

THREE ESSAYS BASED ON EXPERIMENTAL ECONOMICS: INCENTIVES AND RISK

DOCTORAL DISSERTATION

Noemí Herranz Zarzoso

Supervised by:

Dr. Gerardo Sabater Grande

Castelló de la Plana, July 2019





**UNIVERSITAT
JAUME·I**

Programa de Doctorado en Economía y Empresa

Escuela de Doctorado de la Universitat Jaume I

THREE ESSAYS BASED ON EXPERIMENTAL ECONOMICS:

INCENTIVES AND RISK

Memoria presentada por Noemí Herranz Zarzoso para optar al grado de
doctora por la Universitat Jaume I

Noemí Herranz Zarzoso

Gerardo Sabater Grande

Castelló de la Plana, Julio 2019.

Funding

I would like to express my gratitude to the Ministerio de Educación Cultura y Deporte for the financial support received through the FPU14/02966.

Thesis by compendium of publications:

Herranz-Zarzoso, N., & Sabater-Grande, G. (2018). Monetary incentives and self-chosen goals in academic performance: An experimental study. *International Review of Economics Education*, 27, 34-44. <https://doi.org/10.1016/j.iree.2018.02.002>. Impact factor 0.618 (Q2). The doctoral student has run the experiment, analysed the data and write the article.

Herranz-Zarzoso, N., Sabater-Grande, G & Jaramillo-Gutiérrez, A. Framing and Repetition Effects on Risky Choices: A behavioral approach. Revised and resubmitted to *Journal of Behavioural and Experimental Economics*. Impact factor 0.966 (Q2). The doctoral student has run the experiment, analysed the data and write the article.

Herranz-Zarzoso, N., & Sabater-Grande, G. Monetary Incentives and overconfidence in academic performance: An experimental study. Submitted to *Metacognition and Learning*. Impact factor 1.692 (Q1) The doctoral student has run the experiment, analysed the data and write the article.

“This thesis has been accepted by the co-authors of the publications listed above that have waved the right to present them as a part of another PhD thesis”

**THREE ESSAYS BASED ON EXPERIMENTAL ECONOMICS:
INCENTIVES AND RISK**

To my family.

Acknowledgements

First, I would like to express my most deep gratitude to my supervisor Gerardo Sabater Grande. He has instilled me his passion for experimental economics and has taught me to carry out satisfactory research. Thanks for believing in me since the beginning, arousing my research vocation, guiding and supporting me through the entire journey.

My everlasting thankfulness goes to the Laboratorio de Economía Experimental (LEE) members, especially Prof. Aurora García-Gallego, Prof. Nikolaos Georgantzis, Dr. Ainhoa Jaramillo-Gutiérrez and Dr. Iván Barreda-Tarrazona who have always supported and taught me in my professional and personal development. A special appreciation is for Prof. Nikolaos Georgantzís who advised me to carry out my research visit at the University of Nottingham and recommend me for doing so. I will be always grateful for this.

My gratefulness to Prof. Chris Starmer for giving me the opportunity to complete my research visit at the University of Nottingham with him. I cannot thank him enough for his kindness, feedback and insights during the stay and after it. Special thanks to the whole CeDEx research group for their warm welcome, support and invaluable suggestions and comments in a seminar I gave there, which have enriched my thesis. I must extend my sincere appreciation to Dr. José Vicente Guinot Saporta who helped me not only during the stay, but also in programming and running a new experiment with Chris Starmer. My visit to Nottingham is one of the greatest things I have done in my life.

I gratefully acknowledge the financial assistance of the Ministerio de Educación, Cultura y Deporte for a four-year FPU scholarship (FPU014/02966) which has made my doctoral studies possible.

My most heartfelt thanks go to the most important people in my life, my parents. There are not enough words to describe how thankful I am for being your daughter. Thanks for teaching me that in life you must always give your best and set the goals looking at the sky in order to be the highest ones. With their sweet personalities, confidence and incommensurable love, they have guided me to reach my dreams since I was a child, thus this thesis is theirs too. Thanks to my grandparents who take care of me from heaven and earth.

Last but not least, my endless thanks go to the person who brightened my life, my love David. I want to thank you for transmitting me your patience when I needed and supporting me in all the stages of life. Thank you for being the light of my existence.

Contents

List of Tables	iv
List of Figures.....	vii
Introduction	xi

Chapter 1. Monetary incentives and self-chosen goals in academic performance: an experimental study 15

1. Motivation.....	16
2. Literature review.....	16
3. Research Design	21
4. Experimental design	27
5. Data analysis	31
5.0. Sample self-selection: a potential willingness effect	31
5.1. Descriptive statistics	32
5.1.1. Bets	32
5.1.2. Grades	33
5.2. Statistical tests	35
5.2.1. Incentive effects	35
5.2.1.1. New students	35
5.2.1.2. Returning students	37
5.2.2. The effects of different payment mechanisms	38
5.3. Regression Analysis	39
5.3.1. Determinants of bets	39
5.3.2. Determinants of the incentivised grades	41
6. Conclusions	44
References	46

Chapter 2. Framing and repetition effects on risky choices: a behavioural approach 53

- 1. Introduction 54
- 2. Experimental Design 60
- 3. Data analysis 68
 - 3.1. Statistical tests 68
 - 3.1.1. Lottery vs. Lottery method 69
 - 3.1.2. Lottery vs. Certainty equivalent method 74
 - 3.2. Regressions 82
 - 3.2.1. Lottery vs Lottery method 82
 - 3.2.2. Lottery vs Certainty method 83
- 4. Conclusions 85
- References 87
- Appendix 92

Chapter 3. Monetary incentives and overconfidence in academic performance: an experimental study 107

- 1. Introduction 108
- 2. Experimental design 114
- 3. Measures of potential factors affecting students' overconfidence 119
- 4. Data analysis 121
 - 4.1. Sample self-selection: a potential willingness effect 121
 - 4.2. Descriptive statistics and tests 122
 - 4.2.1. Hypothetical incentive calls 123
 - 4.2.2. Real monetary incentive calls 126
 - 4.3. Regression analysis 131
 - 4.3.1. Predictions and postdictions 131
 - 4.3.2. Grades 135

4.3.3. Overconfidence 138

5. Conclusions 144

References 146

General discussion and conclusions..... 159

List of Tables

Table 1. Summary of treatments. ME: Midterm exam; FE: Final Exam	29
Table 2. Averages and standard deviations (included between parentheses corresponding to grades obtained in the midterm exam by new and returning students	31
Table 3. Models explaining bets	40
Table 4. Models explaining incentivised grades.....	42
Table 5. Pairs of lotteries offered in LL_{17}	63
Table 6. Pairs of lotteries offered in CE_{21}^S	65
Table 7. Pairs of lotteries offered in CE_{31}^R	66
Table 8. Summary of treatments	68
Table 9. LL models for the entire sample and for constant and imprecise subjects only	83
Table 10. LC models for the entire sample and for constant and imprecise subjects only	84
Table 11. Pairs of lotteries offered in LL_9	92
Table 12. Pairs of lotteries offered in LL_{17}	93
Table 13. Pairs of lotteries offered in LL_{13}	94
Table 14. Pairs of lotteries offered in LL_5	95
Table 15. Pairs of lotteries offered in LL_7	95
Table 16. Pairs of options offered in LC_{11}^S	96
Table 17. Pairs of options offered in LC_{21}^S	97
Table 18. Pairs of options offered in LC_{16}^S	98
Table 19. Pairs of options offered in LC_5^S	99

Table 20. Pairs of options offered in LC^S_8	99
Table 21. Pairs of options offered in LC^R_{16}	100
Table 22. Pairs of options offered in LC^R_{31}	101
Table 23. Pairs of options offered in LC^R_{23}	103
Table 24. Pairs of options offered in LC^R_{10}	104
Table 25. Pairs of options offered in LC^R_9	105
Table 26. Summary of treatments	117
Table 27. Average grades obtained in the exams for participants and non-participants. NBME: Non-binding midterm exam; BME: Binding midterm exam; BFE: Binding final exam.....	122
Table 28. Descriptive statistics of predictions, postdictions, grades and potential/actual overconfidence for the (non-) binding midterm exam.....	123
Table 29. Descriptive statistics and treatment comparisons of predictions, postdictions, grades, overconfidence, safe choices in the HL task, DAT-AR score, subscales studying and grades of the ABC scale, and academic record	127
Table 30. OLS regressions for predictions and postdictions elicited in the midterm exams	132
Table 31. OLS regressions for predictions and postdictions elicited in the final exam splitting the sample between monetarily incentivised and non- monetarily incentivised subjects.....	134
Table 32. OLS regressions for grades obtained in the midterm and the final exam. The latter regression is split in monetarily incentivised and non- monetarily incentivised subjects.....	135

Table 33. Probit regressions for prediction success and postdiction success. Coefficients are marginal effects 138

Table 34. OLS regressions for POC and AOC exhibited in the midterm exams 139

Table 35. OLS regressions for POC and AOC exhibited in the final exam splitting the sample in monetarily incentivised and non-monetarily incentivised subjects 140

List of Figures

Figure 1. Bets' average (standard deviations included between parentheses)	33
Figure 2. Averages and standard deviations (included between parentheses) corresponding to grades (new students) and grade improvements (returning students)	34
Figure 3. Distribution of grades obtained in the final exam by new students in T1, T2 and T3	36
Figure 4. Distribution of grade improvements obtained by returning students in T1, T2 and T3	37
Figure 5. Average rate of safe choices per pair in the LL task in T1	70
Figure 6. Average rate of safe choices per pair in the LL task in T2.....	71
Figure 7. Average rate of safe choices per pair in the LL task for constant and imprecise subjects in T1.....	72
Figure 8. Average rate of safe choices per pair in the LL task for constant and imprecise subjects in T2.....	73
Figure 9. Average percentage of certain choices per pair preferred to the safe lottery in T3.....	75
Figure 10. Average percentage of certain choices per pair preferred to the risky lottery in T3	75
Figure 11. Average percentage of certain choices per pair preferred to the safe lottery in T4	76
Figure 12. Average percentage of certain choices per pair preferred to the safe lottery in T4	77

Figure 13. Average percentage of certain choices per pair preferred to the safe lottery in T3	78
Figure 14. Average percentage of certain choices per pair preferred to the risky lottery in T3.....	79
Figure 15. Average percentage of certain choices per pair preferred to the safe lottery in T4	80
Figure 16. Average percentage of certain choices per pair preferred to the risky lottery in T4	80
Figure 17. Means of predictions, postdictions, grades, POC and AOC for the non-binding midterm and the binding midterm exams.....	124
Figure 18. Confidence using predictions and postdictions against grades in the midterm exams	125
Figure 19. Predictions, Postdictions, Grades, POC and AOC means for the final exam	128
Figure 20. Confidence using predictions and postdictions against grades in the final exam	129
Figure 21. Confidence using predictions and postdictions against grades in the final exam	131
Figure 22. Grades obtained in each exam for treatment group	137

Introduction

Economics was not considered an experimental science until recent times. Nevertheless, in the last years there has been an excellent growth in experimental economics' research. Nowadays, economists accept this experimental branch and most of them use it to check the theory. The strengthening of this experimental science occurred in 2002 when Professor Vernon Smith was awarded the Nobel Prize in Economic Sciences for his ground-breaking work in the field of experimental markets. The experimental economics' branch origin dates back to Daniel Bernoulli's (1738) study of the San Petersburg paradox. This experiment was the first that pointed out the differences between theory and subjects' behaviour.

In the past, economics and psychology were distant, but with the incursion of experimental economics, both sciences came together in some aspects. For instance, the best point of union of both sciences is the study of human behaviour. However, there are some points in which differences arise and one of them is the role of incentives.

In this thesis, incentives, risk preferences and cognitive biases as overconfidence and framing are in the spotlight.

One of the most solid principles of experimental economics is incentives. This field strongly believes that subjects will only reveal private information and preferences if they are rewarded with real money to do so, and that leads

to trustworthy results. Given that experimental economists firmly believe that monetary incentives have a crucial role motivating subjects to reveal their preferences, they are also convinced that monetary incentives will lead to more effort and higher performance in other areas like, for instance, academic performance. Specifically in this area, economists, policymakers and researchers have set their sights. This thesis goes in depth in the role of monetary incentives in academic performance in order to shed light on their potential motivational effect as a tool to improve academic performance and to remove some biases found in the literature.

Another solid principle of experimental economics is framing. It is a bias by which subjects make different choices depending on the description of a formally identical decision problem. This bias plays an important role in individual decision making under risk. Specifically, Von Neumann and Morgenstern (1944) had a fundamental contribution to individual decision making with their Expected Utility Theory in which they explained decisions under risk and uncertainty. More recently, Kahnemann and Tversky (1979) developed the Prospect Theory, pointing that each individual interprets differently probabilities. Thanks to these contributions and others, nowadays the literature accepts that individuals are not risk neutral and some tools have been provided to researchers to elicit risk attitude. Nevertheless, the risk measure used may be affected by the way in which subjects face it. For this reason, the thesis analyses what risk task measure, if any, does not suffer from this bias and consequently, is are robust to variations in the structure of the task.

This thesis is organised in three chapters that share the experimental economics methodology in order to investigate the role of monetary incentives and different cognitive biases. In terms of incentives, Chapter 1 analyses the effect of incentive-compatible self-chosen goals on academic performance by means of a randomised field experiment with the purpose of being aware of the motivational role of incentives in this new area. In terms of cognitive biases, Chapter 2 examines the framing bias caused by two versions of the most frequent procedure used to elicit risk attitudes, the choice list procedure. The aim of this chapter is to elucidate whether this bias highly reported in the literature is just due to changes in the structure of the tasks or it may be influenced by other factors. Lastly and following with cognitive biases, Chapter 3 investigates the overconfidence bias. It means that a person's subjective confidence in their performance is reliable greater than the objective result of their performance. The aim of this chapter is to analyse how monetary incentives can affect this bias in potential and actual academic achievements.

Chapter 1. Monetary incentives and self-chosen goals in academic performance: an experimental study

Abstract

This paper analyses the effect of incentive-compatible self-chosen goals on academic performance by means of a randomised field experiment. We use two alternative payment mechanisms, a piece-rate and a rank-order tournament, to motivate students depending on their absolute or relative academic performance respectively. Students enrolled in Introductory Microeconomics were classified in two types depending on whether they had a failed background in this course (returning students) or they had not (new students). Controlling for potential confounding factors such as gender, degree, professor and university entrance grade, we find that both payment mechanisms are effective increasing grades of new and returning students.

1. Motivation

Many students are especially prone to focus too much on the present. Assessing the present costs of studying is much easier than evaluating (distant) potential future benefits. Policy makers might see this myopic behaviour as an opportunity of improvement, and might want to try to offset present costs implementing closer benefits. One approach to address the present bias is simply requiring students to think about their academic goals and to formulate them. Another approach to correct the aforementioned bias is offering immediate incentives that trim immediate costs. Adopting both approaches, we conduct a randomised field experiment where over 170 undergraduate students are asked to report their individual goals on academic performance and monetary incentives are delivered to participants who reach their self-chosen goal.

2. Literature review

Our paper is related to two separate lines of research. The first one is the literature on goal setting. Goal setting is a cognitive theory based on the premise that the source of motivation is the desire and intention to reach a goal, i.e. the aim of a task that a person consciously desires to achieve or obtain (Locke and Latham, 2002, 2006). In achievement environments, such as higher education, task goal setting is a function of many variables, including skills and outcome expectations. Following Zimmerman (2011),

outcome expectation can be defined as a belief about the success of a given task, differentiating it from the highly correlated concept of self-efficacy expectation, which is the belief about the personal capability to execute the behaviour needed to produce the desired outcome. Both of them are key for students' motivation: if a student does not consider himself capable or has low outcome expectations, his/her motivation decreases and he/she does not make the effort needed to succeed¹. Although Bandura (2006) developed a guide for constructing scales to measure perceived academic self-efficacy, it is not possible to get an incentive-compatible elicitation of self-efficacy scales. However, outcome expectations can be elicited using betting on outcomes as an incentive-compatible method. Specifically, when students bet on their outcome in a particular task, they take responsibility and ownership for their own goal. As Elliot and Fryer (2008) pointed out, such self-chosen goal setting is empowering and proactive, creating commitment and acceptance. In the same vein, Royer *et al.* (2015) and Samek (2016) concluded that giving the option to subjects to choose their own goal “acts as an internal commitment device meant to overcome problems of self-control” (Samek 2016, 2).

From the pioneering work of Latham and Locke (1979) to the present, there has been an extensive body of empirical literature testing goal-setting theory in controlled environments²³. However, when it comes to self-chosen

1. See Pajares (2008) for a literature review on self-efficacy and regulation learning.

2. These authors were the first to report evidence that goals lead to a better performance as compared to not setting goals.

3. See Dykstra (2015) for a review on empirical evidence supporting goal setting as a tool to increase individuals' performance.

goals, the number of empirical contributions is limited and not all of them use incentive compatibility. Without financial incentives, Falk and Knell (2004) presented a social comparison model where people chose their own reference standards in order to accomplish goals of self-enhancement and self-improvement. The model's prediction about people tending to compare themselves to similar others was confirmed through a questionnaire where students only received a small show-up fee. Using the same methodology, Sackett *et al.* (2014) presented a questionnaire where marathoners were either asked or not asked to provide a time goal prior to their race. They found that the mere fact of asking runners (without potential financial reward) about their goals prior to the race improved performance among experienced but not novice marathoners.

Using a large sample of college students, Clark *et al.* (2016) analysed the effect of self-chosen goals based on course performance and a specific task (completing online practice exams). They found that course goals had no significant impact on the performance of college students but, in contrast, task-based goals had a large positive effect on the level of task completion. However, students' extrinsic motivation in these two field experiments could be critically undermined because students were not financially rewarded when they met their self-chosen goals. The importance of financial incentives was analysed in Goerg and Kube (2012). By means of a natural field experiment where workers were hired to re-organise a library, these authors presented the workers with an incentive's contract that combined self-chosen goals and

monetary rewards in such a way that all of them preferred to set themselves a non-trivial goal. They found that self-imposed goals could work even in the absence of corresponding monetary incentives.

Empirical evidence on self-chosen goals where subjects' performance is based on financial incentives is even scarcer. Using a field experiment in an Indian data-entry firm, Kaur *et al.* (2010) tested whether workers demand self-disciplining devices. They found that a fraction of workers voluntarily agreed to incur in a monetary loss when falling short of a self-chosen production goal. Moreover, Dalton *et al.* (2015) proposed a simple model of self-chosen goals and tested its predictions in the laboratory, finding that only men confirmed their model's predictions: they exerted greater effort under the self-chosen goal contract system than under a piece-rate contract.

More related to our experiment and involving first-year university students, van Lent and Souverijn (2016) analysed the effect of setting a goal and increasing its ambitiousness using mentor-student meetings. They found out that treatment group students performed better as compared to students in the control group. Nevertheless, students who were challenged to set a higher goal performed significantly worse than comparable students in the goal treatment. Contrary to van Lent and Souverijn (2016), we ask students to formulate their own goal and bet on it without any external influence, rewarding them financially. Theoretically, in our experiment, students' self-chosen goals must generate more effective incentives than those given by

a third party because students set their goals based on their capabilities and knowledge, creating their own individual motivation.

Second, our paper is related to the literature based on financial incentives on academic performance. Experimental economists are convinced that higher incentives will lead to more effort and higher performance. However, psychologists claim that incentives improve performance in “algorithmic” or repetitive tasks, but they are less effective, or even counterproductive, in “heuristic” tasks requiring creativity, concentration, or intuition. Because learning has generally been classified as heuristic, extrinsic grade incentives may not be effective motivators. The argument behind this claim is that the use of incentives could crowd out intrinsic motivations that are important to produce the desired behaviour⁴. Leaving aside ethical issues about the convenience of using financial incentives to improve students’ grades, incentives have become object of interest of economists, policymakers, and researchers in the last years. In the literature, mixed evidence has been obtained using randomised field experiments and natural experiments (using databases from state programs) to analyse the effects of financial incentives on students’ academic performance⁵.

4. Gneezy *et al.* (2011) state that a potential conflict arises between the extrinsic and the intrinsic effects of the incentives, especially in areas like education, contributions to public goods and forming habits, in the short run and in the long run.

5. See Lavecchia *et al.* (2016) for a survey of effectiveness of financial incentives in education.

3. Research Design

All the evidence analysed in the previous section is far from being conclusive because it has been generated in experiments or quasi-experiments with a wide array of incentive specifications, incentivised performances, target students and timing of performances and payments. The following is an elaboration of how such characteristics are related to the effectiveness of financial incentives on academic performance, comparing our study to the literature.

To begin with, we deal with the specification of the incentives provided. This specification includes two general categories: incentives for inputs and incentives for outputs. The former means anything (reading books, doing homework, attending school, etc.) that can contribute to learning under the student's control. The latter refers to student achievement, generally measured through test scores or class grades. Although Fryer (2011) concluded that output experiments demonstrated less-promising results than the input ones, we focus on output incentives given that traditional price theory predicts that they are socially optimal⁶. Output incentives include financial aid intended for payment of education expenses like tuition, fees and books, and freely

6. Using different award schemes for primary and middle school students, Fryer (2011) found that paying for performance on standardised tests had little or no effect on the outcomes for which students received financial incentives.

available monetary rewards.⁷ We prefer to use money as incentive because, as Croson (2005) pointed out, everyone values it and it is non-satiable (more is always better). Additionally, it is worth to note that our field experiment has been specifically designed for research purposes and it is not part of any state program dedicated to improve academic results. This fact allows us to manipulate independent variables, choosing timing incentives and payment mechanisms (piece-rate system and rank-order tournament) under controlled conditions⁸.

Second, we focus on the incentivised performance. In most cases, achievements required are referred to accomplishing a minimum composite score over an academic course, involving different subject matters and skills.⁹ Given our interest in self-chosen goals, we offer incentives in a specific subject and over a limited time horizon, trying to facilitate students' thinking over their goals. With this objective in mind, we choose to incentivise the Introductory Microeconomics final exam like in Leuven *et al.* (2011). Similar to us, Levitt *et al.* (2011) incentivised a specific task (a standardised test), focusing on short-term effort in achievement.

Third, we address the issue of target students, differentiating (a) primary/

7. Many colleges and universities offer financial incentives in the form of merit scholarships. However, these incentives are of a different nature to freely disposable money. See, for example, papers by Henry *et al.* (2004), Cornwell *et al.* (2005), Angrist and Lavy (2009), Angrist *et al.* (2009), Scott-Clayton (2011), Sjoquist and Winters (2012) and Castleman (2014).

