From Measurements to Measures: Learning Risk Preferences under Different Risk Elicitation Methods

Caferra, Rocco and Morone, Andrea and Pierno, Donato

Department of Law and Economics, UnitelmaSapienza - University of Rome, Piazza Sassari 4, 00161, Roma, Italy, University of Bari Aldo Moro, Bari, Italy, Department of Economics and Finance, Tor Vergata - University of Rome

13 June 2024

From Measurements to Measures: Learning Risk Preferences under Different Risk Elicitation Methods

Rocco Caferra
UnitelmaSapienza University of Rome

Andrea Morone
University of Bari Aldo Moro

Donato Pierno
Tor Vergata University of Rome

June 13, 2024

Abstract

This study explores how people learn and adapt their risk preferences using different elicitation methods, challenging the neoclassical theory that suggests preferences are fixed. Instead, we show that preferences can change. However, we aim to explain whether the observed changes are due to a real change in the measure, i.e. individuals' risk preferences, or if they are attributable to the limitations of the measurement tool, i.e. the specific risk elicitation method employed. We use a detailed experimental design to examine the stability and consistency of risk preferences using a hands-on learning experience. Our main goals are to assess how consistent risk choices are, understand how preferences remain stable or change over time, and evaluate the effectiveness of different elicitation methods like the Multiple Price List and Ordered Lottery Selection ones. On the one hand, results demonstrate that risk preferences are variable and adaptable, and this can be partly due to the role of experience-based learning. On the other hand, we observe how Multiple Price List methods, even if more complex, are more accurate in identifying risk preferences and then in improving measurement stability and accuracy.

Key words: Risk preferences; Experiments; Elicitation Methods; Learning.
JEL Classification: D90; D81

1 Introduction

Are Risk Preferences Stable? Changes in risk preferences could explain the variation in several economic outcomes. The way we choose to take risks defines various aspects of real life, including the labor market, where risk-averse individuals are less likely to be self-employed, and investments, where risk-averse individuals might be less inclined to invest in stock markets.

This study explores how individuals discern and adjust their risk attitudes through different elicitation methods. By integrating insights from experimental economics, decision theory, and behavioral finance, we aim to unravel the dynamics of personal risk preference learning and adaptation. Neoclassical theory claims that people have rational and fixed preferences; however, the consistency over time of such preferences have been widely examined in dynamic decision theory. The debate towards the dynamic consistency of such preference have originated different strand of discussion (see McClennen, 1990) recognized crucial in economics and philosophy. Experimental evidence (both laboratory and field) suggests that preferences can change over time, challenging their dynamic consistency. This study builds on the works of Bradbury et al. (2015), Ert and Haruvy (2017), and Charness et al. (2023), which have shown that risk preferences can shift based on experience-based learning.

The emerging "Risk Elicitation Puzzle" (Pedroni et al., 2017, then revised by Holzmeister and Stefan, 2021), well-known in experimental literature, refers to the difficulty in accurately identifying subjects' risk preferences. Based on this, the present work focuses on studying and comparing various experimental tools used to assess risk aversion: *Are changes in preferences related to actual changes in the measures or to limitations of the measurement tool?*

As noted by Charness et al. (2013), different experimental techniques vary in their complexity for measuring risk. The simplest way to declare one's preferences is by responding to survey questions about individual risk-taking propensity (Dohmen et al., 2011). However, since this method is not incentivized, it is straightforward but fails to address the intention-behavior gap, where stated preferences do not match preferences in real-choice contexts with actual outcomes. You can ideally believe and declare that you are a person who, in reality, does not reflect your choices.

Moving a step forward from unincentivized surveys, the experimental methods proposed by Eckel and Grossman (2008) and Charness et al. (2023), known as Ordered Lottery Selections (OLS), require individuals to make simple choices to identify their risk levels. In this case, the subject chooses in a one-shot manner which type of lottery they want to play and therefore what level of risk they want to take on. This method makes progress over surveys of stated risk preferences, as it allows for the evaluation of choices in incentivized contexts and the estimation of a risk aversion parameter. However, subjects risk preferences may be more sophisticated than a single isolated choice, and it may be more appropriate to ask the subject to make multiple choices in risky contexts to better define the shape of their risk propensity. In this case, tools like Multiple Price Lists (MPL) (Holt and Laury, 2002) ask subjects to make various risky choices, providing multiple confirmations of their risk tolerance. Consequently, these tools may be more effective in consistently identifying subjects' risk preferences.

Let us clarify with an example: choosing to take a risk occasionally with methods proposed by Eckel and Grossman (2008) does not necessarily mean the subject is always willing to do so. Going to a casino once does not indicate that one subject is consistently risk-loving over time. With the MPL methods, in order to be considered a risk-lover, subjects must repeatedly choose high-risk options within the same framework. Coming back to the casino example, they would need to

decide to enter and bet multiple times. This approach, even if complex, can more reliably track risk preferences.
As a result, the first issue related to measuring risk preferences is linked to the choice and variability of the tool. We can now consider a second aspect related to risky choices: variations in the measure. In this case, behavioral studies consider a crucial factor related to changes in risk preferences: experience. Various contributions in the literature, such as Ert and Haruvy (2017), Charness et al. (2023) and Charness and Chemaya (2023) , study how preferences evolve over time and are not static. In their cases, the analyzed change varies based on the technique used. Charness et al. (2023) use a variation of the OLS, while Ert and Haruvy (2017) use repetitions of the MPL (Holt and Laury, 2002). While the former finds an increase in risk tolerance linked to experience, the latter identifies a shift toward risk neutrality. These reported examples support the research idea proposed in the present work. Indeed, we notice changes in the measure that seem to be also linked to the measurement tool used. It becomes difficult to understand whether the changes in choices are due to experience or simply "corrections" of errors or adjustments of initial choices.
We aim to explain whether the observed changes are due to a real change in the measure, i.e., individuals' risk preferences, or if they are attributable to the limitations of the measurement tool, i.e., the specific risk elicitation method employed. We use a detailed experimental design to examine the stability and consistency of risk preferences using a hands-on learning experience. Our main goals are to assess how consistent risk choices are, understand how preferences remain stable or change over time, and evaluate the effectiveness of different elicitation methods like the MPL and OLS. On the one hand, results demonstrate that risk preferences are variable and adaptable, partly due to the role of experience-based learning. On the other hand, we observe how MPL methods, even if more complex, are more accurate in identifying risk preferences and then in improving measurement stability and accuracy.
In addition, we further disentangle the results by considering one of the most debated stereotypes in the analysis of risk preferences: the gender effect. The mainstream result identified in behavioral studies points to a higher risk aversion among women (Charness and Gneezy, 2012) and generally less daily experience with risky choices (Harris and Jenkins, 2006). Consequently, the advantage in terms of knowledge and experience might be reflected in the level of risk tolerance. It is interesting to observe the potential gender gap in risk preferences under repeated measurement with different tools, where both individuals are given the opportunity to gain experience in risky contexts.

