionnaire, where your answers are not relevant for your payoff. All your decisions and answers during the experiment will stay completely anonymous to everyone.

In 'Part 1', you will make decisions in twelve different situations, where in each situation you will make one choice.

A sample decision screen is provided above and on the piece of paper



Option 'LEFT'					Option 'RIGHT'			
100%	8.00 €	-	-	1 - ALWAYS RIGHT	36.0%	14.00 €	64.0%	5.00 €
100%	8.00 €	-	-	2	41.6%	14.00 €	58.4%	5.00 €
100%	8.00 €	-	-	3	46.0%	14.00 €	54.0%	5.00 €
100%	8.00 €	-	-	4	49.8%	14.00 €	50.2%	5.00 €
100%	8.00 €	-	-	5	53.3%	14.00 €	46.7%	5.00 €
100%	8.00 €	-	-	6	56.9%	14.00 €	43.1%	5.00 €
100%	8.00 €	-	-	7	60.6%	14.00 €	39.4%	5.00 €
100%	8.00 €	-	-	8	65.7%	14.00 €	34.3%	5.00 €
				ALWAYS LEFT				

Figure 1 *This is the first sample screen subjects saw; note that it does not match any of the nine methods.*

in front of you. Please feel free to take notes on this paper.

On each screen you see two columns ('LEFT' and 'RIGHT'). In each column you have eight rows with different payment possibilities. You have to decide which option you prefer in each row.

For example on the screen provided above, in the first row you will have to choose between:
'LEFT': receive 8€ with 100% probability
'RIGHT': receive 14€ with 36.0% probability or 5€ with 64.0% probability

In the second row, you would make a choice between:
'LEFT': receive 8€ with 100% probability
'RIGHT': receive 14€ with 41.6% probability or 5€ with 58.4% probability

In the center of the screen you will find a number of radio buttons. You can only click one of those buttons on each screen. This button indicates at what row you want to switch from 'LEFT' to 'RIGHT'.

If you choose row 3, that means you prefer 'LEFT' for the first 2 rows,

Option 'LEFT'				Suppose that your earnings depend on the following lottery.	Option 'RIGHT'			
24.0%	11.36 €	76.0%	0.00 €	1 - ALWAYS RIGHT	23.0%	12.00 €	77.0%	0.00 €
34.0%	9.76 €	66.0%	0.00 €	2	27.0%	11.36 €	73.0%	0.00 €
44.0%	8.16 €	56.0%	0.00 €	3	29.7%	10.93 €	70.3%	0.00 €
54.0%	6.56 €	46.0%	0.00 €	4	32.3%	10.51 €	67.7%	0.00 €
64.0%	4.96 €	36.0%	0.00 €	5	36.6%	9.83 €	63.4%	0.00 €
74.0%	3.36 €	26.0%	0.00 €	6	42.5%	8.88 €	57.5%	0.00 €
84.0%	1.76 €	16.0%	0.00 €	7	56.0%	6.72 €	44.0%	0.00 €
94.0%	0.16 €	6.0%	0.00 €	8	91.7%	1.01 €	8.3%	0.00 €
				ALWAYS LEFT				

Figure 2 *Example Screen with colorcoding*

but you prefer 'RIGHT' for rows 3-8.
If you choose row 6, that means you prefer 'LEFT' for the first 5 rows, but you prefer 'RIGHT' for rows 6-8.
If you choose row 1 - ALWAYS RIGHT, that means you prefer 'RIGHT' in every row.
If you choose row ALWAYS LEFT, that means you prefer 'LEFT' in every row.

When you are finished making your decision for a screen, click 'OK' and you will get to the next screen where you will see your choice again. In case you are not satisfied with your choice, you can change your choice once if you wish to. Your second choice is final and cannot be changed afterwards.

After each decision screen, we will ask you how difficult it was for you to make a decision on the previous screen. These questions do not affect your payment. Still, we ask you to answer truthfully.

For your payoff of Part 1, one screen of the twelve is randomly selected by the computer. The computer will also select one of the rows at random.

In this row, you have chosen 'LEFT' or 'RIGHT'. Finally the computer will randomize between the two possible outcomes based on the given probabilities.

For example, the computer chooses at random the following row to be relevant for your payoff:
'LEFT': 'receive 20€ with 70% probability or 5€ with 30% probability.'
'RIGHT': 'receive 15€ with 60% probability or 10€ with 40% probability.'
You have chosen left in that particular row. Therefore, you either get 20€ or 5€, but the probability to get 20€ is higher.

Every screen and every row on each screen has an equal chance to be chosen by the computer to be relevant for your payoff. Considering that every decision you make matters, we advise you to think carefully about each decision you make.

After you made these twelve decisions, 'Part 2' will start. Instructions for 'Part 2' will be given on the screens themselves. One situation in 'Part 2' will be randomly selected to affect your payment. 'Part 1' and 'Part 2' are completely independent of each other.

Your final payoff will then be the sum of your payoffs from 'Part 1' and 'Part 2'.

If you have no questions, please click 'OK' to answer a couple of control questions. These questions will make sure that you have understood the setup. You cannot commence with the experiment unless you answer the control questions correctly.

The experiment will start afterwards!

If you have questions, please raise your hand at any time and an experimenter will provide assistance.

Auction Instructions

You will now participate in an auction against a computer opponent over a good that has a value of 20€. You can bid any amount from 0€ to 20€, and you can specify your bid down to the exact Cent.

The computer will bid a random number from 0€ to 20€, down to the exact Cent, and each number has an equal probability to be chosen.

If your bid is higher than the computer's, you will get 20€ minus your bid as your payoff. If your bid is lower than the computer's, you will get 0€.

If your bids are tied, the winner of the auction is selected randomly and you will receive the payoff of 20€ minus your bid with 50% probability.

Example 1: If you bid 12.41€ and the computer bids 16.53€, your payoff is 0€ as the computer's bid is higher than yours.
Example 2: If you bid 18.8€ and the computer bids 0.17€, your payoff is 1.2€ (=20-18.8) from this auction, as your bid is higher than the computer's.

Remember, you can bid any amount from 0€ to 20€. If you win the auction, your payoff is 20€ minus your bid.

Now, please type in how much you want to bid for the good.

Investment Game Instructions

You will now have the opportunity to invest an endowment of 10.00€.

There are two assets you can invest in: STOCKS and BONDS.

The amount you invest in bonds does not give returns. You will get the amount you invested as your payoff for sure.

STOCKS: STOCKS can have higher gains than BONDS, but are more risky. The amount you invest in STOCKS has a 50% chance to be multiplied by 1.5, and a 50% chance to be lost.

You can freely allocate your endowment of 10.00€ between the two assets, down to the exact Cent.

Example 1: You invest 10€ in BONDS and 0€ in STOCKS. Your payoff will be 10€.
Example 2: You choose to invest 2.58€ in BONDS and 7.42€ in STOCKS.

Your payoff will either be 13.71€ (=7.42*1.5+2.58) with 50% probability, or 2.58€ with 50% probability.
Example 3: You choose to invest all 10€ into STOCKS. Your payoff will either be 15€ (=10*1.5) with 50% probability, or 0€ with 50% probability.

Remember:
You will receive the amount you invest in BONDS as your payoff.

You will receive the amount you invest in STOCKS times 1.5 with 50% probability, and 0€ with 50% probability.

Please choose how much you want to invest in STOCKS (the rest of your endowment will be invested in BONDS):