8. The vast majority of the papers use only one payment mechanism, generally the piece-rate system.

9. See, for example, papers by Fryer (2011), Bettinger (2012), Angrist *et al.* (2014), Barrow *et al.* (2014) and Castleman (2014).

middle education and higher education students, (b) higher and low ability students and (c) voluntary and non-voluntary students¹⁰. Regarding (a), we are interested in higher education students given that monetary incentives' provision to under-age subjects can be a problematic issue. This is so given that when participants are minors, experimentalists have to consider the ways children of different ages view the value of a payment, and to ensure that the amount and method are age-appropriate, and it does not present undue influence. With respect to (b), students are classified based on their scores in different tasks, which can be more or less related to the incentivised performance. Using scores in two programs of mathematics in Dutch secondary education, Leuven *et al.* (2010) sorted students in high and low ability types. These authors found that high-ability students had higher pass rates and collected significantly more credit points whenever they were assigned to (larger) reward groups. In contrast, low-ability students appeared to achieve less when they were assigned to the large reward group. These negative effects for less-able students were consistent with the work by Camerer and Hogarth (1999), who found that the performance threshold could result in a binding participation constraint at the bottom of the ability distribution, resulting in zero incentive effects for low-ability students. In the same line and following a tournament rule to the 30 best performing students, Paola *et al.* (2012) obtained that financial rewards increased high ability (above the median high school grade) students' performance while the effect was null for low ability participants. Looking for the closest relationship

10. See Laveccia *et al.* (2016) for a literature classification according to education level.

between students' academic background and their incentivised task, we classify them in two types: returning students (if they had a failed background in Introductory Microeconomics) and new students (if they had not). Contrary to the aforementioned literature, where the effect of incentives is low or null for low ability students, we obtain that incentives are effective, increasing returning students' academic performance. Regarding (c), we randomly assign students to a treatment group (where incentives are offered) or to a control group (where no incentives are offered once they have explicitly declared their will to participate). Our design, requiring voluntary participation, ensures that subjects are really interested in being included in the experiment. In the terminology we use below, only subjects interested in the offered incentive will be considered participants. In the vast majority of the literature, experiments included as participants subjects who did not explicitly declare any interest in the offered incentive. Exceptions are Leuven *et al.* (2011), where students had to select themselves into different tournaments, Jackson (2010), Cha and Patel (2010), Scott-Clayton (2011), Pattel and Rud (2012), where students had to enrol in the corresponding state program, and Paola *et al.* (2012), where students were asked to fill a participation form¹¹.

Lastly, we deal with another significant issue: time intervals. Specifically, we focus on the waiting time between the call and the starting of the task, and

11. In our paper, incentive effects are disentangled from sorting effects by means of an experimental design that allows us to obtain individual data on academic performance with and without incentives in the same Introductory Microeconomic course. Contrary to Leuven *et al.* (2011), we find that the difference in performance between students can be attributed entirely to incentive effects.

the waiting time between the accomplishment of the task and the reception of incentives (rewards). On one hand, most of the studies that reward incentives based on test performance announced the incentives well in advance of the test. On the other, studies that announced incentives immediately before the test distributed the payoffs with an appreciable delay. The evidence on such delayed rewards is mixed. In this sense, O'Neil *et al.* (1995, 1997) found that students' effort may be increased by financial rewards offered at the time of the test. In the same vein, Levitt *et al.* (2012) found that all motivating power of the incentives vanished when rewards were handed out with a delay. Therefore, shortening as much as possible both time intervals is a recommendable strategy in order to obtain a significant impact on student performance.¹² In our case, both waiting times are reduced to a minimum: incentives are offered after the midterm exam, allowing a maximum one month time span between the announcement of the experiment and the day of the final exam (similar to Leuven *et al.*, 2011).¹³ After the publication of the definitive grades (two weeks after the exam), participants are immediately paid in cash.

Summing-up, our experimental design aims to maximise the efficacy of monetary incentives on academic performance in a specific task, offering monetary incentives (according to a piece-rate system or a competitive ranking) on self-chosen goals to volunteer higher education students. Based on this design strategy, we propose the following hypotheses:

12. Braun *et al.* (2011) is an example of good results in which the incentive was announced immediately before the test and the reward was distributed immediately after the test.

13. Students were allowed to bet until the day before the final exam.

Hypothesis 1 (H1): Monetary incentives based on self-chosen academic goals should increase the academic performance obtained by both new and returning students.

Hypothesis 2 (H2): A piece-rate payment mechanism should be more effective upgrading average academic performance of all participants than a rank-order tournament system. Contrary to piece-rate payment schemes, under rank-order tournaments students' payoffs are discontinuous in the level of exerted effort. This is so given that a marginal unit of effort increases the expected student payoff by increasing the probability to win, but it does not necessarily generate a higher payoff. This feature can produce complicated behavioural effects affecting its effectiveness.

Our results confirm H1 and reject H2 using a clean experimental design where confounding factors such as gender, degree, professor and university entrance grades are controlled.

Onwards, the structure of the paper is organised as follows: first, we present the design of the implemented experiment; after that, we analyse the empirical evidence collected; and lastly, we present our conclusions.

4. Experimental design

We conducted a randomised field experiment aiming to improve academic performance in an introductory course of Microeconomics through monetary incentives offered on the basis of self-chosen goals.

New and returning students enrolled in the 2017/2018 course of Introductory Microeconomics at the University Jaume I were offered the possibility of taking part in a monetary incentive program asking them through an invitation call about their willingness to participate. We opened our call once students had been midterm examined.¹⁴ In the call they were informed that students responding affirmatively would be randomly assigned¹⁵ to one of three groups: a control group, Treatment 1 (T1), or two alternative treatment groups: Treatment 2 (T2) and Treatment 3 (T3). In T1 participants would not be monetarily incentivised. Alternatively, participants assigned to T2 would be paid according to their absolute (piece-rate) academic performance. Lastly, in T3 a rank-order tournament would be used as payment mechanism to reward participants.¹⁶ Additionally, we notified that participants would receive information about their corresponding group before they were invited to choose a goal for their final exam grade. This design discards possible

14. Subjects had not been informed about any incentive program before the midterm.

15. We randomise the treatments within each type (new and returning) of student.

16. Rank-order tournaments as analysed in Lazear and Rosen (1981) are characterised by the evaluation of individual performance relative to the performance of competitors. Regarding this issue van Dijk *et al.* (2001) find that tournaments lead to a higher effort on average but more variable compared to the other payment schemes.

willingness effects given that, both the treatment groups and the control group are integrated only by students declaring their willingness to participate in the program. Moreover, students not responding our call were included in Treatment 0 (T0) in order to compare non-participants' midterm grades with the corresponding ones obtained by participants. In doing so, we can check for the actual existence of a (potential) willingness effect in our sample.

Like in Clark *et al.* (2016), we allowed participants to bet¹⁷ according to their personal goal (as opposed to goals set by a professor).¹⁸ In this way, the goal was tailored to each student's degree of self-control and ambition. In order to do so, we used an experimental design strategy similar to Jackson (2010), opening a call named "Bet for your grade and win", right after students had been informed about their grades in a midterm exam.¹⁹ This within-subject strategy aims to collect individual grades with and without incentives for the same subject. Among 496 students enrolled in Introductory Microeconomics (406 new students and 90 returning students), 177 of them (111 new students and 66 returning students) attended the call to participate in this experiment. These rates of acceptance contrast with De Paola *et al.* (2012) where about 90% of students assigned to treated groups decided to participate in the experiment.

17. Note that we use the meaning of "bet" as a guess or opinion, given that our participants do not risk their own money.

18. Unlike van Lent and Souverijn (2015), where a mentor-student meeting was used to induce students to set a course-specific grade goal, we choose to implement a website to gather students' bets in order to avoid any kind of bias in their outcome expectations.

19. In order to participate in the experiment, subjects were asked to register on a betting system based on PHP+MySQL through our lab's (LEE) website. Once registered they were randomly assigned to the control or the treatment groups. Then, all students could bet on the highest grade they thought they could get.

Table 1 lists the summary of the treatments discussed above.

Table 1

Summary of treatments. ME: Midterm exam; FE: Final Exam

<i>Treatments</i>	<i>Number of subjects</i>		<i>Treatment variables</i>		
	<i>New</i>	<i>Returning</i>	<i>Willingness to participate</i>	<i>Incentive ME / FE</i>	<i>Payment Mechanism</i>
<i>T0</i>	295	24	No	No / No	-
<i>T1</i>	41	19	Yes	No / No	-
<i>T2</i>	31	22	Yes	No / Yes	Piece-rate
<i>T3</i>	39	25	Yes	No / Yes	Tournament

In the experiment, the purpose of bettors was to maximise their monetary reward (R). Subjects were informed that their reward (R) would depend on their bet (B), the grade (G) they obtained in the Introductory Microeconomics final exam, and, only for returning students, the average grade (AG) in the same subject-matter final exams in previous semesters, according to the following function:

$$R = \left(B - \left[\frac{AG}{2} \right] \right)^2 \quad \forall G \geq B$$

Students were rewarded with R euros only if G was higher than or equal to B .²⁰

²⁰. For new students $R=B^2$ given that $AG=0$.

Therefore, in T2 all students whose grade was higher than or equal to their bet (successful bettors) earned R euros.²¹ Alternatively, in T3 we implemented two rank-order identical tournaments, one for the new students and another one for returning students, offering the aforementioned rewards only for the top three students in each tournament²². Thus, new and returning students compete separately in two independent tournaments for prizes. In the experimental instructions, participants²³ were informed that the two rankings would be published soon after the notification of the final exam actual grades, and prizes would be delivered to participants immediately after that.²⁴

Our experimental design allows us: (a) to make between-subjects comparisons between grades obtained in the midterm exam by non-participants (T0) and participants (T1+T2+T3) in order to test for willingness effects, (b) to make between-subjects comparisons between grades obtained in the final exam by T1 students and T2/T3 students and (c) to analyse the factors driving the incentivised grades and bets, controlling for potential confounding factors, such as gender, degree, professor and University Entrance Grade (UEG).

21. Note that, if $G \geq B$, the reward was the same for all students with identical bets, independently of the obtained grade. Additionally, students were only paid if their final grade was at least 5 out of 10.

22. According to Vandegrift *et al.* (2007), a tournament in which second and third-place performers also receive a payment should induce lower performance than a winner-take-all tournament.

23. New and returning students were informed about the number of participants included in their group before they were allowed to bet.

24. The instructions are available upon request.

5. Data analysis

5.0. Sample self-selection: a potential willingness effect

Because of our design requires students' willingness to participate in the program in both the control group and the treatments group, potential self-selection problems are ruled out. In spite of this, it can be interesting to analyse the real existence of this potential effect comparing grades obtained in the midterm exam between students who are willing and who are not willing to participate in the program.

Table 2

Averages and standard deviations (included between parentheses) corresponding to grades obtained in the midterm exam by new and returning students.

Students	New		Returning	
	Yes	No	Yes	No
Participants	4.35 (2.72)	3.59 (2.58)	4.28 (2.45)	3.78 (2.28)

Table 2 shows that the average grades obtained in the midterm exam by new/returning students declaring their willingness to participate in the program are higher than non-participants. However, these differences are

statistically significant in median only for new students.²⁵ In consequence, requiring voluntary participation in the control group avoids an actual self-selection problem in our sample of new students.

Additionally, as double-check tool for rule out self-selection problems, we compare grades obtained by students in the midterm exam between our treatment groups (T2/T3) and the control group (T1) finding no significant differences in any case.²⁶

5.1. Descriptive statistics

This section presents descriptive statistics of students' bets and grades corresponding to the final exam in the Introductory Microeconomics course.

5.1.1. Bets

Figure 1 shows new and returning students' bets averages. For both, new and returning students, we observe that the bets' average is slightly higher in T3 compared to T2.²⁷ Moreover, the bets' averages obtained by new students are higher than the ones obtained by returning students.

25. Mann-Whitney test p-values for new (returning) students: 0.013 (0.291).

26. Using a t- (Mann-Whitney) test we find no significant differences between T1 and T2, and T1 and T3, for both new (returning) students, p-values: 0.771 and 0.760 (0.236 and 0.398) respectively.

27. Using a Mann-Whitney test we find no significant differences between T2 and T3 bets for both new (p-value: 0.935) and returning students (p-value: 0.270)

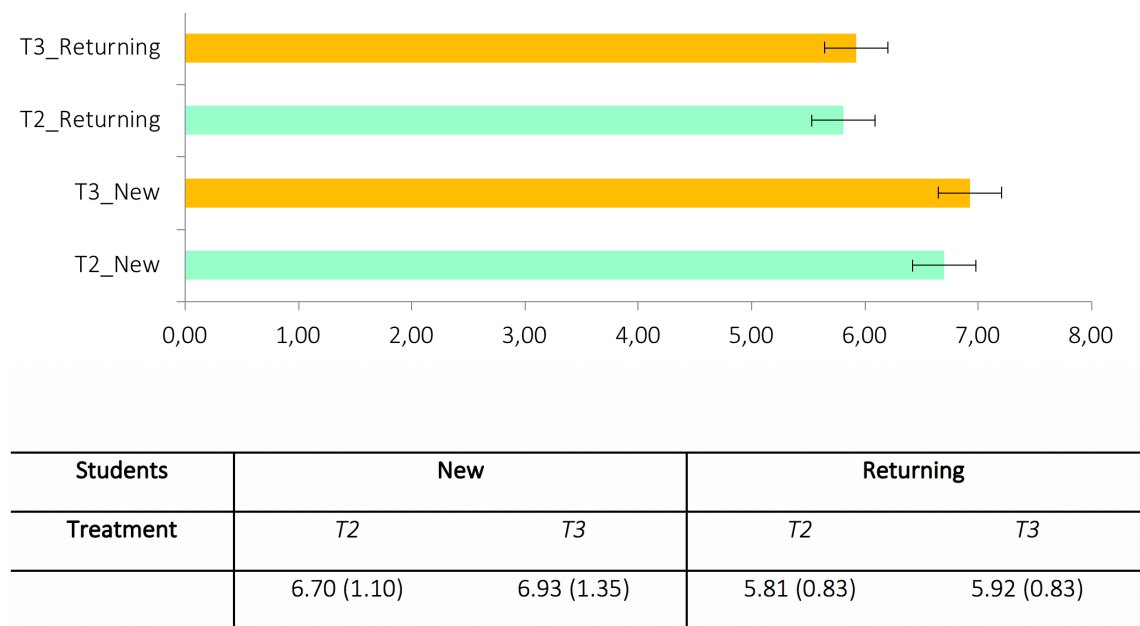


Fig 1. Bets' average (standard deviations included between parentheses)

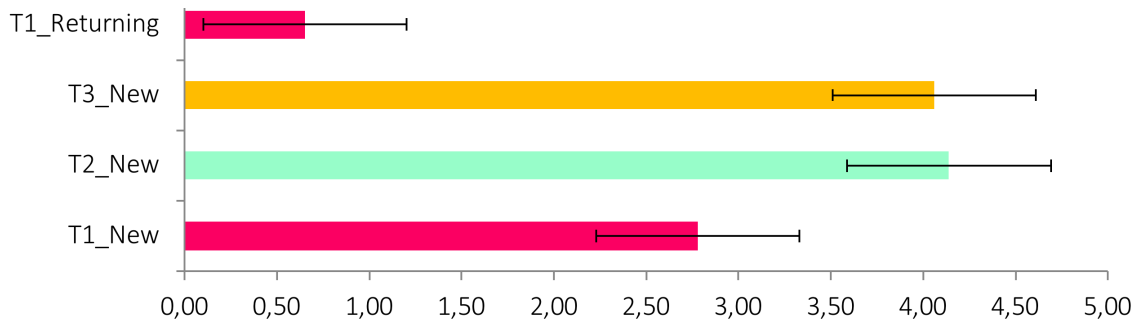
In view of the fact that only the top three students were monetary rewarded in T3, bettors in this treatment could be compelled to overestimate their goals, resulting in higher bets.

5.1.2. Grades

For returning students, we define “grade improvement” as the students’ final exam grade in the semester in which the incentive program was implemented minus the average of final exam grades in previous semesters.

Figure 2 displays descriptive statistics corresponding to new students’ grades and returning students’ grade improvements after our call. We observe

that the average of grades obtained by T2 (T3) new students is 1.36 (1.28) points higher than the average of grades obtained by T1 students. A similar pattern is observed for returning students given that the difference between T2 (T3) and T1 average of incentivised improvement grades is 1.23 (1.43). Graphically, we can see these aforementioned differences in grades and grade improvements between the control group and the treatment group students for new and returning students respectively are noticeable.



Students	New			Returning		
	T1	T2	T3	T1	T2	T3
Treatment	2.78 (1.65)	4.14 (2.96)	4.06 (2.63)	+0.65 (2.30)	+1.89 (2.20)	+2.09 (1.62)

Fig 2. Averages and standard deviations (included between parentheses) corresponding to grades (new students) and grade improvements (returning students).

Additionally, for new students in the two incentivised treatment groups, we observe that the standard deviation of incentivised grades is higher than the one corresponding to T1. Nevertheless, for returning students we observe the opposite effect.

Comparing the treatment groups, T2 and T3, we observe no noticeable differences between grades under the two incentive systems (piece-rate and rank-order tournament) for both new and returning students.

5.2. Statistical tests

5.2.1. Incentive effects.

In this subsection, we analyse the effect of monetary incentives on new and returning students' final exam grades. Since all datasets are normally distributed for a 95% level of confidence, parametric tests are used in order to compare incentivised grades: (a) a t-test comparing means from the two populations, (b) a Levene test comparing variances and (c) a Kolmogorov-Smirnov test comparing distributions.

5.2.1.1. New students

Distributions of grades obtained in the final exam by students in T1, T2 and T3 are presented in Figure 3. For grades higher than five, the percentage of treatment group students is higher than the controls' percentage. Particularly, only incentivised students reach the highest grades. Additionally, for a 90% level of confidence, grades obtained by new students in T2 can be considered not normally distributed.²⁸ For this reason, we implement additional non-parametric tests in order to confirm our results.

28. Kolmogorov-Smirnov test p-values for T1, T2 and T3 are 0.084, 0.119 and 0.115.

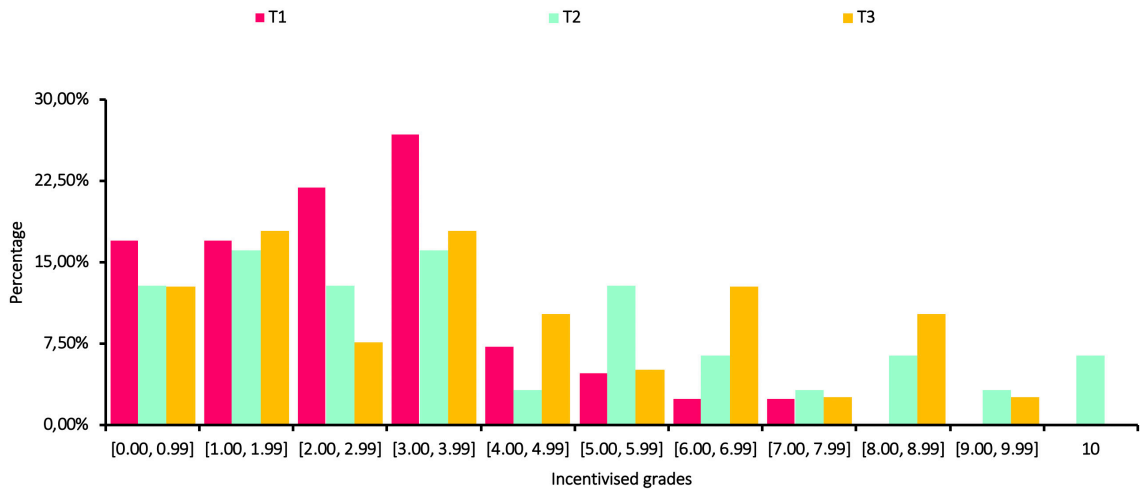


Fig 3. Distribution of grades obtained in the final exam by new students in T1, T2 and T3.

For both treatments, T2 and T3, from (a) we find that the median of grades obtained by incentivised groups is significantly higher than the median of grades obtained by students in our control group.²⁹ In addition, we obtain from b) that the variance of T2/T3 grades is significantly higher than the variance of T1 grades³⁰, while from (c) we find that there are no significant differences between the distributions of grades obtained by T2/T3 students and the corresponding distribution for T1 students.³¹

Result 1: *For both payment schemes, piece-rate and rank-order tournament, monetary incentives based on self-chosen goals are effective to increase the grades obtained by voluntary new students.*

29. T-test p-values: 0.027 for T2 and 0.012 for T3. Additionally, Mann-Whitney test p-values are 0.096 for T2 and 0.042 for T3.

30. Levene test p-values: 0.000 for T2 and 0.002 for T3.

31. Kolmogorov-Smirnov test p-values: 0.104 for T2 and 0.079 for T3.

5.2.1.2. Returning students.

Figure 4 presents distributions of improvement grades obtained in the final exam by returning students in T1, T2 and T3. In this figure, we can observe that, with the exception of improvement grades between 4 and 4.99, for any positive grade improvement the percentage of controls is lower than the percentage of incentivised students.

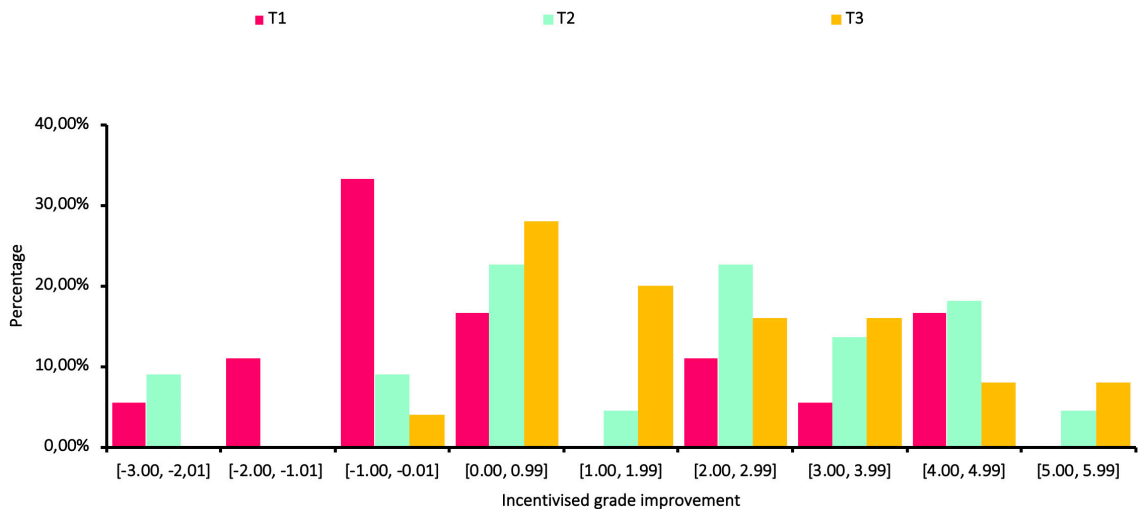


Fig 4. Distribution of grade improvements obtained by returning students in T1, T2 and T3.