All these aspects will be analyzed in the present work. Section 2 reviews the existing literature on the various mentioned methods, specifically describing those that will be used in the conducted experiment. Section 3 presents the experimental design, Section 4 the results, and Section 5 concludes.

2 Risk Elicitation methods background

Numerous experimental methodologies have been developed in economics to elicit and estimate risk attitudes. These methodologies vary in complexity: complex methods require more understanding and mathematical reasoning from participants and multiple confirmation choices to be classified based on your risk attitude, while simpler methods are more accessible. This experimental investigation employs different schemes for eliciting risk preferences, including the MPL and the OLS, along with a self-reported willingness to take risks through a questionnaire. We provide a detailed look at the factors that shape risk preferences and how these can be changed through specific learning experiences, under the lens of different elicitation methods that can accurately disentangle the learning effect from simple adjustment or mistake due to the intrinsic characteristics of the experimental tool used. In this section we review the main advances in the experimental research related to the risk elicitation methods (Sec 2.1), and then the main features of the methods used in the experiment conducted are reported (Sec 2.2).

2.1 The risk-elicitation puzzle

Accurately identifying individual risk attitudes is essential for economic decision-making, as highlighted by Dohmen et al. (2011). This issue has been further analyzed by Attanasi et al. (2018), who propose a multi-choice elicitation task to bridge the gap between lottery-based and survey-based risk attitudes, offering a more sophisticated understanding that captures a broader spectrum of decision-making scenarios.

Such advancements underscore the variability and potential fluidity of risk preferences, suggesting that personal and simulated experiences can significantly influence risk attitudes. This concept of variability in risk preferences is further explored by Ert and Haruvy (2017) and supported by studies such as those by Bradbury et al. (2015) and Kaufmann et al. (2013), which demonstrate how experiences and exposure to risk representations can shift investment risk appetites.

Comprehending risk preferences introduces the "risk elicitation puzzle"[1] , highlighting inconsistencies across various elicitation methods and questioning the external validity of laboratory measures. Charness et al. (2020) investigate this issue, examining the predictive power of risk behavior both in and outside the laboratory. Ert and Haruvy (2017) and Charness et al. (2023) demonstrate that preferences can change over time following experience, highlighting the dynamic nature of risk preferences. At the same time, the direction of the change seems to depend on the risk-elicitation method used.

This raises the question of whether the change in the measure depends on the real changes of the phenomenon being measured, i.e., the acquired experience-based learning of risk preferences, or on the ability to measure the phenomenon, i.e. the accuracy of the risk-elicitation method.

To capture risk preferences, we utilize qualitative measures akin to those proposed by Attanasi et al. (2018), paralleling self-assessment techniques from Dohmen et al. (2011) and the Global Preference Survey introduced by Falk et al. (2018). This allows us to explore individuals' general willingness to take risks and how they may vary based on different learning experiences and elicitation methods.

As anticipated in the introduction, the main idea is to introduce different methods with different level of complexity and accuracy, moving from simple non incentivized questions, where the

[1] Preliminary identified by Pedroni et al. (2017), subsequently revised by Holzmeister and Stefan (2021).

hypothetical bias can result in a discrepancy between stated and revealed preferences, to more accurate methods, from OLS to MPL. We employ a ten-row list pair of lotteries with the same payoff structure as Attanasi et al. (2014) (hereafter MPL10). Presented as a test for risk aversion where each subject faces ten choices, one for each row, between two lotteries (A or B) with probabilities ranging from 0.1 to 1. Then, a variation of the lists of paired lotteries, as proposed by Attanasi et al. (2014) (hereafter MPL20), where two lotteries are presented, doubling the probabilities of the outcomes. Employing this variation, we can have a more precise picture of the extent of the drop in the risk preference parameter as reported by Ert and Haruvy (2017), using a ten-row choice list. The second risk preference elicitation method, OLS, is presented as a sequence of lotteries, where the subjects are asked to pick one. The first is a safe option, while the other alternatives are increasing in expected value and variance, from lottery one to the next, using the same list of lotteries and outcomes in Eckel and Grossman (2008) (hereafter EG) is implemented in our experiment. A variation of the latter mechanism is presented by Charness et al. (2023), which implements a new attractive gamble (moving from five to six), which allows the capture of risk-seeking behavior.

2.2 Risk Elicitation Methods

In this section, we will discuss three prominent elicitation methods, presenting them according to their degree of complexity: Questionnaires, the OLS, and MPL. These methods vary in complexity, risk classification assignment mechanism, the type of data they generate, and their applicability in different experimental settings, offering a comprehensive toolkit for researchers studying risk preferences.

Let us start with the simplest risk assessment method. Commonly used in psychology, *Questionnaires* rely on self-reported measures of willingness to take risks. As highlighted by Charness et al. (2013), individual risk preferences, elicited through this method, are strong indicators of tendencies to engage in risky behaviours across various domains, from financial to health-related decisions. The structure of the question typically involves rating one's willingness to take risks. Dohmen et al. (2011) employed this method, making it the simplest way to elicit individual risk preferences. Following the wording used by Bernasconi et al. (2014) and Attanasi et al. (2018), our question is: "On a scale from 1 to 10, how would you rate your attitude towards risk? Are you a person who always avoids risk or someone who loves taking risks?" Here, 1 corresponds to "I always choose the safest option and try to avoid any possible risk," and 10 refers to "I love risk and always choose the riskier alternative."

Incentivized experimental methods allow the passage from stated preferences to revealed preferences, involving risky choices with real economic gains. The simplest design, based on *OLS*, as reported by Harrison and Rutström (2008), was developed by (Binswanger, 1980, Binswanger, 1981) and later popularized by Eckel and Grossman (2008). It is designed to easily capture and estimate risk preference parameters by asking subjects to make a single choice among a number of gambles. In our experiment, we employ five and six gambles. Specifically, we use the five-gamble choice for treatment 1 and the six-gamble choice for Treatment 0. Each gamble is presented with a 50% probability assigned to both the low and high payoff. The first row represents a sure gamble; from the first to the last, the gambles are characterized by increasing expected payoffs and standard deviations (risk).

Our risk elicitation choices with six rows are presented in Table 1. The second set of risk elicitation choices in our study involves five gambles, following the scheme of Eckel and Grossman (2008), and is presented in Table 2.