In both treatments, from (a) we find that mean of grade improvements in T2/T3 is significantly higher than the mean of grade improvements obtained by non-incentivised students.³² Moreover, from (b) we find significant differences between the variances of grade improvements only comparing T1 and T3³³, while from (c) significant

32. T-test p-values: 0.080 for T2 and 0.027 for T3.

33. Levene test p-values: 0.708 for T2 and 0.088 for T3.

differences are obtained between distributions of incentivised and control students.³⁴

Result 2: *Both relative and absolute academic monetary incentives based on self-chosen goals are effective to increase the grade improvements of returning voluntary students.*

Results 1 and 2 confirm our H1. However, result 2 contrasts with Grove and Wasserman (2006) that using a natural experiment concluded that grade incentives to practice economics throughout the semester boosted the average freshman exam performance, but not that of academically above- or below- average students, or of any other category of students.

5.2.2. The effects of different payment mechanisms

In this section, we compare the treatments groups (T2 and T3) to analyse the effect of the two payment mechanisms implemented for both new and returning students finding that there are no significant differences between medians, variances and distributions of (improvement) grades obtained by new (returning) students in T2 and the corresponding ones in T3.³⁵

In accordance with the findings obtained by van Dijk *et al.* (2001), that workers with relatively low ability do not seem to realise that they have little

34. Kolmogorov-Smirnov test p-values: 0.066 for T2 and 0.011 for T3.

35. For new (returning) students, t, Levene and Kolmogorov-Smirnov tests p-values are 0.903 (0.716), 0.458 (0.092) and 0.982 (0.357) respectively.

chance of winning a rank-order tournament, we obtain that subjects do not bet differently owing to the payment system. Namely, differences in the median/distribution of bets are not statistically significant between treatments³⁶.

Result 3: *Incentives based on relative academic performance are equally effective than those based on absolute performance to increase the average of students' grades.*

5.3. Regression Analysis

5.3.1. Determinants of bets

In this subsection, we estimate different models to shed light on the determinants of the bets.

The models include as covariates: (1) the fact of being assigned randomly to T1 (where students are paid according to a piece rate) or to T2 (where a tournament is used to reward students), (2) the non-incentivised (midterm) grades collected before the starting of the incentives' program, (3) the fact of being a returning student, (4) the degree they are enrolled in ³⁷, (5) professor³⁸ and (6) the University Entrance Grade (UEG).

36. Mann-Whitney test p-value is 0.761 and Kolmogorov-Smirnov test p-value is 0.249.

37. The subject Introductory Microeconomics is shared by students of 4 different degrees: Management, Economics, double degree of Economics and Law and Accounting.

38. Each professor teaches one of the 6 class groups from A to F.

Table 3.

Models explaining bets. Standard errors in parentheses. *** $p < 0.01$,
 ** $p < 0.05$, * $p < 0.1$

Bet	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
T2			0.807** (0.331)	0.804** (0.324)	0.628* (0.326)	0.635* (0.325)	0.621* (0.322)
T3			0.907*** (0.321)	0.914*** (0.314)	0.883*** (0.307)	0.926*** (0.320)	0.925*** (0.317)
Non-incentivised grade	0.0906** (0.0424)	0.0908** (0.0408)	0.0725* (0.0402)	0.0842** (0.0397)	0.0638 (0.0394)	0.0450 (0.0437)	0.0533 (0.0436)
Returning		-0.722*** (0.254)	-0.984*** (0.262)	-0.811*** (0.268)	-0.647** (0.268)	-0.618** (0.301)	-0.824** (0.325)
Gender				0.524** (0.238)	0.582** (0.245)	0.560** (0.262)	0.491* (0.264)
Management					-0.173 (0.279)	-0.237 (0.282)	-0.264 (0.280)
Economics					-0.214 (0.304)	-0.503 (0.330)	-0.451 (0.329)
Economics + Law					1.219** (0.492)	1.290** (0.534)	1.831*** (0.628)
Class group A						-0.0331 (0.329)	0.207 (0.359)
Class group B						-0.0858 (0.405)	-0.0377 (0.402)
Class group D						0.233 (0.487)	0.268 (0.483)
Class group E						0.122 (0.386)	0.151 (0.383)
Class group F						1.090** (0.478)	1.000** (0.477)
UEG							-0.147 (0.0923)
Constant	5.882*** (0.248)	6.104*** (0.251)	5.602*** (0.298)	5.185*** (0.348)	5.331*** (0.415)	5.398*** (0.419)	6.575*** (0.847)
Observations	94	94	94	94	94	94	94
R-squared	0.047	0.125	0.201	0.243	0.314	0.369	0.389

The last model presented (Model 7) controls for all the potential confounding factors. The results obtained in this analysis are quite conclusive. The midterm exam grades do not influence subjects' bets. Otherwise, other factors such as treatment group, the fact of being returning, gender, degree and class group affect significantly the self-chosen goals of students. Particularly,

those who have failed the subject previously choose less-demanding goals than students facing the subject the first time do. Furthermore, given that in the tournament mechanism (T3) only the top three students are rewarded, their bets are higher than the ones of those betting under a piece rate mechanism (T2). In the explanation of bets, gender plays a role, pointing that women are more conservative in terms of bets than men are. In addition, students enrolled in the double degree bet significantly higher than those enrolled in accounting do. A possible explanation may rest on the fact that they are, theoretically, cleverer given that the cut-off mark for entering in this degree is higher. Furthermore, one class group is statistically significant respect to class group C. We use class group C as a reference category because the professor is one of the experimentalists. The results help us to avoid potential critics about the problem of having a professor in the subject who has design the experiment given that only the bets fixed by students belonging to the class group F are significantly higher. The remaining class groups bet in a similar way. Finally, the UEG does not affect students' self-chosen goals.

5.3.2. Determinants of the incentivised grades

Now, we estimate different models to shed light on the key determinants of the incentives' program effects on students' incentivised grades. This analysis not only supports our previous results based on descriptive analysis, but also controls for the potential influences of confounding factors. We include the same covariates as before.

Table 4

Models explaining incentivised grades. Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Incentivised grade	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
T2	1.354*** (0.431)	1.252*** (0.440)	1.146** (0.442)	1.146** (0.443)	0.928** (0.406)	0.891** (0.365)	0.863** (0.372)
T3	1.357*** (0.414)	1.395*** (0.419)	1.303*** (0.420)	1.303*** (0.421)	1.334*** (0.384)	1.141*** (0.348)	1.115*** (0.354)
Non-incentivised grade		0.196*** (0.0669)	0.199*** (0.0665)	0.199*** (0.0668)	0.136** (0.0620)	0.257*** (0.0606)	0.262*** (0.0627)
Returning			0.626* (0.367)	0.626* (0.377)	1.065*** (0.353)	0.868** (0.341)	0.839** (0.370)
Gender				0.00143 (0.360)	0.221 (0.341)	-0.240 (0.314)	-0.253 (0.319)
Management					-0.394 (0.416)	-0.156 (0.380)	-0.153 (0.385)
Economics					-0.123 (0.463)	0.135 (0.428)	0.130 (0.432)
Economics + Law					3.952*** (0.773)	3.849*** (0.740)	3.983*** (0.828)
Class group A						0.0634 (0.445)	0.123 (0.487)
Class group B						2.177*** (0.470)	2.187*** (0.475)
Class group D						-0.136 (0.553)	-0.142 (0.558)
Class group D						-1.430*** (0.525)	-1.389** (0.540)
Class group F						-0.368 (0.573)	-0.393 (0.580)
UEG							-0.0438 (0.123)
Constant	2.955*** (0.298)	2.181*** (0.405)	1.999*** (0.416)	1.998*** (0.479)	2.043*** (0.559)	1.600*** (0.531)	1.952* (1.088)
Observations	177	158	158	158	158	158	156
R-squared	0.073	0.134	0.150	0.150	0.310	0.472	0.469

As long as more covariates are included in the models the explanation power increases and, step-by-step the effect of each explanatory variable can be analysed in a detailed way. The first result obtained is the power of the incentives provided increasing grades. Specifically, the tournament is more

powerful increasing grades than the piece rate, although both of them reach their main purpose. The career path of the student in the same subject (non-incentivised grade) as a predictor of their grade is verified. Another determinant explaining incentivised grades is the fact of being a returning student in the subject: given that, it is not the first time they face the subject's concepts, their grade increases in average more than the one obtained by subjects enrolled in the subject the first time. Gender does not influence the grades under incentives. Additionally, another determinant is the degree they are enrolled in, pointing that those coursing the double degree in economics and law obtained significantly higher grades than those enrolled in accounting. The last factors explaining the grades obtained in the incentivised scenario are the class groups they belong to or in other words, the professor. As we have explained previously, we use class group C as a reference category because the professor is one of the experimentalists. The grades obtained by his/her students are in the mean, finding groups with significantly more and fewer points in the incentivised grades analysed thus, our results are not influenced by this fact. Lastly, the UEG has not any effect in the incentivised grades obtained by the students.

The previous results based on descriptive analysis are corroborated by means of these regressions making our results about the role of the incentives' program designed more convincing.

Result 4: *Both payment mechanisms are effective increasing grades of new and returning students also controlling potential confounding factors. Furthermore, although grades obtained in the midterm exam do not affect students' self-chosen goals, they do explain their final exam grades.*

6. Conclusions

Our paper is the first study that introduces self-chosen academic goals in an incentive-compatible field experiment including two alternative payment mechanisms. We have analysed the effectiveness of monetary incentives based on absolute and relative academic performance using a piece-rate and a rank-order tournament system respectively. New and returning students in Introductory Microeconomics at the University Jaume I were offered the possibility to participate in the incentives program explaining them beforehand the different groups they could be assigned to. Those accepting and showing explicitly their will to participate were randomly assigned to a control (where no incentives were offered) or to two alternative treatment groups (where incentives were offered) and then, they were asked to bet on their own final exam grade. Our experimental design aims to maximise the efficacy of monetary incentives based on self-chosen goals, choosing to a specific task and a limited time horizon, and providing the incentives with immediacy.

Our results suggest that non-incentivised grades obtained in the midterm exam do not influence students' self-chosen goals, but they explain the incentivised grades obtained in the final exam.

Controlling for potential confounding factors as gender, degree, professor and university entrance grade, we find that incentives based on a tournament mechanism are as effective as those based on a piece-rate scheme to increase grades of new and returning students.

Although this paper highlights the importance of providing incentives to students in order to increase their academic performance, policymakers could be worried by the financial resources needed to guarantee its effectiveness. Our results show that a (low-cost) rank-order tournament payment mechanism based on self-chosen goals can be as effective as a (resource intensive) piece-rate mechanism. Consequently, government policies should be able to support the development of effective incentive programs, allowing students to compete for rewards based on their self-chosen goals.

References

- Angrist, J., Daniel L., & Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *Journal of Applied Economics*, 1 (1), 136–63.
- Angrist, J., & Lavy, V. (2009). The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review*, 99 (4), 1384–1414.
- Angrist, J., Oreopoulos, P. & Williams, T. (2014). When Opportunity Knocks, Who Answers? New Evidence on Collage Achievement Awards. *Journal of Human Resources* 49 (3), 572–610.
- Bandura, A. (2006). Guide for Constructing *Self-Efficacy Scales*. In *Self-Efficacy Beliefs of Adolescents*, edited by Frank. Pajares and Timothy C. Urdan, 307–37. Greenwich, CT: Information Age Pub.
- Barrow, L., & Rouse, C. E. (2016). Financial Incentives and Educational Investment: The Impact of Performance-Based Scholarships on Student Time Use. *Education Finance and Policy*, 1–55.
- Bettinger, E. P. (2012). Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores. *Review of Economics and Statistics*, 94 (3), 686–98.
- Braun, H., Irwin, K., & Yamamoto, K. (2011). An Experimental Study of the Effects of Monetary Incentives on Performance on the 12th-Grade NAEP Reading Assessment. *Teachers College Record*, 113 (11), 2309–44.

- Camerer, C. F., & Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19 (1-3), 7-42.
- Castleman, B. L., Long, T., Bridget, A., Chris, M., Richard, G., & Benjamin, L. (2014). The Impact of Partial and Full Merit Scholarships on College Entry and Success: Evidence from the Florida Bright Futures Scholarship Program. 17.
- Cha, P., & Patel, R. (2010). Rewarding Progress, Reducing Debt: Early Results from the Performance-Based Scholarship Demonstration in Ohio.
- Clark, D., Gill, D., Prowse, V., & Rush, M. (2016). Using Goals to Motivate College Students: Theory and Evidence from Field Experiments.
- Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student Responses to Merit Scholarship Retention Rules. *Journal of Human Resources*, XL (4), 895-917.
- Croson, R. (2005). The Method of Experimental Economics. *International Negotiation*, 10 (1), 131-48.
- Dalton, P. S., Gonzalez, V., & Noussair, C. N. (2015). Paying with Self-Chosen Goals: Incentives and Gender Differences.
- De Paola, M., Scoppa, V., & Nisticò, R. (2012). Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment. *Journal of Human Capital* 6 (1), 56-85

Dykstra, A. (2015). A Literature Review of Expectancy & Goal-Setting Theories When Meeting Organizational Goals and Effecting Change for Diverse Populations. Southern New Hampshire University.

Falk, A., and Knell, M. (2004). Choosing the Joneses: Endogenous Goals and Reference Standards. *Scandinavian Journal of Economics*, 106 (3), 417–35.

Fryer, R. G. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *The Quarterly Journal of Economics*, 126 (4), 1755–98.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25 (4), 191–210.

Goerg, S. J., & Kube, S. (2012). Goals (Th)at Work – Goals, Monetary Incentives, and Workers' Performance. MPI Collective Goods Preprint, 2012/19.

Grove, W. A., & Wasserman, T. (2006). Incentives and Student Learning: A Natural Experiment with Economics Problem Sets. *American Economic Review*, 96 (2), 447–52.

Henry, G. T., Rubenstein, R., & Bugler, D. T. (2004). Is HOPE Enough? Impacts of Receiving and Losing Merit-Based Financial Aid. *Educational Policy* 18 (5), 686–709.

Jackson, C. K. (2010). A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program. *Journal of Human Resources*, 45 (3), 591–639.

Kaur, S., Kremer, M., & Mullainathan, S. (2010). Self-Control and the Development of Work Arrangements. *American Economic Review: Papers and Proceedings*, 100 (2), 624–28.

Latham, G. P., & Locke, E. A. (1979). Goal setting—A Motivational Technique That Works. *Organizational Dynamics*, 8 (2), 68–80.

Lavecchia, A., Liu, H. & Oreopoulos, P. (2014). Behavioral Economics of Education: Progress and Possibilities. 20609.

Lazear, E. P., & Rosen, S. (2009). Rank and Order Tournament: An Optimal Labor Contract. *Journal of Political Economy*, 89, 841–456.

Leuven, E., Oosterbeek, H., Sonnemans, J., & Van Der Klaauw, B. (2001). Incentives versus Sorting in Tournaments: Evidence from a Field Experiment. *Journal of Labor Economics*, 29 (3), 637–58.

Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2010). The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment. *Journal of the European Economic Association*, 8(6), 1243-1265.

Levitt, S. D, List, J. A., Neckermann, S., & Sadoff, S. (2011). The Impact of Short-Term Incentives on Student Performance. Becker Friedman Institute for Research in Economics, University of Chicago

Locke, E. A., & Latham, G. P. (2006). New Directions in Goal-Setting Theory. *Current Directions in Psychological Science*, 15 (5), 265–68.

Locke, E. A., & Latham, G. P. (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist*, *57* (9), 705–17.

O'Neil, H. F. (1997). Final Report of Experimental Studies on Motivation and NAEP Test Performance. Los Angeles.

O'Neil, Jr., H. F., Sugrue, B., & Baker, E. L. (1995). Effects of Motivational Interventions on the National Assessment of Educational Progress Mathematics Performance. *Educational Assessment*, *3* (2), 135-157

Pajares, F. (2008). Motivation and Self-Regulated Learning: Beliefs in Self-Regulated Learning. In *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, edited by Dale H Schunk and Barry J. Zimmerman, 111–39. Lawrence Erlbaum Associates.

Patel, R., & Rudd, T. (2012). Can Scholarships Alone Help Students Succeed?

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, *7* (3), 51–84.

Sackett, A. M., Wu, G., White, R. J. & Markle, A. (2014). Harnessing Optimism: How Eliciting Goals Improves Performance. *SSRN Electronic Journal*.

Samek, A. S. (2016). Gifts and Goals: Behavioral Nudges to Improve Child Food Choice at School.

Scott-Clayton, J. (2011). On Money and Motivation: A Quasi Experimental Analysis of Financial Incentives for College Achievement. *Journal of Human Resources*, 46 (3), 614–46.

Sjoquist, D. L., & Winters, J. V. (2012). Building the Stock of College-Educated Labor Revisited. *Journal of Human Resource,s* 47 (1), 270–85.

van Dijk, F., Sonnemans, J., & van Winden, F. (2001). Incentive Systems in a Real Effort Experiment. *European Economic Review*, 45 (2), 187–214.

van Lent, M., and Souverijn, M. (2016). Goal Setting and Raising the Bar: A Field Experiment. 2017/-001/VII.

Vandegrift, D., Yavas, A., & Brown, P. M. (2007). Incentive Effects and Overcrowding in Tournaments: An Experimental Analysis. *Experimental Economics*, 10 (4), 345–68.

Zimmerman, B.J. (2011). Motivational Sources and Outcomes of Self-Regulated Learning and Performance. In *Handbook of Self-Regulation of Learning and Performance*, edited by Barry J. Zimmerman and Dale H. Schunk, 49–64. New York: Routledge.

Chapter 2. Framing and repetition effects on risky choices: a behavioural approach

Abstract

Framing effects play an important role in individual decision making under risk. This investigation revisits framing effects caused by two versions of the choice list procedure, lottery vs. lottery (LL) and lottery vs. certainty (LC). In the former, subjects face pairwise choices between lotteries within a choice list. In the latter, subjects are asked to state a minimum safe amount to give up the lottery they are endowed with. In order to elicit risk preferences, we implement an incentive-compatible experiment that allows imprecision and controls the proper functioning of the random lottery incentive mechanism. Particularly, it is tested whether variations in the number of options offered in a choice list with and without variations in the range of options affect subjects' choices. Our results suggest that changes in framework disturb subjects' risk preferences only in the LC version when the range of options presented has been decreased.

1. Introduction

Risk attitude is known to be a key determinant of various economic and financial choices. Behavioural studies that aim to evaluate the role of risk attitude in contexts of this type require tools for measuring risk aversion at both the individual and aggregate levels. The most frequent procedure to elicit individual risk attitudes is referred as the choice list procedure. The choice list method presents a table of binary choices designed so that as a respondent works through the table, he can be expected to switch at some point from “one side” to the other. Two alternative versions of this procedure are the lottery vs. lottery (LL) and lottery vs. certainty (LC) methods. In the LL method, subjects face pairwise choices between lotteries within a choice list. A famous example of this method was proposed by Holt and Laury (2002, HL hereafter), in which subjects are given a list of 10 decisions between paired lottery choices where payoffs are constant and probabilities vary. Our closed-elicitation LC method is based on a Becker-DeGroot-Marschak (BDM) auction in which subjects are asked to choose a minimum safe amount to give up the lottery they have been endowed with. In order to determine the subject’s payoff in a choice list, the Random Lottery Incentive (RLI) payment mechanism is used to pick randomly one decision from the list. If the isolation hypothesis from prospect theory (which implies that subjects evaluate each risk task independently of the other tasks) is maintained, each pairwise choice a subject makes in the list can be interpreted as if he had faced only a single binary choice.

This paper investigates whether subjects' choices are influenced by framing effects originated by the choice list procedure. Framing effect is a decision-making bias in which subjects choose differently when the same problem is presented in diverse ways. This effect can be referred to multiple issues that may affect the presentation of the risk task implemented. The best-known framing effects are related to variations in the order, the number and range of the options presented in the risk choice task. Using the experimental task named the Asian Disease Problem, the framing effect was first presented by Tversky and Kahneman (1981). They highlighted the importance of changes in individual preferences because of inconsequential changes in the formulation of choice problems. In a subsequent paper, Tversky and Kahneman (1986) argued that framing effects violate the normative condition of description invariance, which stipulates that the same problem should be evaluated in like manner regardless its description. An example of models assuming this principle is the Expected Utility Theory (EUT) model where choice options are evaluated strictly as a function of probability and payoff, with no specification of probability-payoff framing.

Results on framing effects are a mixed bag depending on the method used to elicit subjects' risk aversion and the type of framing effect analysed (ordering effects, changes in the number of options presented with or without affecting their range, simultaneous versus sequential presentation of lotteries, etc.). Hey and Orme (1994) found that when the same 100 pairs of ternary lotteries were repeated two times on separate days (with a possibility to declare

indifference) in different order, subjects chose identical options for each pair in around 75% of all cases. Increasing the number of repetitions with respect to his previous paper, Hey (2001) found that some individuals maintained a constant variability in their responses to a 100 pairwise risky-choice questions in spite of their repetition. Using the HL method, Andersen *et al.* (2006) found that choices were affected by ordering effects and the range of a given lottery. Specifically, they found out that the deletion of the worst pairs (with the lowest expected value) of lotteries increased risk aversion. Additionally, the authors showed that enforcing only one switching point, strict monotonicity and transitivity, had no systematic effect. Lévy-Garboua *et al.* (2012) presented experimental evidence of how framing affected decisions in the context of the HL procedure. They found that presenting lotteries simultaneously induced significantly less inconsistency than showing lotteries in sequential appearance. Additionally, both repetition of identical choices and high payoffs reduced inconsistency too. Bosch-Domènech and Silvestre (2013) found what they called “an embedding bias”. This bias implies that when some specific pairs of alternative lotteries are removed, risk aversion becomes less frequent and the ranking of individuals by risk aversion is not preserved. However, the aforementioned bias was not found when they analysed the certainty equivalent (CE) elicitation method. Contrary to these results, Freeman *et al.* (2016) found that embedding a pairwise choice in a choice list increased the fraction of subjects choosing the riskier lottery when the safer alternative was certain, but it did not significantly affect choices when the safer alternative is risky. Erev *et al.* (2008) and Blavatskyy and Koehler (2009) analysed the

robustness of CE mechanism to elicit risk preferences obtaining that elicited payoffs were systematically affected by the range of certain payoffs to which the lottery was compared. Beauchamp *et al.* (2012) studied how risk aversion parameters were affected by the manipulation of the intermediate pairs of options without affecting the range of options. They found that when the endpoints of the multiple price list were fixed and intermediate outcomes were decreased, participants' choices became significantly more risk averse. Finally, Loomes and Pobregna (2014) used three elicitation methods finding a considerable variability within -and even more, between- the results they produced. This finding suggested that not only different elicitation instruments but also framing-specific issues could interact with imprecise underlying preferences.