Choice (50/50 Gamble)	Low payoff	High payoff	Expected return	Standard deviation	Fraction of Subjects (%)
Gamble 1	28	28	28	0	6,06
Gamble 2	24	36	30	6	11,11
Gamble 3	20	44	32	12	33,33
Gamble 4	16	52	34	18	18,18
Gamble 5	12	60	36	24	17,17
Gamble 6	2	66	34	32	14,14

Table 1: Treatment 0, Charness et al., 2023 gamble choices

Choice (50/50 Gamble)	Low payoff	High payoff	Expected return	Standard deviation	Fraction of Subjects (%)
Gamble 1	16	16	16	0	4,30
Gamble 2	12	24	18	6	16,40
Gamble 3	8	32	20	12	32,38
Gamble 4	4	40	22	18	22,70
Gamble 5	0	48	24	24	23,80

Table 2: Treatment 1, Eckel and Grossman, 2008, gamble choices

By increasing the complexity of the task, we can now talk about *MPL elicitation methods*. According to Harrison and Rutström (2008), this design was initially employed by Miller et al. (1969) and later popularized in the influential paper by Holt and Laury (2002). It is now the most commonly used experimental method (EM) in economics. In this design, each subject is presented with a list of 10 paired lottery choices. The gambles are labelled as Option A (safe choice) and Option B (risky choice), arranged in rows. Participants must choose their preferred gamble for each of the 10 rows. The gambles associated with each option have constant payoffs but changing probabilities. Ideally, a rational subject in the first decision row should prefer the higher payoff by choosing Option B. For the first six rows, the expected payoff of Option B is greater than that of Option A. The constant payments are 12 or 10 Experimental Currency Units (ECU) for Option A, and 22 or 0.50 ECU for Option B. Starting from a certain probability in the first decision row, the higher payoff for both options is paid within a range of 0.9 to 0.1. In our MPL scheme, the expected value for Option A ranges from 12 to 10.20, while Option B ranges from 22 to 2.65. Our ten ordered choices are reported in Table 3. We employ the same payoff scheme as Attanasi et al. (2014), and only one row will be drawn for payment. We also use a variation of the MPL, initially presented in Attanasi et al. (2014) and further developed in Attanasi et al. (2018), which allows subjects to make choices within a wider probability range. This variation presents 20 decision rows, doubling the outcome probabilities. The MPL design characterizes our treatments 2 and 3. In the former, probabilities are framed in 10 percentage points, while in the latter, they are framed in 5 percentage points. Holt and Laury (2002) suggest using the number of safe choices as an aggregate measure. We report the corresponding risk-averse classification and the percentage of subjects for each number of safe choices in Table 4. Similarly, we consider the percentage of safe choices made by subjects in Attanasi et al. (2014), presenting their results in three categories; see Table 6.

Decision	Option A	Option B
1	20/20 of 12; 0/20 of 10	20/20 of 22; 0/20 of 0,50
2	18/20 of 12; 2/20 of 10	18/20 of 22; 2/20 of 0,50
3	16/20 of 12; 4/20 of 10	16/20 of 22; 4/20 of 0,50
4	14/20 of 12; 6/20 of 10	14/20 of 22; 6/20 of 0,50
5	12/20 of 12; 8/20 of 10	12/20 of 22; 8/20 of 0,50
6	10/20 of 12; 10/20 of 10	10/20 of 22; 10/20 of 0,50
7	8/20 of 12; 12/20 of 10	8/20 of 22; 12/20 of 0,50
8	6/20 of 12; 14/20 of 10	6/20 of 22; 14/20 of 0,50
9	4/20 of 12; 16/20 of 10	4/20 of 22; 16/20 of 0,50
10	2/20 of 12; 18/20 of 10	2/20 of 22; 18/20 of 0,50

Table 3: Treatment 2, Holt and Laury, 2002 gamble choices

Number of safe choices	Risk preference classification	Proportion of choices (%)
0-1	highly risk loving	0.01
2	very risk loving	0.01
3	risk loving	0.06
4	risk neutral	0.26
5	slightly risk averse	0.26
6	risk averse	0.23
7	very risk averse	0.13
8	highly risk averse	0.03
9-10	stay in bed	0.01

Table 4: Risk averse classification based on Holt and Laury (2002)

Decision	Option A	Option B
1	20/20 of 12; 0/20 of 10	20/20 of 22; 0/20 of 0,50
2	19/20 of 12; 1/20 of 10	19/20 of 22; 1/20 of 0,50
3	18/20 of 12; 2/20 of 10	18/20 of 22; 2/20 of 0,50
4	17/20 of 12; 3/20 of 10	17/20 of 22; 3/20 of 0,50
5	16/20 of 12; 4/20 of 10	16/20 of 22; 4/20 of 0,50
6	15/20 of 12; 5/20 of 10	15/20 of 22; 5/20 of 0,50
7	14/20 of 12; 6/20 of 10	14/20 of 22; 6/20 of 0,50
8	13/20 of 12; 7/20 of 10	13/20 of 22; 7/20 of 0,50
9	12/20 of 12; 8/20 of 10	12/20 of 22; 8/20 of 0,50
10	11/20 of 12; 9/20 of 10	11/20 of 22; 9/20 of 0,50
11	10/20 of 12; 10/20 of 10	10/20 of 22; 10/20 of 0,50
12	9/20 of 12; 11/20 of 10	9/20 of 22; 11/20 of 0,50
13	8/20 of 12; 12/20 of 10	8/20 of 22; 12/20 of 0,50
14	7/20 of 12; 13/20 of 10	7/20 of 22; 13/20 of 0,50
15	6/20 of 12; 14/20 of 10	6/20 of 22; 14/20 of 0,50
16	5/20 of 12; 15/20 of 10	5/20 of 22; 15/20 of 0,50
17	4/20 of 12; 16/20 of 10	4/20 of 22; 16/20 of 0,50
18	3/20 of 12; 17/20 of 10	3/20 of 22; 17/20 of 0,50
19	2/20 of 12; 18/20 of 10	2/20 of 22; 18/20 of 0,50
20	1/20 of 12; 19/20 of 10	1/20 of 22; 19/20 of 0,50

Table 5: Treatment 3, Attanasi et al., 2014 gamble choices

Number of safe choices	Risk preference classification	Proportion of choices (%)
0-10	risk loving	0.19
10	risk neutral	0.16
11-20	risk averse	0.65

Table 6: Proportion of safe choice in Attanasi et al. (2014).