However, framing effects in the literature rely on two crucial assumptions: (1) to suppose that subjects have a clear preference, choosing always the same answer to exactly the same question and (2) the fulfilment of the isolation hypothesis.

Imprecision occurs when subjects do not have a clear choice between some options, generally implying contiguity of switching points. Lévy-Garboua *et al.* (2012) pointed out that “even in decision experiments where subjects make repeated independent and identically distributed decisions among pairs of lotteries without any alteration” (p. 129) an estimable quantity of subjects reported different options over repetition. Supporting this evidence,

experiments by Ballinger and Wilcox (1997) and Loomes and Sugden (1998) sustained that repetition drove subjects toward increasingly safer choices. Besides, Butler and Loomes (2011) suggested that the violations of the EUT could be explained by imprecision. Later, Loomes and Pogrebna (2014) found that most subjects showed variability when they answered to some questions with the aim of eliciting their risk attitude. In addition, they pointed out that the imprecision that subjects exhibited in their preferences could produce that preferences depended on the effects of the procedure. More recently, Cubitt *et al.* (2015) designed a new measure in which subjects had to compare a given lottery with a safe amount of money, identifying the options for which they preferred the lottery, the safe amount or for which they did not have a clear choice. In order to reward imprecise subjects, a specific switching point within their imprecision interval was asked to them. The authors obtained that there existed imprecision, which was persistent across various lotteries. Their main finding was that the measure they constructed in order to test imprecision varied across lotteries in an “intelligible and systematic way” (p. 5) but in contrast, it did not have a systematic change with repetition or experience. We allow subjects’ imprecision including the repetition of identical risk choice-lists in order to analyse framing effects allowing preference imprecision.

The second crucial assumption is related to the fact that the RLI mechanism provides incentives for truthful revelation of preferences. This standard payment protocol in individual risky choice experiments involves a subject making $K > 1$ binary choices over objective lotteries, and then selecting

one choice at random for payment. If this assumption is sustained, there should not be differences between risk preferences revealed under RLI and in the case where a subject makes only one choice, and then he is being paid with certainty for the single choice (1-in-1 payment procedure). Although the RLI payment system has been widely accepted by experimentalists, there are some studies pointing out that this mechanism could not work properly in some cases. Holt (1986) presented a theoretical objection arguing that, if the reduction of compound lotteries (ROCL) axiom holds, a failure of the expected utility compound independence axiom (CIA) would suffice to reject the RLI mechanism compatibility.

Starmer and Sugden (1991) were the first ones who tested whether the subjects' behaviour in random-lottery experiments was consistent with the ROCL assumption. Showing that the reduction principle did not hold, they discarded Holt's conjecture. Comparing choices of subjects in 1-in-1 with RLI payment procedures in experiments, Beattie and Loomes (1997) and Cubitt *et al.* (1998) supported the conclusion that the RLI payoff mechanism elicited true preferences. Contrary to this result and directly testing the CIA, Harrison and Swarthout (2014) showed that risk preference estimates obtained under RLI mechanism differed from those obtained in a 1-in-1 design. Complementary to this paper, Cox *et al.* (2014) found that risk preferences could be manipulated by integrating a second, asymmetrically dominated choice problem in a RLI mechanism behaviour. In the same vein, Harrison *et al.* (2014) highlighted the apparent problem of inferring preferences using the RLI and treating these

results “as if” they were the same as those from a 1-in-1 scenario. This concern was shared by Cox et al. (2015) showing large differences across mechanisms in subjects’ revealed risk preferences. Lastly, Brown and Healy (2016) found out that RLI was not incentive-compatible when all decisions were displayed in a standard list format but it was restored when the rows of the list were randomised and shown on separate screens.

Our study argues that it cannot be concluded that changes in decisions are necessarily due to changes in framing if subjects make different decisions in identical sequentially repeated risk tasks. In this sense, we depart from previous literature because we analyse framing effects taking into account contamination effects due to preference imprecision and the proper functioning of the RLI mechanism.

Isolating these effects, our results point out that the LL elicitation method is robust to manipulations in the number and/or the range of options offered in the list. Nevertheless, the LC method is not that robust because changes in the task structure modify subjects’ preferences.

2. Experimental Design

In order to study framing effects in the multiple choice list procedure we test for shifts in risk preferences due to: (1) a (a)symmetric increase of the

number of pairs offered keeping constant the range (CR) of options, and (2) a (a)symmetric decrease of the number of pairs varying the range (VR) of options offered.

CR and VR changes are analysed using both elicitation methods. Treatment 1 (T1) and treatment 2 (T2) correspond to the LL elicitation method for CR and VR changes respectively. Treatments 3 (T3) and treatment 4 (T4) are related to the LC method for CR and VR modifications respectively. Following Gonzalez and Wu (1999), we ask subjects to choose which row they want to switch at to fill out the remaining choices for the subject.

A total of 141 subjects (34 in T1, 36 in T2 and T3, and 35 in T4) were recruited among undergraduate students from different economics or business-related courses from the University Jaume I (UJI), using standard recruitment procedures with an open call for subjects through the LEE (Laboratorio de Economía Experimental) website. Before the beginning of each session, subjects were given written instructions, which were also read aloud by the organisers. Any remaining questions were privately answered.

At the end of each session, subjects responded to a questionnaire, asking them to report the main reason why, if this was the case, they have varied choices across the different repetitions. After that, they were privately paid in cash. All sessions were computerised and carried out in a specialised computer laboratory, using software based on the Z-Tree toolbox by Fischbacher (2007).

In the case of lottery vs. lottery, we present LL_{17} in table 5. The name obeys to the fact that in this task subjects face a list of seventeen pairs of lotteries, which we numerate from one to seventeen, each pair involving a “safe” lottery (S) and a “risky” lottery (R). These labels are provided since if we compare lottery R with S, R offers the best payoff and the worst (null) payoff. The last three columns in Table 5 (not shown to the experimental subjects) indicate the expected euro values of the safe lottery in the pair (denoted EVs) and that of the risky lottery (denoted EV_R), as well as the difference between the two. For the first fifteen rows, the risky option offers the higher expected value (EV) while for the last two rows, the safe option offers the higher EV, with the difference between EVs decreasing as we go down the list. Thus, a risk-neutral individual would select the R lottery in all pairs with the exception of the last two. Subjects’ payoffs are selected in order to offer: (1) a sufficient reward to subjects in an experiment with multiple risk task repetitions and random lottery incentive (RLI) as payment mechanism, and (2) a wide number of pairs of lotteries where EVR exceeds EVs.

Table 5Pairs of lotteries offered in LL₁₇

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV_S	EV_R	$EV_S - EV_R$
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
1	10%	17.5€	90%	26.70€	10%	0.0€	90%	100.0€	25.78€	90.0€	- 64.22€
2	15%	17.5€	85%	26.70€	15%	0.0€	85%	100.0€	25.32€	85.0€	- 59.68€
3	20%	17.5€	80%	26.70€	20%	0.0€	80%	100.0€	24.86€	80.0€	- 55.14€
4	25%	17.5€	75%	26.70€	25%	0.0€	75%	100.0€	24.40€	75.0€	- 50.60€
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
6	35%	17.5€	65%	26.70€	35%	0.0€	65%	100.0€	23.48€	65.0€	- 41.52€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
8	45%	17.5€	55%	26.70€	45%	0.0€	55%	100.0€	23.56€	55.0€	- 31.44€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
10	55%	17.5€	45%	26.70€	55%	0.0€	45%	100.0€	21.64€	45.0€	- 23.36€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
12	65%	17.5€	35%	26.70€	65%	0.0€	35%	100.0€	20.72€	35.0€	- 14.28€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€
14	75%	17.5€	25%	26.70€	75%	0.0€	25%	100.0€	19.70€	25.0€	- 5.30€
15	80%	17.5€	20%	26.70€	80%	0.0€	20%	100.0€	19.34€	20.0€	- 0.66€
16	85%	17.5€	15%	26.70€	85%	0.0€	15%	100.0€	18.88€	15.0€	3.88€
17	90%	17.5€	10%	26.70€	90%	0.0€	10%	100.0€	18.42€	10.0€	8.42€

Using LL₁₇, we construct four additional risk tasks: LL₁₃, LL₉, LL₇ and LL₅. LL₁₃ contains the odd pairs of LL₁₇ plus all pairs from ten to seventeen; LL₉ is composed of all odd numbers from one to seventeen, LL₇ comprises odd pairs from five to seventeen and LL₅ contains odd numbers from five to thirteen. Treatment 1 (2) is formed by tasks LL₉, LL₁₇ (LL₅), LL₁₃ (LL₇). As can be seen by the design, the baseline task is LL₉ given that it is used in both treatments. Thus,

task LL₁₇ (LL₁₃) symmetrically (asymmetrically) increases the number of pairs of lotteries keeping constant the range of options from 10-90%. In the same way, task LL₇ (LL₅) decreases symmetrically (asymmetrically) the number of options reducing the range of options to 30-90% and 30-70% respectively.

Subjects are asked to make repeated independent and identically distributed (i.i.d.) decisions among lists of pairs of options without any alteration. Specifically, in treatment 1 (2), each subject faces tasks LL₉, LL₁₇ (LL₅) and LL₁₃ (LL₇) six times each one of them, in a totally random order. Repetition of identical choice lists allows us to test for possible preference imprecision without explicitly ask subjects if they are sure about their risk preferences. However, subjects only were classified as imprecise if they self-reported as a cause of their variability that they were not sure about their preferred switching point in i.i.d. risk tasks. Nevertheless, we do not consider as imprecise to those subjects self-reporting that previous chosen switching points affected their subsequent decision. The latter were treated as inconsistent with the isolation hypothesis that is, a failure in the functioning of the RLI mechanism.

In both treatments, we inform subjects that three draws would be implemented to determine their payment. A first draw is carried out to choose which one of their 18 tasks will be selected; a second draw is used to randomly choose one from all pairs of lotteries contained in the selected task; a third draw given the odds of the lottery preferred by the subject in the pair, will be applied to determine individual payoffs. This design rules out possible wealth

effects due to subjects' (expected) earnings from previous periods.

As regard the LC elicitation method, we use two lotteries: a safe lottery (S) and a risky one (R). In table 6 (table 7), we present LC_{21}^S (LC_{31}^R). In these tasks, subjects must choose between the lottery (S or R) they have been endowed with and safe amounts of money that increase by 0.50€ in LC_{21}^S and 1.50€ in LC_{31}^R for each additional pair. The last columns of these tables (not shown to the experimental subjects) indicate the difference between the expected euro value of the lottery in the pair (denoted EV_S) and the safe amount of money. A risk-neutral individual would select the safe lottery in all pairs of the table 6 excepting the two last ones. However, a risk-neutral subject would choose the R lottery only in the three first pairs of table 7.

Table 6

Pairs of lotteries offered in CE_{21}^S .

Pair		Safe lottery			Safe amount (SA)	EV_S-SA
1	81%	10.0€	19%	0.00€	0.00€	8.10€
2	81%	10.0€	19%	0.00€	0.50€	7.60€
...		
20	81%	10.0€	19%	0.00€	9.50€	-1.40€
21	81%	10.0€	19%	0.00€	10.00€	-1.90€

Table 7

Pairs of lotteries offered in CE^R_{31} .

Pair		Risky lottery			Safe amount (SA)	$EV_R - SA$
1	19%	45.0€	81%	0.00€	0.00€	8.55€
2	19%	45.0€	81%	0.00€	1.50€	7.05€
...		
30	19%	45.0€	81%	0.00€	43.50€	-35.05€
31	19%	45.0€	81%	0.00€	45.00€	-36.55€

Using LC^S_{21} to construct the safe lotteries vs. certainty, we create four additional risk tasks: LC^S_{11} , LC^S_{16} , LC^S_5 and LC^S_8 . Task LC^S_{11} contains all odd numbers from LC^S_{21} ; LC^S_{16} is comprised of all odd numbers of LC^S_{21} and all pairs from one to eleven; LC^S_5 comprises all odd pairs from seven to fifteen; and LC^S_8 contains all odd pairs from one to fifteen. Treatment 3 (4) for safe lotteries contains LC^S_{11} , LC^S_{21} (LC^S_5), LC^S_{16} (LC^S_8). As can be detected, the baseline task for safe lotteries is LC^S_{11} . In this way, task LC^S_{21} (LC^S_{16}) symmetrically (asymmetrically) increases the number of certainty payoffs without varying their range with respect to LC^S_{11} (the range goes from 0.00-10.00€). Additionally, task LC^S_5 (LC^S_8) decreases symmetrically (asymmetrically) the number of options diminishing the range of options offered to subjects with respect to LC^S_{11} . Indeed, the latter goes from 3.00-7.00€ in the case of LC^S_5 and from 0.00-7.00€ in the case of LC^S_8 .

In like manner, we use LC^R_{31} to build four additional risk tasks: LC^R_{16} , LC^R_{23} , LC^R_{10} and LC^R_9 . LC^R_{16} comprises all odd pairs from LC^R_{31} ; LC^R_{23} contains

all odd pairs from LC_{31}^R and all pairs from one to fifteen, LC_{10}^R contains all odd pairs from seven to twenty-five, and lastly, LC_9^R comprises all odd pairs from one to seventeen. Treatment 3 (4) for risky lotteries contains LC_{16}^R , LC_{31}^R (LC_{10}^R), LC_{23}^R (LC_9^R). The task repeated in two treatments and the one that is considered our benchmark is LC_{16}^R . Therefore, task LC_{31}^R (LC_{23}^R) symmetrically (asymmetrically) increases the number of safe amounts without varying their range with respect to LC_{16}^R , the latter goes from 0.00-45.00€. Additionally, task LC_{10}^R (LC_9^R) decreases the range of options offered to subjects with respect to LC_{16}^R by symmetrically (asymmetrically) diminishing the number of certainty payoffs featured.

In treatment 3 (treatment 4), all subjects complete in random order tasks LC_{11}^S , LC_{21}^S (LC_5^S), LC_{16}^S (LC_8^S), LC_{16}^R , LC_{31}^R (LC_{10}^R) and LC_{23}^R (LC_9^R). All tasks are repeated six times in both treatments. In this case, subjects are informed that until three draws could be necessary to calculate their payment, avoiding aforementioned wealth effects. A first draw is used to choose which one of their 36 tasks will be selected; a second draw is put through to choose one from all pairs of options contained in the selected task; in case that the chosen option is the lottery, a third draw will be implemented to obtain subjects payoffs.

To sum up the experimental design, a summary of the treatments is presented in table 8

Table 8

Summary of treatments.

Treatment	Subjects	Tasks	Type of framing effect
T1	34	LL ₉ , LL ₁₇ , LL ₁₃	Constant range (CR)
T2	36	LL ₉ , LL ₅ , LL ₇	Varying range (VR)
T3	36	LC ^S ₁₁ , LC ^S ₂₁ , LC ^S ₁₆ LC ^R ₁₆ , LC ^R ₃₁ , LC ^R ₂₃	Constant range (CR)
T4	35	LC ^S ₁₁ , LC ^S ₅ , LC ^S ₈ LC ^R ₁₆ , LC ^R ₁₀ , LC ^R ₉	Varying range (VR)

3. Data analysis

3.1. Statistical tests

In order to analyse framing effects, we use a Wilcoxon test. Specifically, it is used to compare the percentage of safe choices (in the LL method) or the certainty choices (in the LC method) taking place under two different frameworks for the same sample of subjects. Aiming to treat observations independently, we calculate the percentage of safe choices made by each subject for the six repetitions in a common pair. Then, those percentages are compared between different lotteries for each subject. We apply a Bonferroni

correction to take into account the problem of false positives in multiple pair comparisons.

By repetition of the same risk task six times, we are able to analyse in a within-subject framework the variability of subjects' choices within each i.i.d. task. This allows us to classify subjects depending on their variability within i.i.d. tasks. We name "constant" (C) subjects to individuals who always choose the same switching point in i.i.d. tasks and "inconstant" (IC) subjects to participants who do not choose the same switching point in i.i.d. tasks. Based on answers to our questionnaire, we divide IC subjects in two types: "imprecise" (I) and "inconsistent with the isolation hypothesis" (INC) subjects. We name I subjects to those informing that they do not have a clear choice between some adjacent switching points. INC subjects are those participants who justify their non-adjacent switching points choices reporting to perceive their situation as a grand meta-lottery where they think about all choices in the experiment as a whole. INC subjects violate the isolation hypothesis given that they do not take each list choice separately and for them, the RLI mechanism does not provide incentives for truthful revelation of their preferences. Consequently, we will consider his/her deletion from our sample in part of the analysis.

3.1.1. Lottery vs. Lottery method

Figures 5 and 6 present the average rate of safe choices per pair of options in each LL task included in T1 and T2. In both treatments, the benchmark lottery

is LL₉. In T1 we symmetrically (asymmetrically) increase the number of pairs offered keeping constant the range of options by means of LL₁₇ (LL₁₃). In T2, we symmetrically (asymmetrically) decrease the number of pairs diminishing the range of options offered by means of LL₅ (LL₇).

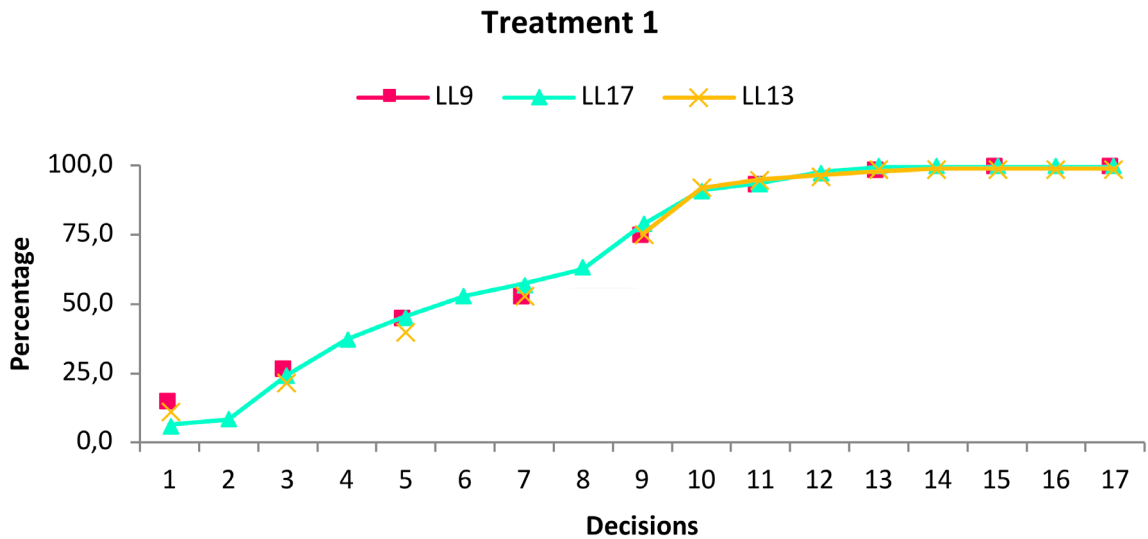


Fig 5. Average rate of safe choices per pair in the LL task in T1.

In T1, where the range is constant, the differences among the average rate of safe choices in the risk tasks is, in general, unnoticeable. Specifically, when we expand symmetrically the number of pairs (from LL₉ to LL₁₇) we do not find significant differences between the percentages of safe lotteries chosen by subjects in the same pair³⁹. An identical result is obtained when the number of options increases asymmetrically (from LL₉ to LL₁₃ or from LL₁₃ to LL₁₇). Therefore, we can conclude that:

39. All Bonferroni-corrected Wilcoxon test p-values corresponding to each pair are above 0.05.

Result 1: An increase (symmetric or asymmetric) in the number of pairs offered in the LL method, keeping constant the range of options, does not produce framing effects.

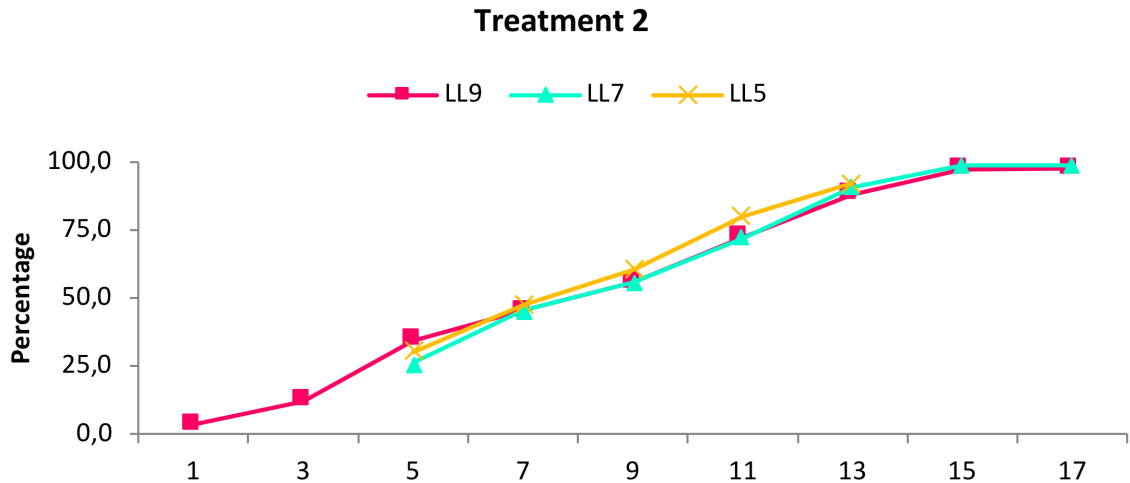


Fig 6. Average rate of safe choices per pair in the LL task in T2.

In T2, we present a symmetric or an asymmetric decrease in the number of pairs offered decreasing the range of options. Specifically, in task LL₇ (LL₅) the range of options offered decreases with respect to the baseline task diminishing symmetrically (asymmetrically) the number of pairs. Comparing LL₉ and LL₇, no significant differences⁴⁰ between the percentages of safe lotteries chosen by subjects in the common pairs are found. The same results are found when we compare LL₉ with LL₅. In consequence, we can state that:

40. All Bonferroni-corrected Wilcoxon test p-values corresponding to each pair are above 0.05.

Result 2: A decrease (symmetric or asymmetric) in the number of pairs offered in the lottery vs. lottery method, reducing the range of options, does not generate framing effects.

The previous analysis is based on the entire sample, including constant and imprecise subjects (C&I hereafter) and those inconsistent. We repeat the above analysis considering C&I subjects and disregarding inconsistent individuals because they do not fulfil the isolation hypothesis.

Figures 7 and 8 reformulate the empirical evidence of figures 1 and 2 presenting the average rate of safe choices per pair of options only for C&I subjects.