3 Experimental design and hypotheses

3.1 Experimental Design

The experiment was conducted in May 2024 at the ESSE laboratory at the University of Bari Aldo Moro.
We recruited 88 participants in two different sessions; to obtain an appropriate sample size, each subject randomly participated in two treatments[2]. We controlled for order and learning effects, finding no differences in aggregate behavior considering the order in which subjects faced the two tasks.[3] For each subject, one of the two treatments was randomly selected for payment to avoid the hedging effect, with the average payment being approximately 5 euros. The instructions for each treatment are provided in Appendix A. As anticipated, we have 2 treatments for the OLS and another 2 for the MPL mechanism. In addition, we use a measure of self-reported risk to study how it correlates with the real behavior of the subjects, analyzing the intention-behavior gap also in light of the gender effect.
Treatment 0 (T0) replicates Charness et al. (2023), as an adaptation of Dave et al. (2010), with six-row gamble choices that are relatively easy with 50-50 gambles (see Table 1). **Treatment 1 (T1)** replicates the elicitation method of Eckel and Grossman (2008), consisting of five-row gamble choices. The crucial difference between these two treatments lies in the last row of our treatment 0, which is really attractive for risk-seeking subjects. **Treatment 2 (T2)** refers to the most commonly employed MPL method, which is Holt and Laury (2002), characterized by ten decision rows (see table 3). **Treatment 3 (T3)** is a variant proposed by Attanasi et al. (2014), allowing us to observe choices in twenty decision rows where the assigned outcome probabilities are framed in 5 percentage points rather than 10 (see table 5 and Appendix A).
Following Charness et al. (2023), experimental subjects were accurately informed by the experimenter about the tables of gambles, with the relative treatment scheme for high (low) payoff. The OLS design, characterized by 50%-50% probability gamble, has a simple distinction between even(odd) results from a dice rolled; while the MPL design ranging from 5% to 100% probabilities (see Appendix A). We used twenty-face dice for all four treatments, providing one dice for a subject.
The two sessions are held by the same experimenter, who reads and explains the instructions of the experiment. Each treatment is formed by four different phase:

- **Initial Risky Choice.** In the first phase, subjects, after receiving and reading the instructions, are asked to make a choice in order to elicit their first measure of risk attitude, according to the assigned treatment;
- **Learning phase.** In the second phase consists of 24 unpaid learning periods. These 24 periods are characterized by 12 periods with pre-selected choices; here, each person rolled a dice and reports on a record sheet the dice outcome and the would-be payoff, while for the 12 free choices, each subject chooses the preferred gamble(s) and follows all the previous steps.
- **Final Risky Choice.** In the third phase, subjects are asked, as made in the first phase, to

[2]This allows us to have 164 total observations.
[3]For each treatment, we controlled for the probability of changing risk preferences in each treatment considering the subgroups that played it as the first task and those that played it as the second, not finding statistically significant differences (Kolmogorov-Smirnov test: T0, p-value: .999; T1, p-value: .944; T2, p-value: .292; T3, p-value: .945.

make their choice

- **Stated Preferences (Ex-post Questionnaire).** Finally the last phase is a survey where they are asked to report their self-reported measure of risk attitude, and demographics information.

After these four phases each subjects comes in front of the desk, first roll the dice to determine which of the two treatments is valid for payment, then roll the dice to determine if the first or the third phase will be paid, finally roll the dice to determine the gamble outcome.
More deatils can be found in the Online Supplementary Materials.

3.2 Hypotheses

As previously mentioned, neoclassical theory states that subjects do not change their risk preferences over time. The stability of risk preferences has been criticized in subsequent experimental studies, noting how experience-based learning (Bradbury et al., 2015; Charness et al., 2023; Eckel et al., 2009) can cause subjects to reconsider their choices. In this work, we consider three key aspects of this phenomenon: i) the measurement of preferences and their potential temporal variation due to learning; ii) the accuracy of the preference measurement and elicitation tool; iii) the heterogeneity of the subjects' characteristics. Based on this, we can formulate our hypotheses considering the three key aspects.

Measures. Individuals change their preferences over risk from initial to final choices. Following existing studies, this can be (partly) attributed to experience. Experimental evidence shows that preferences towards risk can change after experiencing shocks (Eckel et al., 2009). In their study, Bradbury et al. (2015) examine how investment decisions under risk are influenced by an experience-based learning approach, suggesting that people reconsider their investments in favor of higher expected returns, choosing riskier financial products. Another important piece of evidence, closely related to our work, comes from Ert and Haruvy (2017). They explored the effect of an experience-based learning approach through 200 paid repetitions, with feedback payoffs for each period, using the Holt and Laury (2002) risk elicitation method. Their findings showed changes in preferences towards the lottery with the higher expected value, indicating a shift to more risk-neutral preferences over time. A similar approach is provided by Charness et al. (2023), employing an OLS design that supports previous evidence of changes towards higher risk and expected return choices. From these insights, we state our first hypothesis, which challenges the standard assumption of fixed risk preferences.
H1: Experience leads to better self-awareness and clearer definition of one's risk preferences.

Measurements. A more complex elicitation method presents a more stable classification of subjects' risk preferences after the experience learning periods. We propose that the complexity of the risk elicitation method can impact the stability of risk preference classifications. Simple methods might lead to inconsistent classifications due to their limited capacity to capture sophisticated decision-making processes. Conversely, more complex methods, requiring participants to engage with multiple variables and decision points, are likely to provide a more comprehensive understanding of their risk preferences. This can result in more consistent and stable classifications of risk preferences, even after participants undergo experience-based learning periods. That is, we can disentangle the true impact of experience controlling for the bias related to the accuracy of

the elicitation method. By utilizing a complex elicitation method, we aim to ensure that observed shifts in risk preferences reflect genuine learning and adaptation rather than artifacts of the elicitation process itself. This hypothesis builds on previous findings, such as those by Charness et al. (2023), suggesting that richer experimental designs can better account for the dynamic nature of risk preferences.
H2: Repeating risky tasks after gaining experience may lead to changes in preferences regardless of the method used.

Measures, Measurements, and Stereotypes under risk. As previously mentioned, experience may have a different impact based on gender, as the initial familiarity with risky contexts might differ. This follows evidence on the gender effect (Charness and Gneezy, 2012) and the greater potential familiarity with risky choices for males (Harris and Jenkins, 2006). Consequently, the advantage in terms of knowledge and experience leads to marginal variations for those already experienced with risky tasks. It is interesting to observe the potential gender gap in risk preferences under repeated measurement with different tools, where both individuals are given the opportunity to gain experience in risky contexts. We want to test this by also considering the correlation between stated risk preferences (using the questionnaire question) and revealed risk preferences during the task. In fact, subjects with a low initial level of experience might distort their reported potential attitudes towards risk, making the intention-behavior gap more salient.
H3: The consistency of risk preferences and the magnitude of experience in the specific task depend on individual-specific factors.

4 Results

To compare the risk preferences of subjects across treatments, we identified a uniform categorization among treatments, classifying *Risk-averse, Risk-Neutral, and Risk-Loving* subjects. We classify as *Risk Averse* subjects in Treatment 0 and 1 as those who select Gamble 1 and 2 (lower risk expressed as standard deviation); for Treatment 2 we simplify the classification based on the number of safe choices presented in Holt and Laury (2002), grouping the different shades of risk aversion into one macrocategory. Finally, in Treatment 3, we follow the same scheme as in Treatment 2, doubling the number of safe choices for each of the risk classifications.
The classification for *Risk Neutral* subjects in Treatment 0 is based on the choice of Gambles 3 and 4, while for Treatment 1, we include only Gamble 3. The classification of these subjects in Treatment 2 follows the number of safe choices equal to 4, while in Treatment 3, it is equal to 8.
Finally, we classify *Risk Loving* subjects in Treatment 0 as those choosing Gambles 5 and 6; in Treatment 1, we consider in this category those who choose Gambles 4 and 5. Subjects falling into this category for Treatment 2 are those who make a number of safe choices equal to 3 and less than 7 for Treatment 3. Table 7 summarizes the main classification proposed to standardize the analysis across treatments.
As a first step of the analysis, we compare the aggregate Switching Rate across treatments. We compute the switching rate considering the share of people changing their risk-profile category moving from their first to final choice. Table 8 reports summarizes the percentage of subjects remaining or changing their risk-category at the end of the game. We first start to investigate whether there is any kind of relationship between changes in category and our experimental design. For each treatment, we perform a proportion test on the percentage of subjects disclosing

Type	Choices			
	T0	**T1**	**T2**	**T3**
Risk Averse	1,2	1,2	5-10	9-20
Risk Neutral	3,4	3	4	8
Risk Lover	5,6	4,5	1-3	0-7

Table 7: Choices refer to the selected rows for the OLS, and to the safe choices for the MLP cases.

changes in risk classification between treatment orders. Treatment 0 (z = -0.709, p=0.478 two-tailed test), Treatment 1 (z = -0.027, p=0.978 two-tailed test), Treatment 2 (z = -1.272, p=0.203 two-tailed test) and Treatment 3 (z = 0.281, p=0.778 two-tailed test).
Our evidence highlights changes in preferences over choice gambles, as in the recent findings in the literature; moreover, we report evidence of changes in risk classification between the initial and final choice. This evidence occurs in both our designs but with different intensity. Specifically, it can be observed that MPL elicitation methods (T2 and T3) report the lowest level of switching rate, as reported in Figure 1.

	T 0	T1	T2	T3
Change (%)	19.51	29.79	14.29	7.69
Do not change (%)	80.49	70.21	85.71	92.31
Total	41	47	42	39
Risk Attitude Initial choice				
Risk Averse (%)	21.95	51.06	83.33	89.74
Risk Neutral (%)	60.98	38.30	16.67	5.13
Risk Loving (%)	17.07	10.64	0.00	5.13
Risk Attitude Final choice				
Risk Averse (%)	26.83	46.81	73.81	89.74
Risk Neutral (%)	53.66	25.53	26.19	2.56
Risk Loving (%)	19.51	27.66	0.00	7.69

Table 8: Summary statistics for treatments

A Wilcoxon signed-ranks test on the individual changes in the chosen lottery between the initial and final choice across treatments and risk classification suggests the following: *Risk averse* subjects present mixed evidence, with choices statistically different in Treatment 1(z=-3.132, p=0.0017) and Treatment 2(z=1.778, p=0.0754)[4]. Another category revealing statistically different choices are *Risk loving* subjects, with significant changes in lottery choices($p < 0.05$) in Treatment 0, while a ($p < 0.10$) in Treatment 1. Subjects classified as *Risk Neutral* do not present statistically different choices in all our treatments.

In line with Charness et al. (2023), we observe a tendency towards final choices with higher expected values. Among our 21 subjects in Treatment 0 who change the lottery, 47.62% disclose

[4]In Treatment 2 we test the number of safe choices between the initial and final choice.

this pattern, while in Treatment 1, where 23 subjects change the lottery, the figure is 69.57%. The movements across risk classifications in these two treatments are presented in Figure 2.
Due to the heterogeneity in our sample, we can see the stickiness disclosed by Risk averse and Risk neutral subjects, rather than Risk loving, which are more widely dispersed between categories.
MPL elicitation methods suggest the same pattern: Risk neutral subjects do not show significant changes in terms of ex post risk classification, except for a couple of observations. Risk averse subjects show a tendency to reduce the number of safe choices, approaching Risk neutral behavior. Similarly, in Treatment 3, those classified as Risk loving display the same pattern described in Treatments 0 and 1. We should note that in this case the number of risk lover subject is very low. This is another point in favor of MPL because they tend to classify subjects as risk-averse or at most risk-neutral, whereas with OLS there is a more pronounced bias towards risk-loving behavior (as suggested in Charness et al., 2013).
Considering Figure 2, we can only observe a counter-intuitive result because the risk-neutral individuals in T3 are reallocated between risk-lovers and risk-averse, expressing their preferences more clearly. We note that there are only 2 ex-ante risk-neutral subjects in the specific task, so we are talking about a relatively marginal and small phenomenon that could be due to error adjustments made in the first stage by a small number of subjects.

These changes in gambles and the number of safe choices support both Hypotheses 1 and 2, stating that there are changes in preferences over risk after a rolled dice learning period, regardless of the elicitation method used. However, we can attribute part of this behavior to the bias introduced by the specific elicitation method, since this pattern is evidently reduced with MPL methods. Then, we can conclude that:
R1 and R2. With experience, a real change in the measure, and therefore in risk preferences, is observed. However, controlling for the accuracy of the measurement instrument, and thus the elicitation method used, we notice that this effect is less evident.