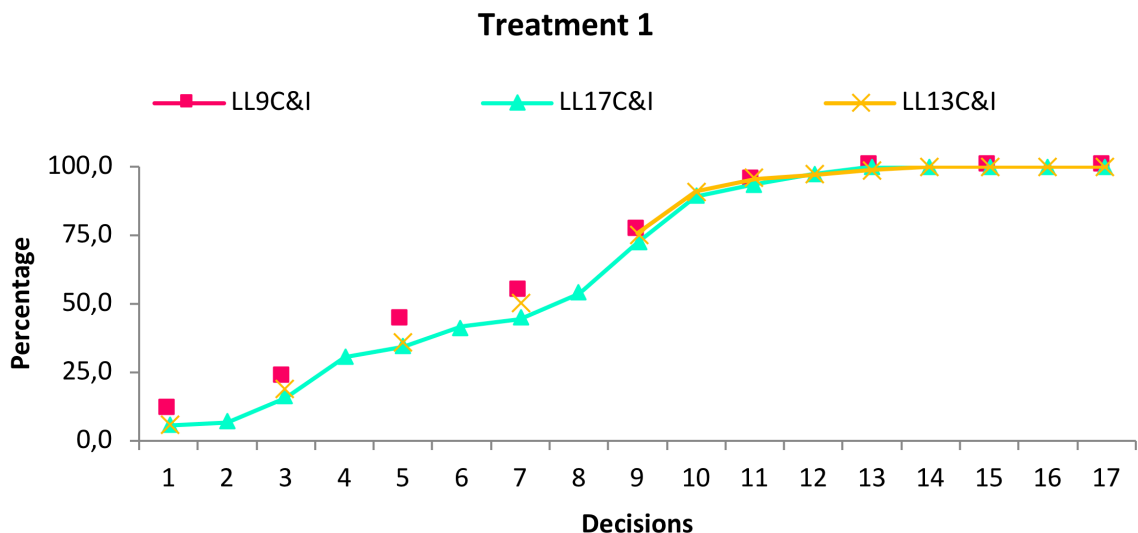


Fig 7. Average rate of safe choices per pair in the LL task for constant and imprecise subjects in T1.

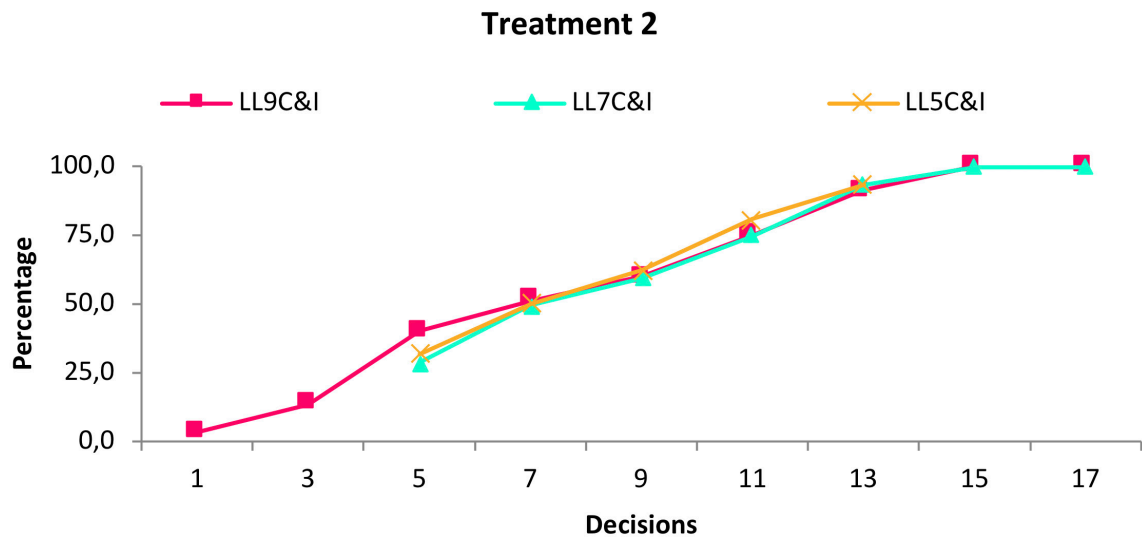


Fig 8. Average rate of safe choices per pair in the LL task for constant and imprecise subjects in T2.

In all aforementioned comparisons, we obtain identical results to the full sample case ones, i.e. no framing effects are found.

Result 3: *In both cases, considering and disregarding inconsistent subjects, no framing effects are found in the lottery vs. lottery method.⁴¹*

These results contrast with some authors who have analysed the same method searching for framing effects. Andersen *et al.* (2006) found that choices were affected by order and lottery range when they deleted the two worst pairs. More recently, Bosch-Domènech and Silvestre (2013) found out that when some pairs were removed the subjects' choices change, what they called embedding bias.

41. The same conclusion is reached when only the first decision in each pair is used instead of the mean of the six repetitions.

Lastly, we check whether our non-significant differences are due to a lack of statistical power running an ex-post power analysis using power set at 90% and probability at 5%. For the case of lottery vs. lottery, sample sizes would have to increase up to at least $N=318$ in order to find framing effects.

3.1.2. Lottery vs. Certainty method

Framing effects in the LC method are analysed by means of T3 and T4. In T3, we symmetrically/asymmetrically increase the number of certainty payoffs keeping constant their range respect to the baseline tasks (LC_{11}^S or LC_{16}^R). In T4, we decrease the range of options offered to subjects respect to the baseline tasks symmetrically/asymmetrically diminishing the number of certainty payoffs.

Figure 9 and 10 display the average percentage of certain choices in both the safe and the risky lottery respectively per pair in the LC tasks presented in T3.

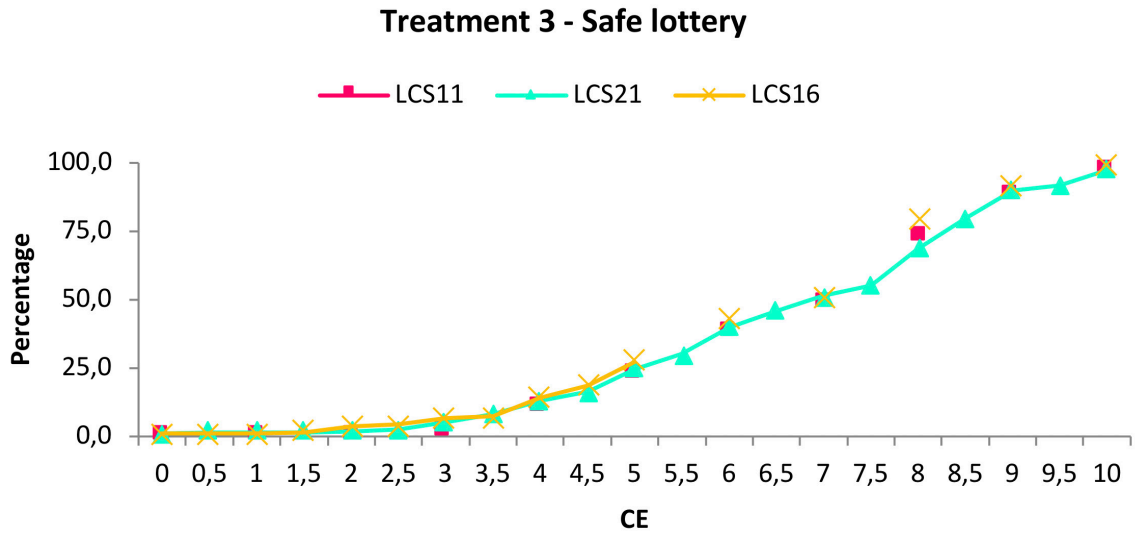


Fig 9. Average percentage of certain choices per pair preferred to the safe lottery in T3.

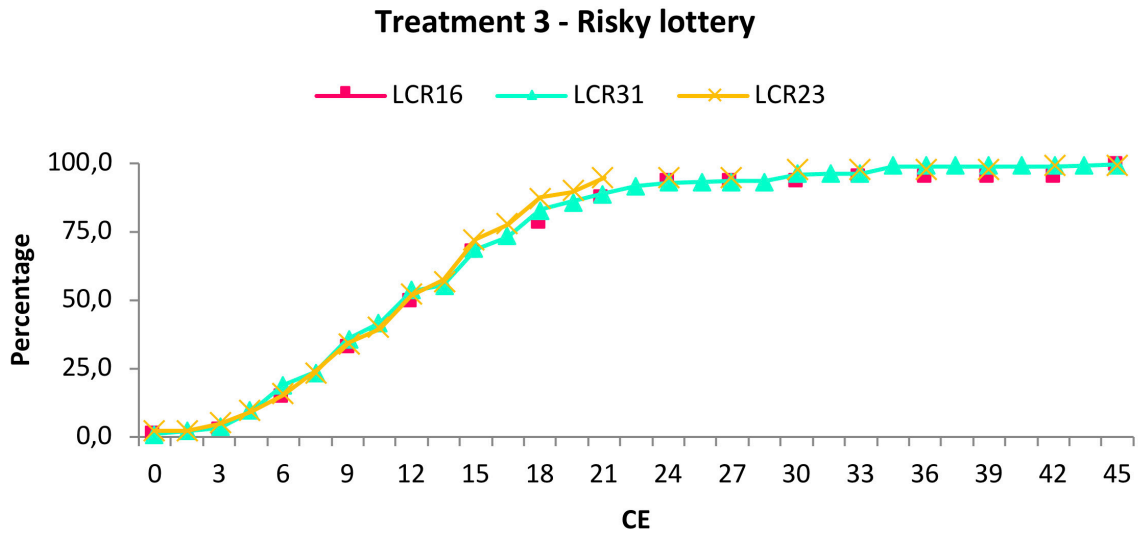


Fig 10. Average percentage of certain choices per pair preferred to the risky lottery in T3.

In general, when we symmetrically (from LC^S_{11} to LC^S_{21} and from LC^R_{16} to LC^R_{31}) or asymmetrically (from LC^S_{11} to LC^S_{16} or from LC^S_{16} to LC^S_{21} and from LC^R_{16} to LC^R_{23} or from LC^R_{23} to LC^R_{31}) increase the number of certainty payoffs, without changing the range of options, we do not find significant differences

between the percentage of safe amounts chosen by subjects in the same pair. However, a framing effect is found when we compare $LC^{R_{16}}$ and $LC^{R_{23}}$ for a safe amount of 18€.⁴²

Result 4: *In some cases, an asymmetric expansion of the number of pairs offered keeping constant the range of options may lead to a framing effect in the lottery vs. certainty method.*

Figures 11 and 12 present the average percentage of certain choices in both the safe and the risky lotteries respectively per pair in the LC task offered in T4, in which the range of options the subject copes to has been reduced.

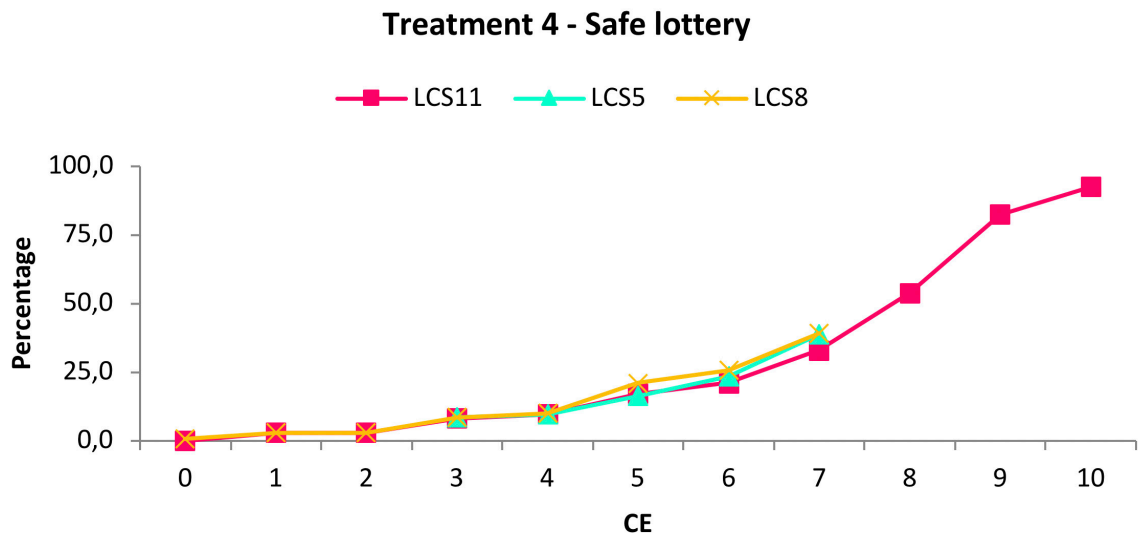


Fig 11. Average percentage of certain choices per pair preferred to the safe lottery in T4.

42. There is a framing effect after computing the Bonferroni-corrected Wilcoxon at a 10% level of significance.

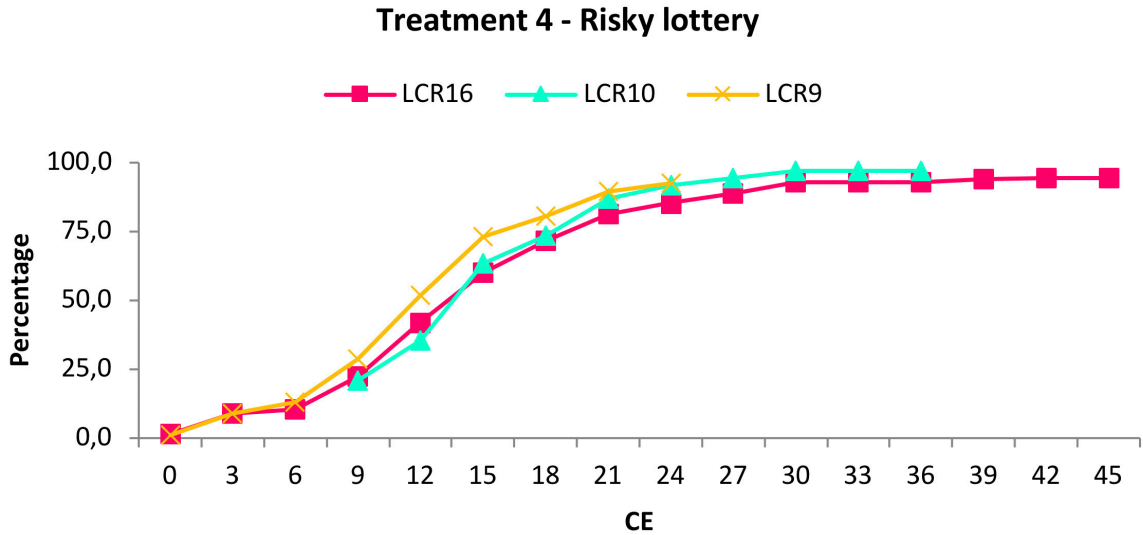


Fig 12. Average percentage of certain choices per pair preferred to the safe lottery in T4.

It is important to note that in the safe (risky) lottery, we symmetrically reduce the number of pairs between $LC^{S_{11}}$ and LC^S_5 ($LC^{R_{16}}$ and $LC^{R_{10}}$), whereas the number of certainty payoffs is asymmetrically decreased between $LC^{S_{11}}$ and LC^S_8 ($LC^{R_{16}}$ and LC^{R_9}) and between LC^S_5 and LC^S_8 ($LC^{R_{10}}$ and LC^{R_9}).

We find significant differences between the percentage of certain choices selected only for the risky lottery in the following cases: (a) comparing $LC^{R_{16}}$ and $LC^{R_{10}}$ for safe amounts 24€, 27€, 30€, 33€ and 36€⁴³; (b) comparing $LC^{R_{10}}$ and LC^{R_9} for safe amounts of 12€ and 18€⁴⁴; and (c) in the comparison

43. Bonferroni-corrected Wilcoxon test p-values are 0.07, 0.06, 0.07, 0.07 and 0.07 respectively after multiplying the original p-values by 10.

44. Bonferroni-corrected Wilcoxon test p-values are 0.012 and 0.090 after multiplying the original p-values by 6.

between LC^{R}_{16} and LC^{R}_{9} for safe amounts of 9€, 12€, 15€ and 18€⁴⁵.

Result 5: A reduction (symmetric or asymmetric) in the number of safe amounts diminishing the range of options produces marginal framing effects for a large number of pairs in the lottery vs. certainty method.

Like in the lottery vs. lottery method, we remove from our sample the inconsistent subjects, thus only constant and imprecise subjects are considered.

Figures 13 and 14 present the average percentage of certain choices per pair preferred to the safe and the risky lottery respectively, for constant and imprecise subjects in T3.

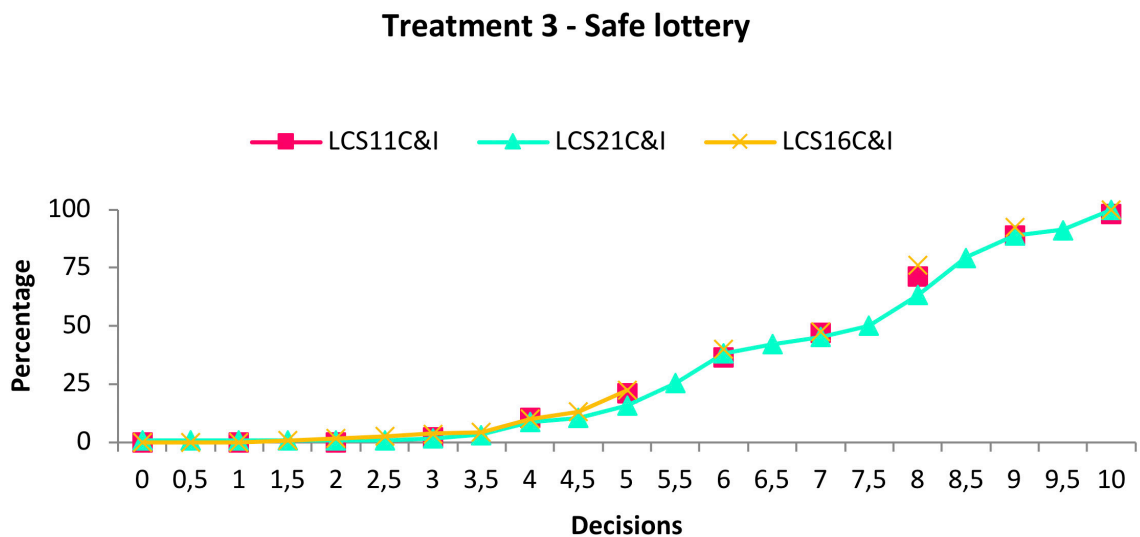


Fig 13. Average percentage of certain choices per pair preferred to the safe lottery in T3.

45. Bonferroni-corrected Wilcoxon test p-values are 0.063, 0.012, 0.090 and 0.072 after multiplying the original p-values by 9.

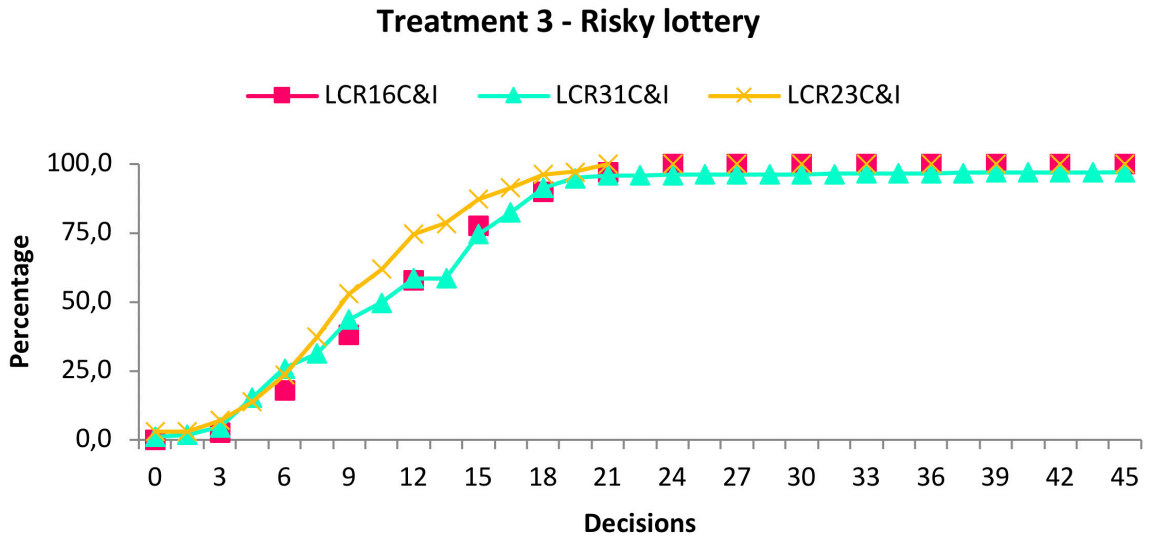


Fig 14. Average percentage of certain choices per pair preferred to the risky lottery in T3.

A (a)symmetric increase in the number of certainty payoffs without changing the range of options and removing inconsistent subjects does not produce framing effects.⁴⁶

Result 6: *An increase (symmetric or asymmetric) in the number of safe amounts keeping constant the range of options does not produce framing effects in the lottery vs. certainty method⁴⁷ when inconsistent subjects are removed.*

In figures 15 and 16, we present the average percentage of certain choices per pair preferred to the safe and the risky lottery respectively, for constant and imprecise subjects in T4.

46. All Bonferroni-corrected Wilcoxon test p-values are above 0.05

47. Although in figure 10, there seems to be framing effects, these disappear once we apply the Bonferroni correction to avoid the existence of false positives.

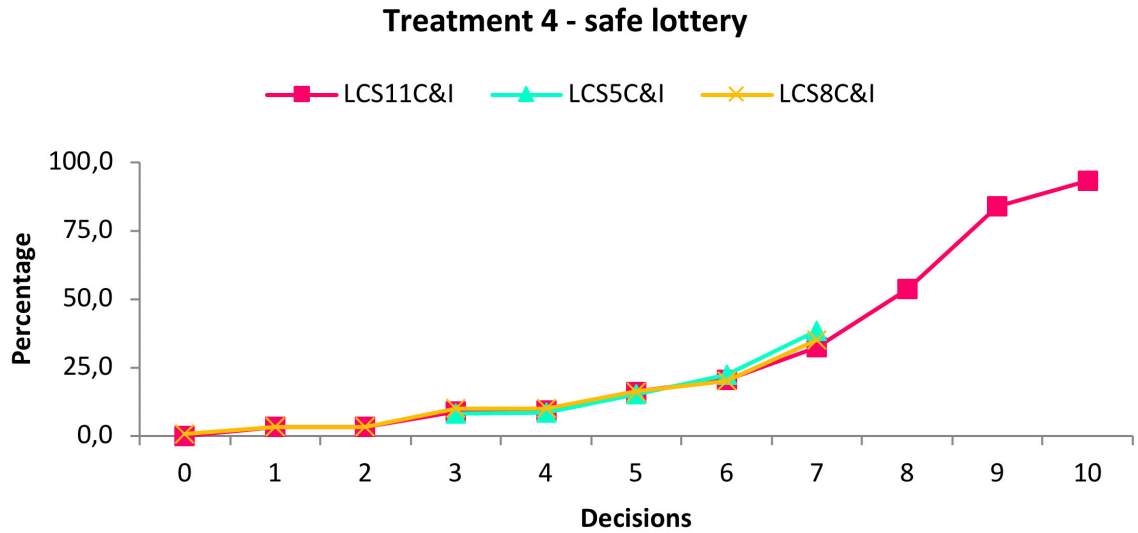


Fig 15. Average percentage of certain choices per pair preferred to the safe lottery in T4.