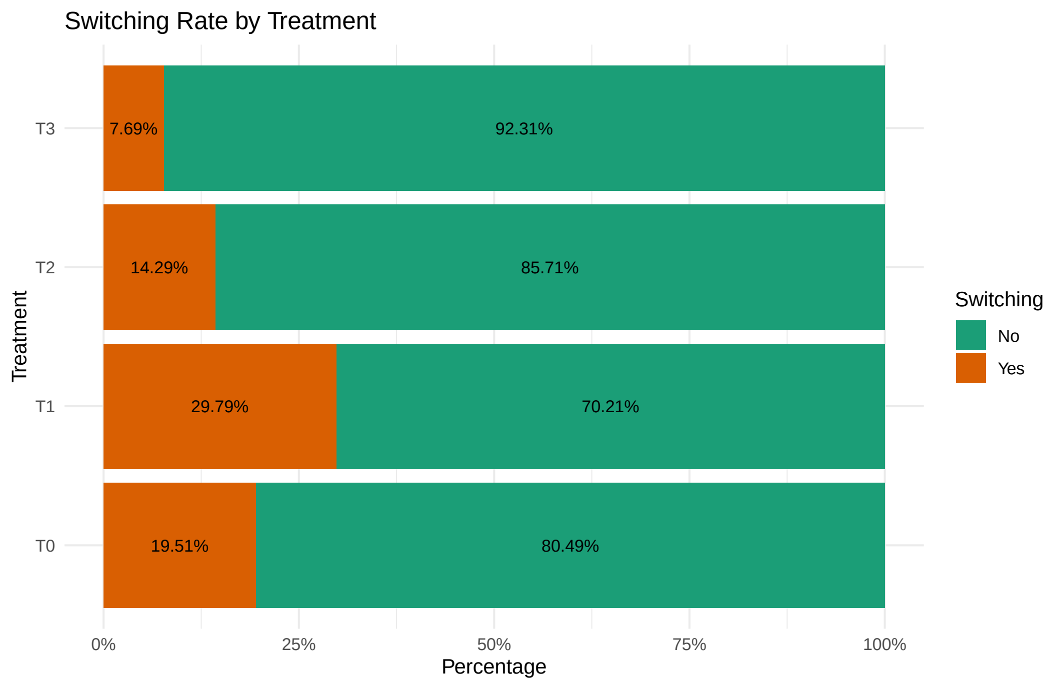

Figure 1: Switching Rate by Treatments

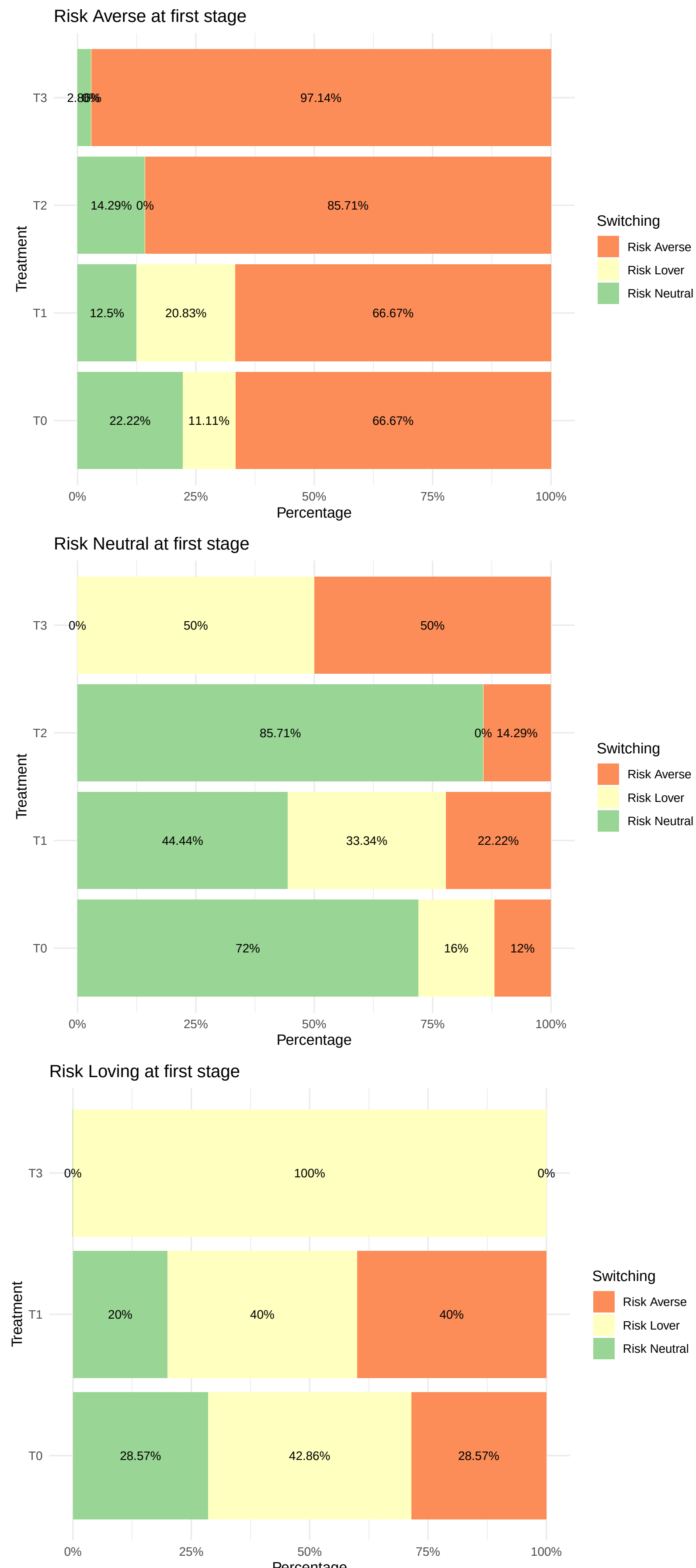

Figure 2: Subjects' categorization across treatments.

Based on the results obtained, we study the relationship between the risk preference of subjects revealed in the different tasks of the experiment and their stated preferences, declared through the survey question (hereafter: Survey Measure).
Therefore, we discuss the correlation coefficients for each treatment across initial and final choices. We compare the selected lottery for Treatments 0 and 1, while for Treatments 2 and 3, we use the number of safe choices. Considering the two designs employed in our experimental work, we note that a simpler EM shows a lower correlation between the initial and final gamble choice, being less significant in statistical terms. The preferred gamble after the non-paid learning period shows a higher correlation coefficient with the self-reported measure of risk attitude compared to the initial gamble choice. This indicates that an experience learning period aligns the experimental behavior in the final choice closer to the stated preference, reducing the intention-behavior gap. In Treatment 0, the stated preference is better represented by laboratory measures; this is likely due to the 6-row gamble, an adaptation of Dave et al. (2010), as reported by Charness et al. (2023), designed to capture risk-loving behaviour. Our analysis reveals that a more complex laboratory elicitation method correlates the initial and final choices with stronger significance and a higher coefficient than a simpler method. Another interesting difference in our data is the negative correlation between the stated preference and the revealed preference elicited with a more complex method. This gap decreases after the learning period, highlighting the effects of experience over subjects.

However, with the percentage of subjects changing risk classification after the learning period (see Table 8), and the findings in Table 9, we confirm Hypothesis 3, suggesting that a more complex elicitation method with finer subject classification is less affected by the learning period's impact on subjects' preferences over risk. This piece of evidence, so strong as interesting, shows the inconsistency between these two type of preferences, stated and revealed, unlike a repeated experience will occur.

Considering previous findings on risk preference and gender effect (see Charness and Gneezy, 2012), we present our observed differences in gender. In Table 8, we report the percentage of subjects who change risk classification after the learning period, here we split by gender and test for possible differences. Starting with Treatment 0, among the 8 subjects who change risk classification, we find more females than males, with a proportion test suggesting this difference is statistically significant ($z=-1.349$, $p<0.10$ one-tailed test). Treatment 1 shows the same pattern, but the proportion of females and males is not statistically different. In the MPL treatments, a greater percentage of change is observed among males, but the difference with females is not statistically significant. We do not find any gender differences in the initial and final number of safe choices concerning Treatments 2 and 3. However, in Treatment 1, we find a gender difference in the distribution of the initial choice. A Kolmogorov-Smirnov test one-tailed $p<0.10$ suggests an initial higher risk aversion among males; this difference disappears after the learning period. No differences are observed in Treatment 0.