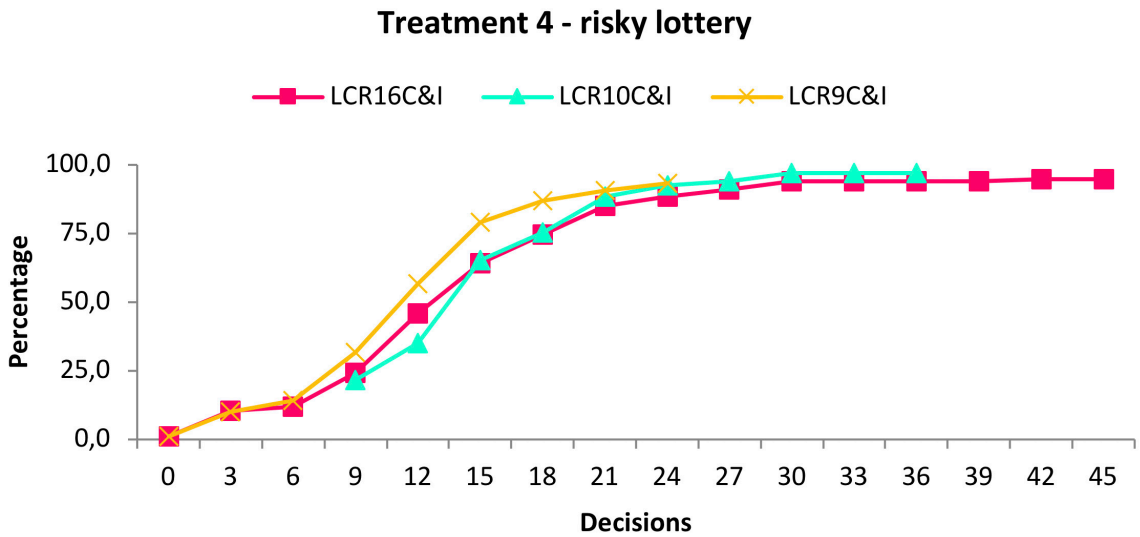


Fig 16. Average percentage of certain choices per pair preferred to the risky lottery in T4.

Unlike the full sample case, when we exclude inconsistent subjects and reduce symmetrically the number of certainty payoffs decreasing the range of options offered, no significant differences⁴⁸ are found in the average rate of

48. All Bonferroni-corrected Wilcoxon test p-values are above 0.05.

adoption of the safe amount. However, the removal of these subjects cannot completely eliminate all framing effects generated by reducing asymmetrically the number of safe payoffs offered decreasing the range: we find that the previous differences obtained in the comparison of $LC^{R_{16}}$ and LC^{R_9} disappear, but the ones between $LC^{R_{10}}$ and LC^{R_9} still remain⁴⁹.

Result 7: *Removing inconsistent subjects, a symmetric decrease in the number of safe amounts reducing the range of options does not generate framing effects in the lottery vs. certainty method. Nevertheless, the elimination of these subjects reduces, but not completely deletes framing effects if the number of safe amounts is reduced asymmetrically, diminishing the range of options.*⁵⁰

Like in the previous method, we check in the lottery vs. certainty method whether our previous non-significant differences are due to a lack of statistical power. In this regard, we run an ex-post power analysis using power set at 90% and probability at 5%. For the lottery vs. certainty method, sample sizes would have to increase up to at least $N=353$ in order to find framing effects in the ones in which there are not.

49. Bonferroni-corrected Wilcoxon test p-values corresponding to 9€, 12€, 15€ and 18€ are 0.066, 0.018, 0.054 and 0.042 respectively after multiplying the original p-values by 6.

50. The same conclusion is reached when only the first decision in each pair is used instead of the mean of the six repetitions.

Our results nuance those of Blavatskyy and Koehler (2009) inferring that the range of feasible minimum safe amounts systematically affects elicited prices, and those of Bosch-Domènech and Silvestre (2013) concluding that the CE method is robust.

3.2. Regressions.

In this section, we estimate different logit models to shed light on the determinants of framing effects and to corroborate our previous results based on statistical tests.

3.2.1. Lottery vs Lottery method

Table 9 includes as explanatory variables the tasks subjects face in random order in each treatment and period, meaning all the tasks that modify our baseline lottery (LL₉) and the dependent variable is the number of safe choices. The modifications are based on changes in the number of options offered with and without decreasing the range. Additionally, we have two different models for each treatment: one includes the entire sample and the other includes only C&I subjects.

Table 9

LL models for the entire sample and for constant and imprecise subjects only. Standard errors in parentheses adjusted within-subjects; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

LL choices	T1	T1 (C&I)	T2	T2 (C&I)
LL ₁₃	0.0560 (0.0789)	0.0215 (0.104)		
LL ₁₇	-0.0473 (0.0794)	-0.0270 (0.104)		
LL ₅			-0.163 (0.114)	0.175 (0.140)
LL ₇			0.0851 (0.114)	0.263 (0.140)
Period	-0.0222 (0.0189)	-0.000924 (0.0248)	0.00823 (0.0274)	0.0117 (0.0333)
Constant	-0.846 (0.206)***	-0.762 (0.292)***	-1.015 (0.469)**	-1.426 (0.603)**
N	31	17	36	26

These models corroborate our Results 1, 2 and 3. Any modifications in the number of pairs offered with and without changing the range encompassing the options do not produce framing effects in binary lotteries. These results hold not only for C&I subjects, but also for the entire sample. Thus, this version of the multiple-choice list procedure is robust to framing effects.

3.2.2. Lottery vs Certainty method

The aim of this subsection is the same that in the previous one, but now for the lottery vs. certainty version.

Table 10 includes as explanatory variables the different tasks faced by subjects in random order in each different treatment (for safe and risky lotteries) and the period and the dependent variable is the number of safe choices. In other words, all the modifications made to our baseline lotteries (LC_{11}^S and LC_{16}^R) in the number of options offered with and without decreasing the range. Furthermore, we have two different models for each treatment: one includes the entire sample and the other includes only C&I subjects.

Table 10

LC models for the entire sample and for constant and imprecise subjects only. Standard errors in parentheses adjusted within-subjects; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

LC decision	T3 Safe	T3 Safe C&I	T3 Risky	T3 Risky C&I	T4 Safe	T4 Safe C&I	T4 Risky	T4 Risky C&I
LC_{16}^S	-0.123 (0.0644)*	-0.0559 (0.0927)						
LC_{21}^S	-0.0258 (0.0651)	0.0790 (0.0937)						
LC_{23}^R			-0.134 (0.0554)*	0.0143 (0.111)				
LC_{31}^R			-0.110 (0.0551)*	0.00703 (0.109)				
LC_5^S					-0.180 (0.172)	-0.366 (0.230)		
LC_8^S					-0.378 (0.168)**	-0.665 (0.222)**		
LC_{10}^R							-0.740 (0.0897)**	-0.815 (0.123)**
LC_9^R							-0.246 (0.0752)**	-0.266 (0.106)**
Period	-0.00800 (0.0156)	0.00571 (0.0223)	-9.81e-05 (0.0132)	-4.78e-05 (0.0262)	0.0974 (0.0401)*	0.0764 (0.0521)	0.00378 (0.0195)	0.000547 (0.0274)
Constant	0.706 (0.144)**	0.856 (0.203)**	-0.835 (0.127)**	-1.386 (0.191)**	3.429 (0.638)**	4.177 (0.687)**	0.620 (0.236)**	0.646* (0.351)*
N	36	18	36	11	34	26	32	16

From the previous models our Result 4, 5 and 7 are corroborated with this additional analysis. The only difference found is respect to our Result 6. The deletion of inconsistent subjects from the sample does not eliminate all the framing effects in the regression analysis. This fact is because in the statistical tests, the Bonferroni correction was applied and it is quite restrictive. Nevertheless, in general terms, the same conclusion is found: this version of the multiple-choice list procedure is not robust to framing effects. Modifications in the structure (number of options and range) of the LC used produce modifications in the risk attitude of subjects.

4. Conclusions

In this study, the robustness of two different choice list methods has been analysed: the LL method, where subjects face pairwise choices between gambles within a choice list, and the LC method, where subjects were asked to state a minimum safe amount to give up the lottery they have been endowed with. In order to analyse framing effects, we have implemented a within subjects experiment, allowing for preference imprecision and controlling for the proper work of the Random Lottery Incentive Mechanism, by means of the repetition of i.i.d. risk elicitation tasks. The framing effects analysed include shifts in risk preferences due to a (a)symmetric increase in the number of pairs offered keeping constant the range of options, and a (a)symmetric decrease in the number of pairs varying the range of options offered.

By means of a six times repetition of each identical risk task, we classify subjects depending on their variability within i.i.d. tasks in two categories: constant and inconstant subjects. The latter subjects were labelled as imprecise subjects or inconsistent subjects depending their variability self-reported cause. Inconsistent subjects violate the isolation hypothesis, driving a bad functioning of the RLI mechanism used to reward them in the experiment. For this reason, they are disregarded of the sample in part of the analysis. In that way, we can analyse if their presence in the sample is the fact that may drive framing effects.

In the LL elicitation method, we do not find framing effects when inconsistent subjects are removed from the sample. However, the LC method does not seem as robust as the LL method. Particularly, if only constant and imprecise subjects are considered, all framing effects found in the full sample analysis disappear with the exception of the ones when the range is diminished.

Summing up, some changes in risk preferences attributed to framing effects in the literature can really correspond to a malfunctioning of the RLI mechanism used to reward subjects in most experiments.

References

- Andersen, S., Harrison, G. W., Lau, M. I., & Elisabet, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4), 383–405.
- Andersson, O., Holm, H. J., Tyran, J. R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or Noise? *Journal of the European Economic Association*, 14(5), 1129-1154
- Ballinger, T. P., & Wilcox, N. T. (1997). Decisions, error and heterogeneity. *Economic Journal*, 107, 1090–1105.
- Bartling, B., Engl, F., & Weber, R. A. (2015). Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. *Journal of the Economic Science Association*, 1(1), 72–85.
- Beattie, J. & Loomes, G. (1997). The Impact of Incentives upon Risky Choice Experiments. *Journal of Risk and Uncertainty*, 14(2), 155-168.
- Beauchamp, J. P., Benjamin, D. J., Chabris, C. F., & Laibson, D. I. (2012). *How malleable are risk preferences and loss aversion?* Harvard University Mimeo.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226–232.

- Blavatskyy, P. R. & Kohler, W. R. (2009). Range effects and lottery pricing. *Experimental Economics*, 12(3), 332–349.
- Bosch-Domènech, A. & Silvestre, J. (2013). Measuring risk aversion with lists: a new bias. *Theory and Decision*, 75(4), 465-496.
- Brown, A. L., & Healy, P. J. (2018). Separated decisions. *European Economic Review*, 101, 20-34.
- Butler, D. J., & Loomes, G. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior and Organization*, 80(3), 511–522.
- Cason, T. N., & Plott, C. R. (2014). Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing. *Journal of Political Economy*, 122(6), 1235–1270.
- Cox, J. C., Sadiraj, K., & Schmidt, U. (2014). Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms. *PLOS ONE*, 9(3), e90742.
- Cox, J.C., Sadiraj V. & Schmidt U. (2015). Paradoxes and mechanisms for choice under risk, *Experimental Economics*, 18(2), 215-250.

- Cubitt, R. P., Navarro-Martínez D. & Starmer, C. (2015). On preference imprecision. *Journal of Risk and Uncertainty*, 50(1), 1-34.
- Cubitt, R. P., Starmer, C. & Sugden, R (1998). On the validity of the Random Lottery Incentive System. *Experimental Economics*, 1(2), 115-131.
- Erev, I., Glozman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty*, 36(2), 153–177.
- Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171-178.
- Freeman, D., Y. Halevy, & T. Kneeland (2016). Eliciting risk preferences using choice lists. Technical report, Vancouver School of Economics.
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38(1), 129–166.
- Harrison, G.W., Lau, M. I., Rutström, E. E., & Sullivan, M. B. (2005). Eliciting risk and time preferences using field experiments: Some methodological issues. In Carpenter, J., Harrison, G.W., & List, J.A., (Eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10).

- Harrison, G. W., Martínez-Correa, J. & Swarthout, J. Todd (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization*, 101, 128-140.
- Harrison, G. W., & Swarthout, J. T. (2014). Experimental payment protocols and the Bipolar Behaviorist. *Theory and Decision*, 77(3), 423–438.
- Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6), 1291–1326.
- Hey, J. D. (2001). Does repetition improve consistency? *Experimental economics*, 4(1), 5-54.
- Holt, C. A., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- Holt, C. A. (1986). Preference reversals and the independence axiom. *The American Economic Review*, 76(3), 508-515.
- Kahneman D & Tversky A (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 263–291.
- Lévy-Garboua, L., Maafi, H., Masclet, D., & Terracol, A. (2012). Risk aversion and framing effects. *Experimental Economics*, 15(1), 128–144.

Loomes, G. & Pogrebna G. (2014). Measuring individual risk attitudes when preferences are imprecise. *The Economic Journal*, 124(576), 569-593.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, 65(260), 581-598.

Starmer, C. & Sugden, R., (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review*, 81(4), pp. 971-78.

Tversky, A. & Kahneman D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211, 453-458.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, 251-278.

Appendix

Lottery vs. Lottery (LL) method

Table 11

Pairs of lotteries offered in LL₉

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV _S	EV _R	EV _S - EV _R
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
1	10%	17.5€	90%	26.70€	10%	0.0€	90%	100.0€	25.78€	90.0€	- 64.22€
3	20%	17.5€	80%	26.70€	20%	0.0€	80%	100.0€	24.86€	80.0€	- 55.14€
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€
15	80%	17.5€	20%	26.70€	80%	0.0€	20%	100.0€	19.34€	20.0€	- 0.66€
17	90%	17.5€	10%	26.70€	90%	0.0€	10%	100.0€	18.42€	10.0€	8.42€

Table 12Pairs of lotteries offered in LL₁₇

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV _S	EV _R	EV _S - EV _R
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
1	10%	17.5€	90%	26.70€	10%	0.0€	90%	100.0€	25.78€	90.0€	- 64.22€
2	15%	17.5€	85%	26.70€	15%	0.0€	85%	100.0€	25.32€	85.0€	- 59.68€
3	20%	17.5€	80%	26.70€	20%	0.0€	80%	100.0€	24.86€	80.0€	- 55.14€
4	25%	17.5€	75%	26.70€	25%	0.0€	75%	100.0€	24.40€	75.0€	- 50.60€
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
6	35%	17.5€	65%	26.70€	35%	0.0€	65%	100.0€	23.48€	65.0€	- 41.52€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
8	45%	17.5€	55%	26.70€	45%	0.0€	55%	100.0€	23.56€	55.0€	- 31.44€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
10	55%	17.5€	45%	26.70€	55%	0.0€	45%	100.0€	21.64€	45.0€	- 23.36€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
12	65%	17.5€	35%	26.70€	65%	0.0€	35%	100.0€	20.72€	35.0€	- 14.28€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€
14	75%	17.5€	25%	26.70€	75%	0.0€	25%	100.0€	19.70€	25.0€	- 5.30€
15	80%	17.5€	20%	26.70€	80%	0.0€	20%	100.0€	19.34€	20.0€	- 0.66€
16	85%	17.5€	15%	26.70€	85%	0.0€	15%	100.0€	18.88€	15.0€	3.88€
17	90%	17.5€	10%	26.70€	90%	0.0€	10%	100.0€	18.42€	10.0€	8.42€

Table 13

Pairs of lotteries offered in LL₁₃

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV _S	EV _R	EV _S - EV _R
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
1	10%	17.5€	90%	26.70€	10%	0.0€	90%	100.0€	25.78€	90.0€	- 64.22€
3	20%	17.5€	80%	26.70€	20%	0.0€	80%	100.0€	24.86€	80.0€	- 55.14€
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
10	55%	17.5€	45%	26.70€	55%	0.0€	45%	100.0€	21.64€	45.0€	- 23.36€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
12	65%	17.5€	35%	26.70€	65%	0.0€	35%	100.0€	20.72€	35.0€	- 14.28€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€
14	75%	17.5€	25%	26.70€	75%	0.0€	25%	100.0€	19.70€	25.0€	- 5.30€
15	80%	17.5€	20%	26.70€	80%	0.0€	20%	100.0€	19.34€	20.0€	- 0.66€
16	85%	17.5€	15%	26.70€	85%	0.0€	15%	100.0€	18.88€	15.0€	3.88€
17	90%	17.5€	10%	26.70€	90%	0.0€	10%	100.0€	18.42€	10.0€	8.42€

Table 14

Pairs of lotteries offered in LL₅

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV _S	EV _R	EV _S - EV _R
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€

Table 15

Pairs of lotteries offered in LL₇

Lottery Pair	Safe lottery (S)				Risky lottery (R)				EV _S	EV _R	EV _S - EV _R
	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff	Prob.	Payoff			
5	30%	17.5€	70%	26.70€	30%	0.0€	70%	100.0€	23.94€	70.0€	- 46.06€
7	40%	17.5€	60%	26.70€	40%	0.0€	60%	100.0€	23.02€	60.0€	- 36.98€
9	50%	17.5€	50%	26.70€	50%	0.0€	50%	100.0€	22.10€	50.0€	- 27.90€
11	60%	17.5€	40%	26.70€	60%	0.0€	40%	100.0€	21.18€	40.0€	- 18.82€
13	70%	17.5€	30%	26.70€	70%	0.0€	30%	100.0€	20.26€	30.0€	- 9.74€
15	80%	17.5€	20%	26.70€	80%	0.0€	20%	100.0€	19.34€	20.0€	- 0.66€
17	90%	17.5€	10%	26.70€	90%	0.0€	10%	100.0€	18.42€	10.0€	8.42€

Lottery vs. Certainty (LC) method

Safe lotteries:

Table 16

Pairs of options offered in LC^S₁₁

Pair		Safe Lottery			Safe amount (SA)		EV _S -SA
1	81%	10.0€	19%	0.00€	0.00€	8.10€	
3	81%	10.0€	19%	0.00€	1.00€	7.10€	
5	81%	10.0€	19%	0.00€	2.00€	6.10€	
7	81%	10.0€	19%	0.00€	3.00€	5.10€	
9	81%	10.0€	19%	0.00€	4.00€	4.10€	
11	81%	10.0€	19%	0.00€	5.00€	3.10€	
13	81%	10.0€	19%	0.00€	6.00€	2.10€	
15	81%	10.0€	19%	0.00€	7.00€	1.10€	
17	81%	10.0€	19%	0.00€	8.00€	0.10€	
19	81%	10.0€	19%	0.00€	9.00€	-0.90€	
21	81%	10.0€	19%	0.00€	10.00€	-1.90€	

Table 17

Pairs of options offered in LC^S₂₁

Pair		Safe Lottery			Safe amount (SA)	EV _S -SA
1	81%	10.0€	19%	0.00€	0.00€	8.10€
2	81%	10.0€	19%	0.00€	0.50€	7.60€
3	81%	10.0€	19%	0.00€	1.00€	7.10€
4	81%	10.0€	19%	0.00€	0.50€	6.60€
5	81%	10.0€	19%	0.00€	2.00€	6.10€
6	81%	10.0€	19%	0.00€	2.50€	5.60€
7	81%	10.0€	19%	0.00€	3.00€	5.10€
8	81%	10.0€	19%	0.00€	3.50€	4.60€
9	81%	10.0€	19%	0.00€	4.00€	4.10€
10	81%	10.0€	19%	0.00€	4.50€	3.60€
11	81%	10.0€	19%	0.00€	5.00€	3.10€
12	81%	10.0€	19%	0.00€	5.50€	2.60€
13	81%	10.0€	19%	0.00€	6.00€	2.10€
14	81%	10.0€	19%	0.00€	6.50€	1.60€
15	81%	10.0€	19%	0.00€	7.00€	1.10€
16	81%	10.0€	19%	0.00€	7.50€	0.60€
17	81%	10.0€	19%	0.00€	8.00€	0.10€
18	81%	10.0€	19%	0.00€	8.50€	-0.40€
19	81%	10.0€	19%	0.00€	9.00€	-0.90€
20	81%	10.0€	19%	0.00€	9.50€	-1.40€
21	81%	10.0€	19%	0.00€	10.00€	-1.90€

Table 18

Pairs of options offered in LC_{16}^S

Pair		Safe Lottery			Safe amount (SA)	EV_S-SA
1	81%	10.0€	19%	0.00€	0.00€	8.10€
2	81%	10.0€	19%	0.00€	0.50€	7.60€
3	81%	10.0€	19%	0.00€	1.00€	7.10€
4	81%	10.0€	19%	0.00€	0.50€	6.60€
5	81%	10.0€	19%	0.00€	2.00€	6.10€
6	81%	10.0€	19%	0.00€	2.50€	5.60€
7	81%	10.0€	19%	0.00€	3.00€	5.10€
8	81%	10.0€	19%	0.00€	3.50€	4.60€
9	81%	10.0€	19%	0.00€	4.00€	4.10€
10	81%	10.0€	19%	0.00€	4.50€	3.60€
11	81%	10.0€	19%	0.00€	5.00€	3.10€
13	81%	10.0€	19%	0.00€	6.00€	2.10€
15	81%	10.0€	19%	0.00€	7.00€	1.10€
17	81%	10.0€	19%	0.00€	8.00€	0.10€
19	81%	10.0€	19%	0.00€	9.00€	-0.90€
21	81%	10.0€	19%	0.00€	10.00€	-1.90€

Table 19

Pairs of options offered in LC^S_5

Pair		Safe Lottery			Safe amount (SA)		EV_S-SA
7	81%	10.0€	19%	0.00€	3.00€	5.10€	
9	81%	10.0€	19%	0.00€	4.00€	4.10€	
11	81%	10.0€	19%	0.00€	5.00€	3.10€	
13	81%	10.0€	19%	0.00€	6.00€	2.10€	
15	81%	10.0€	19%	0.00€	7.00€	1.10€	