In this section, we analyze the previous correlation coefficients(see Table 9) and focus on gender differences(see Table 10). Each of the four treatments shares the same pattern: decreasing in magnitude between the initial and final gamble choice, moving from male to female. However, the treatments related to MPL preserve the significance compared to the OLS. The observed lower correlation for females suggests that the learning periods' effect is stronger than for males,

Spearman correlation coefficients ρ				
		Initial Choice	Final Choice	Survey measure
	Treatment 0			
1.	Initial choice	1.00		
2.	Final choice	0.340**	1.00	
3.	Survey measure	0.455***	0.520****	1.00
	Treatment 1			
1.	Initial choice	1.00		
2.	Final choice	0.369**		
3.	Survey measure	0.145	0.255*	1.00
	Treatment 2			
1.	Initial choice	1.00		
2.	Final choice	0.693****	1.00	
3.	Survey measure	-0.082	-0.073	1.00
	Treatment 3			
1.	Initial choice	1.00		
2.	Final choice	0.727****	1.00	
3.	Survey measure	-0.439***	-0.345**	1.00

Notes: $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$, $^{****}p < .001$

Table 9: Spearman correlation coefficients

but employing a MPL design reduces this effect. Regarding the relationship between stated and revealed preferences, looking at the coefficients for gender, we observe an opposite trend between males and females, regardless of the elicitation method. Females, after experiencing the learning period, increase the intention-behavior gap, with a lower correlation coefficient between the survey measure and the final choice. Considering the correlation between initial and final choices, we observe greater consistency and thus a lower learning effect for men. This is also evident when observing the top graph in Figure 3. Looking at the bottom part of the figure, we notice a very interesting pattern: the correlation between stated risk preferences, i.e., the level of risk propensity based on survey measures, is higher for men compared to women when considering the final risky choices made in the T0 and T1 treatments (OLS methods). As mentioned previously, this type of one-shot choice closely resembles the self-reporting done via surveys.

When we use more complex methods (MPL), such as in T2 and T3, we notice different results. The correlation between stated and revealed measures is negative. This can be attributed to the overstatement of risk propensity made through survey measures and the resulting intention-behavior gap observed when making incentivized choices.

Once again, it is possible to notice that the effect is greater for men, as the learning effect for women is more evident, making the declared level of risk preferences less correlated with the revealed one at the end of the experience. This remarks the importance of incentivized experiemenst and experience in determining risk preferences, as non-incentivized surveys might lead to totally hypothetical and contrasting results with the real observed behavior. This issue is more relevant for males.

It is also interesting to note how this correlation follows a different dynamic compared to existing

studies (Attanasi et al., 2018),leaving room for further investigation.
R3: The consistency of risk preferences and the magnitude of experience in the specific task depend on individual-specific factors. Specifically, the intention-behavior gap under risk is more salient for females, since after the learning phase they better understand their risk attitude, which turns out to be different from their stated ones in the questionnaire.

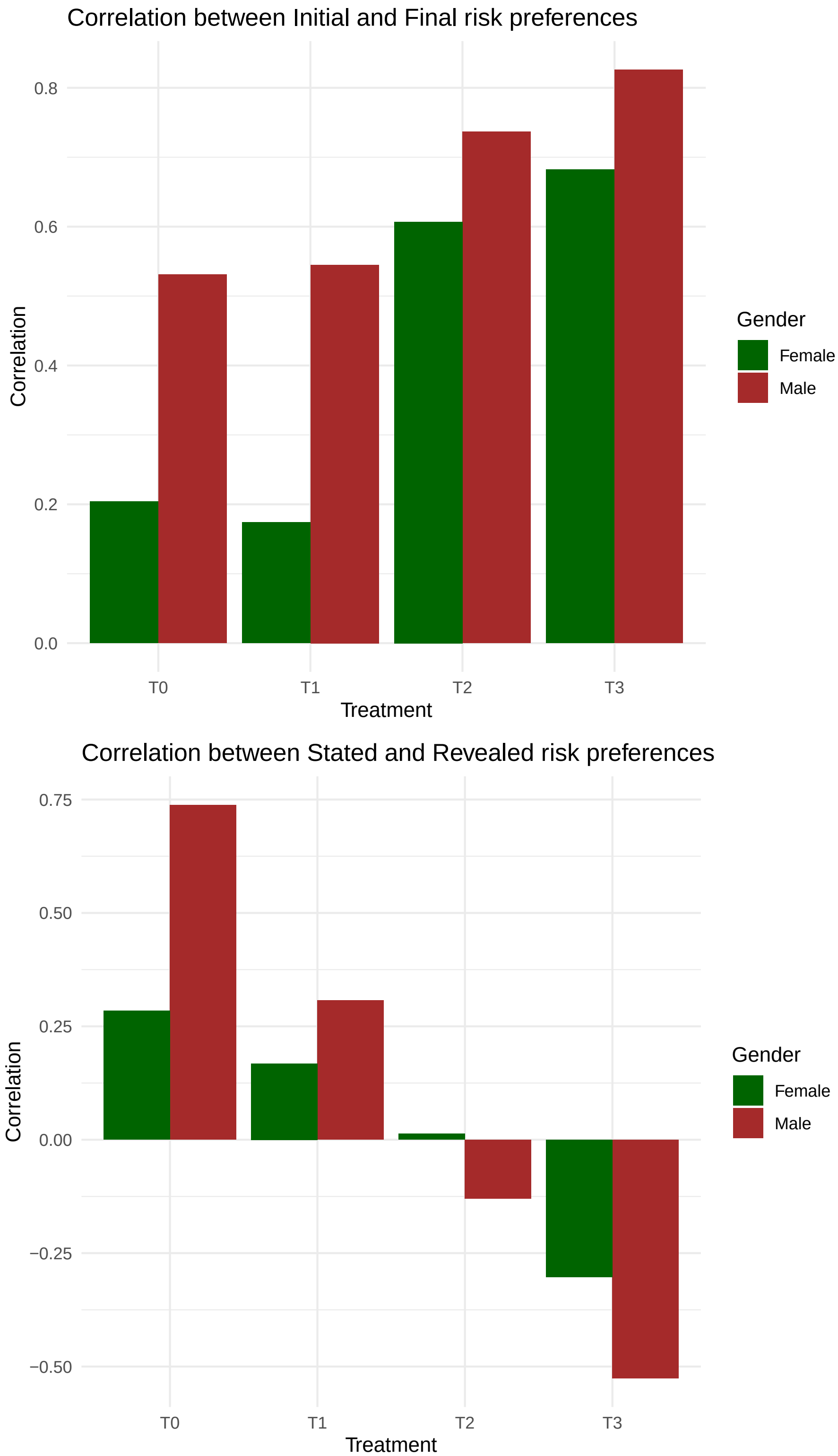

Figure 3: Correlations by gender across treatments.