Table 20

Pairs of options offered in LC^S_8

Pair		Safe Lottery			Safe amount (SA)		EV_S-SA
1	81%	10.0€	19%	0.00€	0.00€	8.10€	
3	81%	10.0€	19%	0.00€	1.00€	7.10€	
5	81%	10.0€	19%	0.00€	2.00€	6.10€	
7	81%	10.0€	19%	0.00€	3.00€	5.10€	
9	81%	10.0€	19%	0.00€	4.00€	4.10€	
11	81%	10.0€	19%	0.00€	5.00€	3.10€	
13	81%	10.0€	19%	0.00€	6.00€	2.10€	
15	81%	10.0€	19%	0.00€	7.00€	1.10€	

Risky lotteries:

Table 21

Pairs of options offered in LC_{16}^R

Pair		Risky Lottery			Safe amount (SA)	EV_R-SA
1	19%	45.0€	81%	0.00€	0.00€	8.55€
3	19%	45.0€	81%	0.00€	3.00€	5.55€
5	19%	45.0€	81%	0.00€	6.00€	2.55€
7	19%	45.0€	81%	0.00€	9.00€	-0.55€
9	19%	45.0€	81%	0.00€	12.00€	-3.55€
11	19%	45.0€	81%	0.00€	15.00€	-6.55€
13	19%	45.0€	81%	0.00€	18.00€	-9.55€
15	19%	45.0€	81%	0.00€	21.00€	-12.55€
17	19%	45.0€	81%	0.00€	24.00€	-15.55€
19	19%	45.0€	81%	0.00€	27.00€	-18.55€
21	19%	45.0€	81%	0.00€	30.00€	-21.55€
23	19%	45.0€	81%	0.00€	33.00€	-24.55€
25	19%	45.0€	81%	0.00€	36.00€	-27.55€
27	19%	45.0€	81%	0.00€	39.00€	-30.55€
29	19%	45.0€	81%	0.00€	42.00€	-33.55€
31	19%	45.0€	81%	0.00€	45.00€	-36.55€

Table 22Pairs of options offered in LC_{31}^R

Pair		Risky Lottery			Safe amount (SA)	EV_R-SA
1	19%	45.0€	81%	0.00€	0.00€	8.55€
2	19%	45.0€	81%	0.00€	1.50€	7.05€
3	19%	45.0€	81%	0.00€	3.00€	5.55€
4	19%	45.0€	81%	0.00€	4.50€	4.05€
5	19%	45.0€	81%	0.00€	6.00€	2.55€
6	19%	45.0€	81%	0.00€	7.50€	1.05€
7	19%	45.0€	81%	0.00€	9.00€	-0.55€
8	19%	45.0€	81%	0.00€	10.50€	-2.05€
9	19%	45.0€	81%	0.00€	12.00€	-3.55€
10	19%	45.0€	81%	0.00€	13.50€	-5.05€
11	19%	45.0€	81%	0.00€	15.00€	-6.55€
12	19%	45.0€	81%	0.00€	16.50€	-8.05€
13	19%	45.0€	81%	0.00€	18.00€	-9.55€
14	19%	45.0€	81%	0.00€	19.50€	-11.05€
15	19%	45.0€	81%	0.00€	21.00€	-12.55€
16	19%	45.0€	81%	0.00€	22.50€	-14.05€
17	19%	45.0€	81%	0.00€	24.00€	-15.55€
18	19%	45.0€	81%	0.00€	25.50€	-17.05€
19	19%	45.0€	81%	0.00€	27.00€	-18.55€
20	19%	45.0€	81%	0.00€	28.50€	-20.05€

21	19%	45.0€	81%	0.00€	30.00€	-21.55€
22	19%	45.0€	81%	0.00€	31.50€	-23.05€
23	19%	45.0€	81%	0.00€	33.00€	-24.55€
24	19%	45.0€	81%	0.00€	34.50€	-26.05€
25	19%	45.0€	81%	0.00€	36.00€	-27.55€
26	19%	45.0€	81%	0.00€	37.50€	-29.05€
27	19%	45.0€	81%	0.00€	39.00€	-30.55€
28	19%	45.0€	81%	0.00€	40.50€	-32.05€
29	19%	45.0€	81%	0.00€	42.00€	-33.55€
30	19%	45.0€	81%	0.00€	43.50€	-35.05€
31	19%	45.0€	81%	0.00€	45.00€	-36.55€

Table 23

Pairs of options offered in LC_{23}^R

Pair		Risky Lottery			Safe amount (SA)	EV_{R-SA}
1	19%	45.0€	81%	0.00€	0.00€	8.55€
2	19%	45.0€	81%	0.00€	1.50€	7.05€
3	19%	45.0€	81%	0.00€	3.00€	5.55€
4	19%	45.0€	81%	0.00€	4.50€	4.05€
5	19%	45.0€	81%	0.00€	6.00€	2.55€
6	19%	45.0€	81%	0.00€	7.50€	1.05€
7	19%	45.0€	81%	0.00€	9.00€	-0.55€
8	19%	45.0€	81%	0.00€	10.50€	-2.05€
9	19%	45.0€	81%	0.00€	12.00€	-3.55€
10	19%	45.0€	81%	0.00€	13.50€	-5.05€
11	19%	45.0€	81%	0.00€	15.00€	-6.55€
12	19%	45.0€	81%	0.00€	16.50€	-8.05€
13	19%	45.0€	81%	0.00€	18.00€	-9.55€
14	19%	45.0€	81%	0.00€	19.50€	-11.05€
15	19%	45.0€	81%	0.00€	21.00€	-12.55€
17	19%	45.0€	81%	0.00€	24.00€	-15.55€
19	19%	45.0€	81%	0.00€	27.00€	-18.55€
21	19%	45.0€	81%	0.00€	30.00€	-21.55€
23	19%	45.0€	81%	0.00€	33.00€	-24.55€
25	19%	45.0€	81%	0.00€	36.00€	-27.55€

27	19%	45.0€	81%	0.00€	39.00€	-30.55€
29	19%	45.0€	81%	0.00€	42.00€	-33.55€
31	19%	45.0€	81%	0.00€	45.00€	-36.55€

Table 24

Pairs of options offered in LC_{10}^R

Pair	Risky Lottery			Safe amount (SA)		EV_{R-SA}
7	19%	45.0€	81%	0.00€	9.00€	-0.55€
9	19%	45.0€	81%	0.00€	12.00€	-3.55€
11	19%	45.0€	81%	0.00€	15.00€	-6.55€
13	19%	45.0€	81%	0.00€	18.00€	-9.55€
15	19%	45.0€	81%	0.00€	21.00€	-12.55€
17	19%	45.0€	81%	0.00€	24.00€	-15.55€
19	19%	45.0€	81%	0.00€	27.00€	-18.55€
21	19%	45.0€	81%	0.00€	30.00€	-21.55€
23	19%	45.0€	81%	0.00€	33.00€	-24.55€
25	19%	45.0€	81%	0.00€	36.00€	-27.55€

Table 25

Pairs of options offered in LC^R_9

Pair		Risky Lottery			Safe amount (SA)	$EV_R - SA$
1	19%	45.0€	81%	0.00€	0.00€	8.55€
3	19%	45.0€	81%	0.00€	3.00€	5.55€
5	19%	45.0€	81%	0.00€	6.00€	2.55€
7	19%	45.0€	81%	0.00€	9.00€	-0.55€
9	19%	45.0€	81%	0.00€	12.00€	-3.55€
11	19%	45.0€	81%	0.00€	15.00€	-6.55€
13	19%	45.0€	81%	0.00€	18.00€	-9.55€
15	19%	45.0€	81%	0.00€	21.00€	-12.55€
17	19%	45.0€	81%	0.00€	24.00€	-15.55€

Chapter 3. Monetary incentives and overconfidence in academic performance: an experimental study

Abstract

In this paper, we analyse students' overconfidence bias on potential and actual academic performance under both hypothetical and real monetary incentives. Students enrolled in a Microeconomics course were offered the possibility to set their own goal before performing different types of exams and immediately after completing them, to postdict their own grade. Controlling for potential driving factors of students' overconfidence such as their cognitive abilities, academic record, risk preferences and self-reported academic confidence, we find that real monetary incentives mitigate overestimation of potential achievements and eliminate overestimation of actual achievements. This finding is compelling given the common interpretation of overconfidence as a conscious bias: if monetary incentives can eliminate subjects' overconfidence as our results point out, it might suggest that overconfidence is not a psychological bias at all.

1. Introduction

Overconfidence is a cognitive bias in which a person's subjective confidence in his/her performance is reliably greater than the objective result of this performance. Moore and Schatz (2017) distinguish three types of overconfidence: overprecision (tendency of individuals to be excessively certain about the accuracy of their beliefs), overestimation (tendency of subjects to exaggerate their absolute actual achievement or skill) and overplacement, which occurs when people perceive their performance in a group as better than it actually is. Previous research has shown that students consistently overestimate their performance on academic exams, especially when their grades are low. Specifically, Hacker *et al.* (2008) obtain that good students are usually more accurate with a tendency to underestimate and bad students usually overestimate their performance.

In this paper, we analyse two types of overestimation: students' overestimation of their potential academic performance (asking them to predict their own performance on a future exam) and students' overestimation of their actual academic performance (asking them to postdict their own performance on an exam immediately after they have completed it). In our paper, students' predictions can be recognised as an endogenous selection of their goals because responses were elicited well in advance of the examination date, allowing students to reconsider their own guesses until the day before the exam. Following Fryer and Elliot (2008), we consider that self-chosen

goals are empowered and proactive, creating commitment and acceptance. Moreover, although Bettinger *et al.* (2018) find evidence that students' beliefs in their ability to learn are malleable, van Lent and Souverijn (2016) show that students perform better when they set their own goal than in a case where mentors challenged them to set a higher goal.

Generally, the literature suggests that students' postdictions of performance are more accurately than predictions. This is because, "whereas predictions are made prospectively and are based on what students think they know, postdictions are made retrospectively and reflect the student's experience of the test" (Hacker *et al.*, 2008). Both subjects' predictions and postdictions, however are often elicited in the literature with non-incentive compatible methods. From Murstein (1965) on, multiple survey data confirming overconfidence has been generated by making no payments to motivate students in their task of forecasting. Exceptions using some kind of incentives are few in number. Feld *et al.* (2017) adopt a lottery draw in which students could win one of two gift vouchers worth €20 if their prediction four weeks before the exam was within 0.25 points of their grade. Their results provide evidence of overconfidence, especially for unskilled students. Blackwell *et al.* (2010) and Magnus and Peresetsky (2018) elicit different types of postdictions using extra grade points in order to encourage students to reveal their honest guesses. The former obtains that students, as a group, are quite adept at guessing their class average score on the assignment. On the contrary, the latter finds that students' expectations on own grades are not rational and

that most students are overconfident, in agreement with general literature. Lastly, Caplan *et al.* (2018) analyse whether incentives can mitigate students' overconfidence implementing an extra credit scheme to incentivise accurate grade forecasting. They find that the incentives tool generally fails to reduce students' overconfidence measured through the mean signed forecast error, as commonly used in the literature.

To the best of our knowledge, this paper is the first to study students' overconfidence using a piece-rate mechanism with real monetary incentives in order to elicit accurate grade forecasts. We consider money as a superior incentive because, as Croson (2005) point out "everyone values it, in contrast with extra-credit points or other grade-related rewards which may be valued only by students who are grade-conscious and/or whose grade may be affected by the outcome" and it is a non-satiabile good. Effectiveness of joint action of self-chosen goals and real monetary incentives in higher education is shown in Herranz-Zarzoso and Sabater-Grande (2018), as a departure point of the present study.

In addition to the type of incentives used to elicit guesses (predictions or postdictions), we consider three potential factors driving students' overconfidence: their skill (potential or actual), their risk aversion and their reported academic self-confidence.

First, we deal with subjects' skill. We use both students' potential and actual skill in order to test for the so-called "Dunning-Kruger effect". This well-known cognitive bias in social psychology holds that when someone is objectively unskilled in a given area, he/she widely overestimates their knowledge of it upon being asked to evaluate themselves. As Dunning (2011) states, this effect has been observed in multiple domains, including academic knowledge. Empirical evidence on the relation between skill and overconfidence is mixed depending on the type of overconfidence analysed and the methods used to measure both variables. In particular, when the same task is used to measure overconfidence and skill the empirical evidence obtained can be distorted by 'regression to the mean' (RTM) effects, that is, it is more likely that individuals with higher skill show less overconfidence. Avoiding this effect, we use different tasks to measure overconfidence (difference between students' guesses and actual grades) and skill (evaluated through a cognitive ability test or by means of the students' academic record depending on whether it is considered respectively as potential or actual).

Defining overconfidence as the difference between the average probability guess and the proportion of correct answers, a positive correlation is found in Wolfe and Grosch (1990) and Bruine de Bruin *et al.* (2007). On the contrary, Stanovich and West (1998) report a negative correlation of different cognitive ability tests with subjects' overconfidence. A more recent study by Hoppe and Kusterer (2011) finds that subjects with higher Cognitive Reflection Test (CRT) scores have a significantly more precise self-assessment of their

performance in a general knowledge questionnaire. However, there is no clear tendency that the more intuitive subjects are more overconfident than the analytical ones. This result may indicate that the effects of cognitive abilities reported in earlier studies may depend on the type of overconfidence analysed.

Duttle (2016) shows that, although overestimation (defined as the difference between a subject's estimated number of correct test items in the Raven Progressive Matrices (RPM) test and his/her actual performance in the test) is not affected by cognitive abilities (as measured by a CRT), the CRT score was associated with a significant decrease in overplacement and overprecision. Nevertheless, Bialek and Domurat (2018) show that the relationship between cognitive abilities and overconfidence disappears after addressing two critiques: (1) the CRT does not measure cognitive abilities but analytic cognitive style (test one's own intuitions), and (2) overconfidence and cognitive ability are artificially correlated since the RPM test, which served as a basis for estimating overconfidence, is also a measure of cognitive abilities.

Second, we focus on the relationship between overconfidence and risk aversion. Although in financial markets it is generally accepted that overconfident investors tend to underestimate risks, in experimental economics results are scarce and mixed. Blavatsky (2009) finds no significant relationship between risk attitudes, measured by means of the Holt and Laury (2002) method, (HL, hereafter), and overconfidence on own performance in a multiple-choice quiz. In the same vein, Michailova (2017) finds that risk

aversion, measured through the HL lottery-choice task, has no explanatory power over subjects' overconfidence on their accuracy in answering quiz questions. However, Murad *et al.* (2016) analyse the effect of risk attitudes on overconfidence based on own absolute performance through two different elicitation procedures: self-reported (non-incentivised) confidence and an incentivised procedure. The results about the former suggest that subjects are overconfident answering difficult questions and under confident in the easy ones. These results have been found in a large number of studies using non-incentivised self-reports. With the latter procedure, filtering out risk attitudes from inferred confidence reduces the degree of under-confidence.

Regarding a possible correlation between our two explanatory variables, cognitive ability and risk aversion, elicited by means of an incentivised method, some experiments suggest that they could be inversely related. However, these studies use either hypothetical choices or small-expected payoffs and Taylor (2013) shows that cognitive ability is unrelated to risk aversion when subjects are rewarded with real incentives. In the same vein, Andersson *et al.* (2016) report evidence that this relation may be spurious suggesting that cognitive ability is related to random decision making rather than to risk preferences.

Third, we examine the relationship between reported academic self-confidence and overconfidence. Following Sander and Sanders (2009), we use the Academic Behavioural Confidence (ABC, hereafter) as a global measure of perceived academic confidence, referring to cognitive judgements

concerned specifically, with study-related behaviours, like independent study's confidence, grades, discussions about course material and questions and academic engagement attending to classes. Using a group test of General Mental Ability by R.K. Tandon (1971) and a Self-Confidence Inventory by Agnihotri (1987), Dhall and Thukral (2009) investigate the relationship between intelligence, self-confidence and academic achievement in schools from Pakistan. They found that intelligence was positively correlated with both self-confidence and academic achievement. However, Saenz *et al.* (2017) find that there is not a strong or consistent relationship between students' grade predictions and other factors like attendance, study habits, preparation, and/or prior performance.

Attending the aforementioned potential driving factors, we are interested in studying whether students' overestimation bias can be reduced using real monetary incentives in order to elicit better guesses on own grades.

From now on, the paper is organised as follows: first, we introduce the experimental design and the measures used; after that, we analyse the empirical evidence collected and present our results; and lastly, we conclude.

2. Experimental design

A randomised field experiment was conducted to analyse students' overconfidence eliciting guesses on their own academic performance

through different (real or hypothetical) monetary incentives in a course of Microeconomics. Students' course evaluation included a non-binding midterm exam (NBME)⁵¹, a binding midterm exam (BME) and a binding final exam (BFE). This structure permits us to implement three calls, one for each exam. It is important to indicate that whereas both midterm exams were based on multiple-choice questions, the final exam consisted of short-answer and computational questions. Moreover, in order to explore the role of incentives in overconfidence, we used hypothetical monetary incentives in both midterm exams and both real and hypothetical ones in the BFE. Furthermore, this design allows us to look for overconfidence differences depending on the type of exam (multiple-choice test vs. non-multiple-choice test and binding vs. non-binding) and type of monetary incentives offered (hypothetical/real).

In each call, 322 students enrolled in a 2018 course of Microeconomics at the University Jaume I were offered the possibility of setting their own grade in two time intervals: before (*prediction*) and immediately after completing the exam (*postdiction*). In the first call, corresponding to the NBME, subjects (non-) expressing their willingness to participate were included in (G1) G2. In our second call, corresponding to the BME, students (non-) expressing their willingness to participate were included in (G3) G4. In the third call, corresponding to the BFE, subjects were informed that students responding

51. Students were allowed to waive the grade if it was not to their liking. In this case, the aforementioned grade would be replaced by the grade obtained in an additional question to be responded in the final exam.

affirmatively to the call would be randomly assigned⁵² to one of two groups: a non-monetarily incentivised group (G6) or to a monetarily incentivised group (G7). In consequence, whereas G6 participants would be incentivised with hypothetical money, participants assigned to G7 would be paid with real money according to their absolute (piece-rate) academic performance. After the publication of the definitive grades (two weeks after the BFE), participants were immediately paid in cash.

In addition, participants were notified about their corresponding group before they were invited to choose a goal for their final exam grade. In this way, this design discards possible willingness effects given that, both G6 and G7 are integrated only by students declaring their willingness to participate in the experiment. Furthermore, students not responding affirmatively to our third call were included in G5 in order to compare non-participants' final grades with the ones obtained by participants. Doing so, we can check for the actual existence of a (potential) willingness effect in our sample.

In (G2/G4/G6) G7 only one of the two elicited guesses, participants' prediction or postdiction, randomly chosen, was used to decide whether subjects would be paid or not with (hypothetical) real money. From here, we define students' potential overconfidence (POC) as the elicited prediction minus the grade obtained and students' actual overconfidence (AOC) as the difference between the postdiction and the grade obtained.

52. We randomise the treatments within each type (new and returning) of student.

Table 26 lists the summary of the treatments discussed above.

Table 26

Summary of treatments.

<i>Groups</i>	<i>Call</i>	<i>Number of subjects</i>	<i>Groups variables</i>				
			<i>Payment offered in the call</i>	<i>Willingness to participate</i>	<i>Randomization</i>	<i>Exam</i>	<i>Payment used</i>
G1	1	102	Hypothetical	No	No	Non-binding	-
G2	1	183	Hypothetical	Yes	No	Non-binding	Hypothetical
G3	2	55	Hypothetical	No	No	Binding	-
G4	2	247	Hypothetical	Yes	No	Binding	Hypothetical
G5	3	16	Hypothetical/Real	No	No	Binding	-
G6	3	43	Hypothetical/Real	Yes	Yes	Binding	Hypothetical
G7	3	60	Hypothetical/Real	Yes	Yes	Binding	Real

In each call, students were informed about the payment mechanism used to reward them depending on their goals and grades obtained in the exam. Like in Herranz-Zarzoso and Sabater-Grande (2018), we allowed participants to set their personal goal (as opposed to goals established by a professor).⁵³ In this way, the goal was tailored to each student's degree of self-control and ambition. This within-subject strategy aims to collect individual grades with hypothetical and real monetary incentives for the same subject in (non-) binding exams.

53. Unlike van Lent and Souverijn (2015), where a mentor-student meeting was used to induce students to set a course-specific grade goal, we choose to gather students' guesses in order to avoid any kind of bias in their outcome expectations.

In this experiment, the aim of participants was to maximise their monetary reward (R). They were informed that their reward would depend on their guess (G) and the obtained grade (GR) in the corresponding exam, according to the following function:

$$R = G^2 \forall GR \geq G$$

Given that the monetary payoff was predetermined by the size of the self-chosen goal, one might argue that risk aversion in combination with uncertainty about the expected performance might influence students in order to set their goals. In consequence, as in Goerg and Kube (2012) we measure subjects' risk aversion using HL lotteries.

Since we are interested in testing whether cognitive abilities promote better calibration of confidence, participants' cognitive ability is measured by means of the Abstract Reasoning part of the Differential Aptitude Test (DAT-AR for PCA, Bennett *et al.*, 1974).

Additionally, we elicit scores in the ABC scale⁵⁴. Nicholson *et al.* (2011) research suggests that undergraduates' confidence in their ability is related to academic performance. Specifically, this study shows that students who are confident at the beginning of the semester in their ability to attain high grades also perform better in their end-of-semester marks. Given the type of

54. See Sander and Sanders (2009).

course considered in this study, we are especially interested in two factors of the ABC scale: grades and studying beliefs.

Summing-up, controlling for potential driving factors such skill, risk aversion, and reported academic self-confidence, our experimental design aims to analyse whether two biases, the overconfidence phenomenon and the Dunning-Kruger cognitive bias, can be mitigated by using real monetary incentives to elicit students' guesses.

3. Measures of potential factors affecting students' overconfidence

In this section, we offer detailed information on the measures used to elicit students' risk aversion, cognitive abilities and self-reported academic confidence respectively.

1) Holt and Laury Lotteries (2002) is perhaps the most popular approach for measuring risk tolerance in the lab. Subjects face a table with 10 rows. In each row they are asked to make binary choices between a safe and a risky lottery in which the probability associated with the high payoff in both lotteries (safe or risky) increases and the probability associated with the low payoff decreases as subjects go down the table. It is expected that when the probability of the high payoff increases enough a subject switches from the

safe to the risky lottery in some point and never switches back. The switching point is the measure of risk aversion. Inconsistent preferences can be seen if a subject switches back or if he/she chooses the safe option in the tenth row given that it is stochastically dominated. In order to detect subjects with inconsistent preferences we allow subjects to switch back and forth.⁵⁵ When participants face these lotteries, they are informed that after making all decisions they will be paid just for one of them that will be selected at random. After that, subjects will be paid according to that decision.