Table 10: Spearman correlation coefficients ρ

	Male	correlations			**Female**	correlations	
		1.	2.	3.	1.	2.	3.
	Treatment 0						
1.	Initial choice	1.000			1.000		
2.	Final choice	0.531*	1.000		0.204	1.000	
3.	Survey measure	0.533*	0.738***	1.000	0.335	0.284	1.000
	Treatment 1						
1.	Initial choice	1.000			1.000		
2.	Final choice	0.545**	1.000		0.174	1.000	
3.	Survey measure	−0.012	0.307	1.000	0.246	0.168	1.000
	Treatment 2						
1.	Initial choice	1.000			1.000		
2.	Final choice	0.737***	1.000		0.607**	1.000	
3.	Survey measure	-0.161	-0.130	1.000	0.018	0.013	1.000
	Treatment 3						
1.	Initial choice	1.000			1.000		
2.	Final choice	0.826***	1.000		0.682**	1.000	
3.	Survey measure	−0.514*	−0.526*	1.000	-0.348	-0.303	1.000

Notes: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

5 Conclusions

Neoclassical economic theory relies on fixed people's preferences and their stability. Paramount is their important role in social science due to the predictability power they exert over several economic outcomes. In the last decades several laboratory and field experimental studies, has shown as this stability is challenged and not a reliable assumption(see Eckel et al., 2009, Bradbury et al., 2015, Ert and Haruvy, 2017 and Charness et al., 2023). This emerging debate questions the validity and the power of the tools employed in Experimental Economics literature with real-world behaviour(see Charness et al., 2020). We challenge these findings, shedding light on the changes in preference over risk after experiencing learning periods. Our experimental study, in line with previous findings, discloses actual changes in preferences over risk. These changes are observed across different elicitation methods, OLS, where we employ two treatments Eckel and Grossman (2008) and Charness et al. (2023), and MPL employing Holt and Laury (2002) and Attanasi et al. (2014). The measure of changes in preferences are relevant, but the instrument used to measure the changes in risk classification is crucial, since more complex and elaborated risk elicitation methods(MPL) result more stable in subjects' classification of risk preferences. On the one hand, results demonstrate that risk preferences are variable and adaptable, and this can be partly due to the role of experience-based learning. On the other hand, we observe how MPL methods, even if more complex, are more accurate in identifying risk preferences and then in improving measurement stability and accuracy.

Our experimental investigation further increases our existing knowledge of the behaviour-intention gap. We clearly observe a gap between stated and revealed risk preferences. Particularly, females stated preferences are uncorrelated with their final risk preferences, suggesting a higher intention-behavior gap. Males stated preferences has a stronger link with incentived risky choices, but this pattern is following opposite directions depending on the risk elicitation method. That is, while OLS and survey measures appear positively correlated, the opposite pattern is observed under MPL measurements. In this case, men believe they have a higher risk tolerance than they actually do, as evidenced by the negative correlation between the two measures.

Finally, we conclude our experimental study by recognizing the need of further investigation in estimating the subject's risk parameter employing different elicitation method designs, such as the one presented in Hey and Orme (1994). Extending our knowledge over the effect exerted by the learning period over laboratory risky choice consistency (see Hey, 2001).

However, since changes in preference over risk occurred, the choice of the elicitation method that better fit the aims of the investigation should be done in light of the stability and the classification characteristics of each design. Here, we shown how a more complex design as MPL, can deal better with changes in preference compared to the more simpler OLS.

Data availability: The author confirms that all data generated or analysed during this study are included in this published article. Furthermore, primary and secondary sources and data supporting the findings of this study were all available at the time of submission.

Ethical Statement: This manuscript complies with the Ethical Rules of this journal.

References

Attanasi, G., Corazzini, L., Georgantzís, N., and Passarelli, F. (2014). Risk aversion, overconfidence and private information as determinants of majority thresholds. *Pacific Economic Review*, 19(3):355–386.

Attanasi, G., Georgantzís, N., Rotondi, V., and Vigani, D. (2018). Lottery-and survey-based risk attitudes linked through a multichoice elicitation task. *Theory and Decision*, 84:341–372.

Bernasconi, M., Corazzini, L., and Seri, R. (2014). Reference dependent preferences, hedonic adaptation and tax evasion: Does the tax burden matter? *Journal of Economic Psychology*, 40:103–118. Special Issue on Behavioral Dynamics of Tax Evasion.

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.

Bradbury, M. A. S., Hens, T., and Zeisberger, S. (2015). Improving Investment Decisions with Simulated Experience*. *Review of Finance*, 19(3):1019–1052.

Charness, G. and Chemaya, N. (2023). Repeated experience and consistent risk preferences. *Economics Letters*, 233:111375.

Charness, G., Chemaya, N., and Trujano-Ochoa, D. (2023). Learning your own risk preferences. *Journal of Risk and Uncertainty*, 67(1):1–19.

Charness, G., Garcia, T., Offerman, T., and Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty*, 60:99–123.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior Organization*, 83(1):50–58. Gender Differences in Risk Aversion and Competition.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of economic behavior and organization*, 87:43–51.

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

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

Eckel, C. C., El-Gamal, M. A., and Wilson, R. K. (2009). Risk loving after the storm: A bayesian-network study of hurricane katrina evacuees. *Journal of Economic Behavior and Organization*, 69(2):110–124.

Eckel, C. C. and 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.

Ert, E. and Haruvy, E. (2017). Revisiting risk aversion: Can risk preferences change with experience? *Economics Letters*, 151:91–95.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Harris, C. R. and Jenkins, M. (2006). Gender differences in risk assessment: why do women take fewer risks than men? *Judgment and Decision making*, 1(1):48–63.

Harrison, G. W. and Rutström, E. E. (2008). Risk Aversion in the Laboratory. In *Risk Aversion in Experiments*, Research in Experimental Economics, pages 41–196. Emerald Group Publishing Limited.

Hey, J. D. (2001). Does repetition improve consistency? *Experimental economics*, 4:5–54.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, pages 1291–1326.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

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

Kaufmann, C., Weber, M., and Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management science*, 59(2):323–340.

McClennen, E. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Paperback re-issue. Cambridge University Press.

Miller, L., Meyer, D. E., and Lanzetta, J. T. (1969). Choice among equal expected value alternatives: Sequential effects of winning probability level on risk preferences. *Journal of Experimental Psychology*, 79(3p1):419.

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., and Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, 1(11):803–809.