2) Abstract Reasoning (AR) part of the Differential Aptitude Test for Personnel and Career Assessment. In this experiment, we use the AR part which is included in the DAT-5 Spanish adaptation by the publisher TEA (Cordero and Corral, 2000). This test contains 40 multiple-choice questions and subjects have 20 minutes to answer. The aforementioned questions are used as a non-verbal measure of abstract reasoning ability given that subjects think logically to find the relationships between abstract figure sequences. Colom *et al.* (2007) consider this test as a marker of fluid intelligence whilst McGrew, 2009 considers it the component of intelligence more liked to general intelligence or g factor.

3) Academic Behavioural Confidence scale (Sander and Sanders, 2009). The ABC scale used in this research was the 24-statement scale. These

55. We find a negligible (2%) percentage of inconsistent subjects in our sample. Moreover and contrary to Galarza and Bejarano (2016), we do not find a significant correlation between cognitive abilities and propensity to make inconsistent choices.

statements elicit the beliefs about achieving good grades in assessments (grades subscale), engaging in independent study (studying subscale), attending lectures, tutorials and other taught sessions (attendance subscale) and discussing material with lecturers and peers (verbalizing subscale). However, all analyses presented in this paper were computed only for two ABC subscales: grades and studying. In Sander *et al.* (2011), the ABC scale show cross-cultural validity when translated into Spanish and administered to over two thousand Spanish psychology students.

4. Data analysis

4.1. Sample self-selection: a potential willingness effect

Given that our design requires students' willingness to participate in the experiment in both the hypothetically incentivised group and the monetarily incentivised group, potential self-selection problems are ruled out. Notwithstanding, it is interesting to analyse the real existence of this potential effect comparing grades obtained in the midterm and final exams between students who are willing (participants) and who are not willing to participate in the experiment (non-participants).

Table 27

Average grades obtained in the exams for participants and non-participants NBME: Non-binding midterm exam; BME: Binding midterm exam; BFE: Binding final exam

	<i>Participants</i>	<i>Non-participants</i>	<i>Mann-Whitney Test p-value</i>
<i>NBME</i>	3.36	3.38	0.7899
<i>BME</i>	4.14	3.62	0.1134
<i>BFE</i>	4.75	4.30	0.1517

For each exam, table 27 displays both participants and non-participants average grades. In addition, the Mann-Whitney test p-values included comparing medians between participants and non-participants show that these differences are not statistically significant in any case. In consequence, we do not find a significant willingness effect in our sample.

4.2. Descriptive statistics and tests

In this section, we present descriptive statistics of students' predictions, postdictions, grades and overconfidence corresponding to each call of our experiment. Moreover, statistical tests comparing treatment variables are offered.

4.2.1. Hypothetical incentive calls

Table 28 displays descriptive statistics of the aforementioned variables for both midterm exams, where the incentives provided are exclusively hypothetical.

Table 28

Descriptive statistics of predictions, postdictions, grades and potential/actual overconfidence for the (non-) binding midterm exam.

	Prediction	Postdiction	Grade	POC	AOC
Non-Binding Midterm	5.94	5.15	3.66	2.29	1.51
	(1.34)	(1.55)	(2.02)	(2.62)	(2.38)
N	69	69	94	69	69
Binding Midterm	6.09	5.68	4.72	1.43	1.00
	(1.19)	(1.43)	(2.18)	(2.33)	(2.05)
N	97	97	100	97	97

Figure 17 displays previous means including Wilcoxon tests p-values showing within-subjects' differences between comparable variables. In the first row of figure 1, we show means of predictions, postdictions and grades for the NBME and BME respectively. In the second row, average POC and AOC for both midterm exams are presented.

The first row panels look like stairs, where the prediction is significantly higher than the postdiction and the latter significantly higher than the corresponding grade obtained. This pattern shows that although students adjust their guesses after the exam, they do not do it sufficiently, showing a significant AOC. However, as can be observed from the second row of panels, in both midterm exams the AOC is significantly lower than the POC.

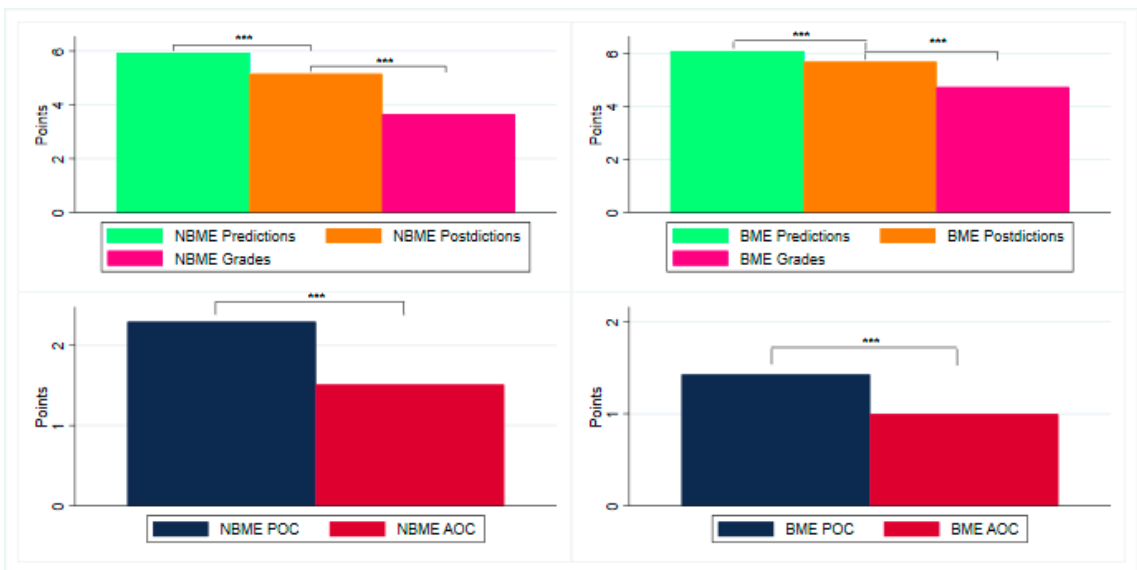


Fig. 17. Means of predictions, postdictions, grades, POC and AOC for the non-binding midterm and the binding midterm exams. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 18 displays the confidence measured using predictions and postdictions against the grade obtained in the midterm exams. The 45-degree line provides a benchmark given that points above the line would represent overconfidence whereas points below the line would represent under confidence. The dashed line is a grade reference point, meaning the minimum grade required to pass an exam.

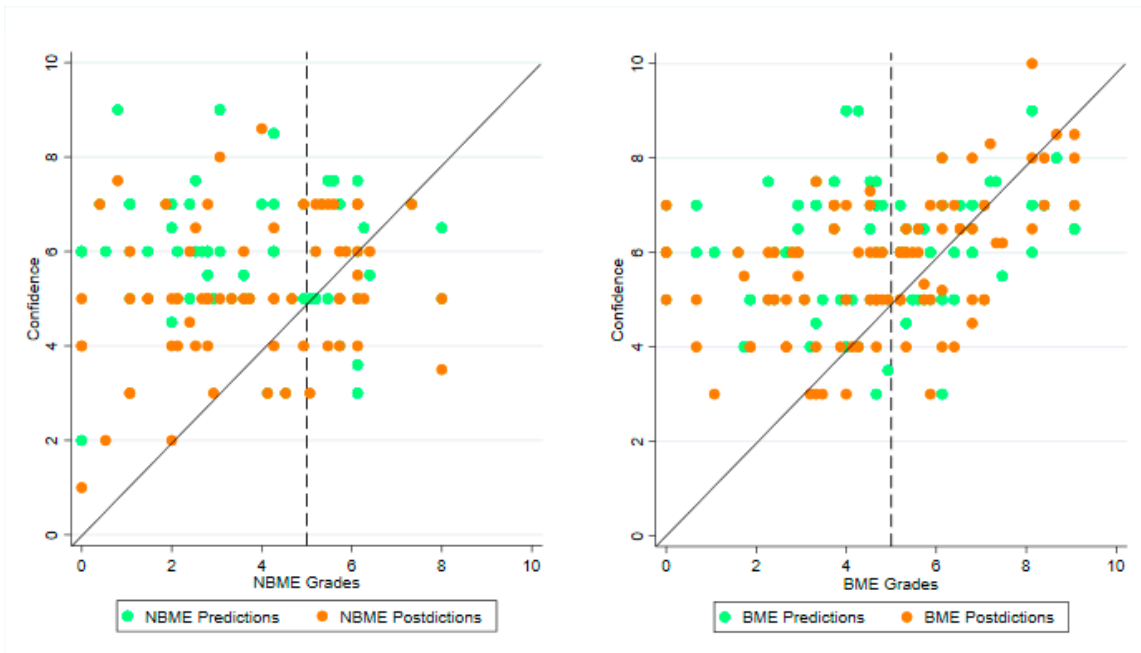


Fig. 18. Confidence using predictions and postdictions against grades in the midterm exams.

In these panels, we observe that most of the subjects show both POC and AOC. Unsurprisingly, under confidence is shown more often in students obtaining at least a pass in the exam, since it is more likely that individuals with higher grades show less overconfidence.

Furthermore, our experimental design allows us analyse how the exam binding consideration affects subjects' overconfidence when guesses are elicited with hypothetical incentives. Comparing overconfidence between the NBME and the BME exam, we find that both POC and AOC are significantly⁵⁶ higher when the exam is non-binding. In consequence, allowing students to treat an exam as if they had never taken it increases their overconfidence

56. Wilcoxon test p-values are 0.0000 for POC and 0.0008 for AOC.

Result 1: *Students exhibit a significant potential and actual overconfidence (being the former significantly higher than the latter) in their academic performance when guesses are elicited using hypothetical monetary incentives. These grade overestimations are even higher when students face a non-binding exam.*

4.2.2. Real monetary incentive calls

We focus now in the third call of our experiment, corresponding to the BFE. Splitting the sample between a non-monetarily incentivised (NMI) group and a monetarily incentivised (MI) group, table 29 presents descriptive statistics of: (1) subjects' predictions, (2) postdictions, (3) grades, (4) POC and AOC, (5) the number of safe choices per pair of options in the HL task, (6) score in the DAT-AR test, (7) scores in the grades and studying subscales of the ABC scale, and (8) academic record. Additionally, we present a t-test or a Mann-Whitney (M-W) test p-value comparing⁵⁷ the NMI group and MI group for all variables above-mentioned.

We observe that although predictions/postdictions in the BFE are similar in the NMI and the MI group, both types of overconfidence are significantly higher when subjects' guesses are elicited with hypothetical (NMI group) instead of real monetary incentives (MI group). Specifically, a POC of almost

57. We use t-tests for comparisons between normally distributed variables (BFE grade, POC and AOC, and DAT-AR score) and M-W test for comparisons between the remaining (non-normal) variables.

two points is reduced to a quarter when subjects' predictions are elicited using real money. In the same manner, an AOC of 1.3 points disappears when subjects' postdictions are obtained under real monetary incentives. This reduction effect caused in both POC and AOC is free of potential influences of factors like risk aversion, cognitive ability or academic confidence/record since these variables do not present significant differences between the NMI and the MI group.

Table 29

Descriptive statistics and treatment comparisons of predictions, postdictions, grades, overconfidence, safe choices in the HL task, DAT-AR score, subscales studying and grades of the ABC scale, and academic record. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	N	N	Mean NMI	Mean MI	S.D.	S.D.	M-W test/ T-test
	NMI	MI			NMI	MI	
BFE Prediction	42	58	6.01	5.83	0.98	0.99	0.4348
BFE Postdiction	42	58	5.34	5.32	1.58	1.23	0.7625
BFE Grade	42	58	4.04	5.26	2.23	2.41	0.0116**
BFE POC	42	58	1.97	0.58	2.16	2.40	0.0037***
BFE AOC	42	58	1.30	0.06	2.06	2.32	0.0072***
Safe Choices HL	42	58	4.52	4.57	1.86	2.09	0.7974
DAT-AR score	42	58	25.19	22.74	7.62	7.71	0.1186
Studying Confidence	42	58	14.31	15.10	2.87	2.16	0.1539
Grades Confidence	42	58	21.62	22.03	3.81	4.02	0.5320
Academic Record	42	58	6.40	6.40	0.66	0.49	0.4402

In Figure 19, above means are displayed and medians compared through a Wilcoxon test. Specifically, in the first row we display means of predictions, postdictions and grades corresponding to the BFE, split between NMI subjects and MI subjects.

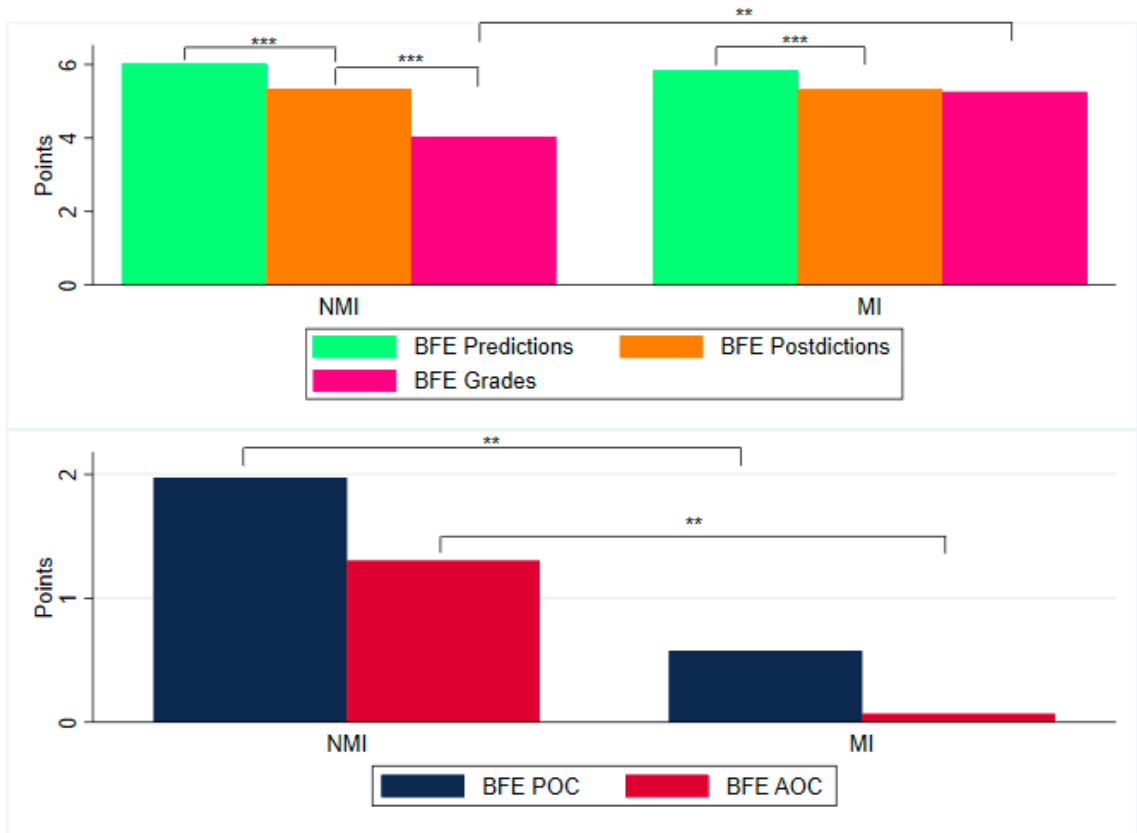


Fig. 19. Predictions, Postdictions, Grades, POC and AOC means for the final exam.
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

For both groups, we obtain that the median of students' predictions is significantly higher than the median of their postdictions. However, only NMI subjects obtain a median grade in the BFE significantly lower than their median postdiction, since MI subjects postdict their grade accurately. The

second row of figure 19 shows mean values of POC and AOC for both groups of subjects, showing that real monetary incentives significantly reduce both POC and AOC medians, being the latter near zero.

As in figure 18, figure 20 displays the confidence measured using pre and postdictions against the grade obtained in the BFE for subjects included in both the NMI and the MI group.

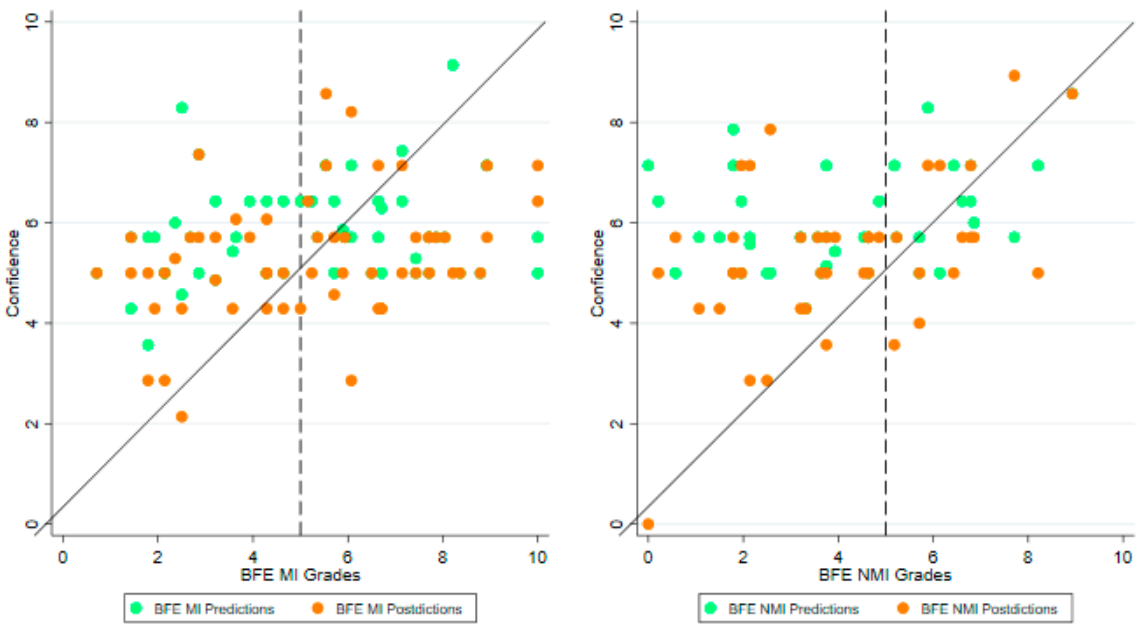


Fig. 20. Confidence using predictions and postdictions against grades in the final exam.

In the right panel of figure 20, we can observe that most of NMI subjects show both POC and AOC in line with the behaviour observed in

the (hypothetically incentivised) midterm exams.⁵⁸ Thus, this right panel reproduces the effect of overconfidence in general terms. Nevertheless, this pattern is not observed for MI subjects (left panel). In fact, they show overconfidence for low grades and under confidence for high grades, as can be expected when a regression to the mean effect is present.

Result 2: *The use of real monetary incentives to elicit students' guesses mitigate potential overconfidence and even eliminate actual overconfidence on their academic performance.*

We analyse now how the type of exam affects subjects' overconfidence when guesses are elicited with hypothetical incentives. If we compare overconfidence between the BME and the BFE for NMI subjects, we do not find statistically significant differences⁵⁹ for both POC and AOC.

Lastly, we examine the relationship between subjects' confidence and elicited variables like cognitive ability and risk aversion. Figure 21 displays both potential and actual confidence for subjects included in both the NMI and the MI group against the aforementioned variables. For both groups, NMI and MI subjects, the left panel shows no tendency relating confidence and cognitive ability.⁶⁰ Regarding risk taking from the right panel, we can observe that no

58. If we compare subjects' overconfidence between the BME and the BFE for NMI subjects, we do not find statistically significant differences for both POC and AOC (Wilcoxon test p-values are 0.9174 and 0.8256 respectively).

59. Wilcoxon test p-values are 0.9174 for POC and 0.8256 for AOC

60. Spearman's coefficients show a non-significant low correlation

relationship is found between confidence and risk taking.⁶¹

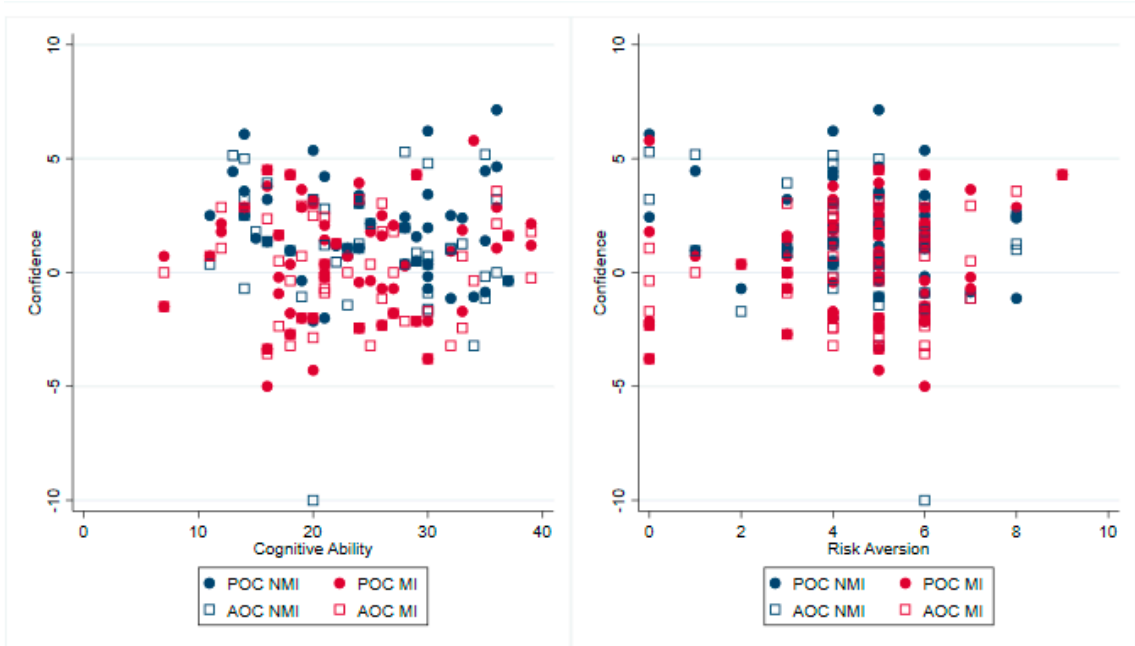


Fig. 21. Confidence using predictions and postdictions against grades in the final exam.

4.3. Regression analysis

4.3.1. Predictions and postdictions

In this section, potential driving factors for both predictions and postdictions in our three calls are analysed. In first place, we study the determinants of predictions and postdictions elicited through hypothetical incentives. With this aim, OLS models for midterm exams are estimated. The potential driving factors used are: (1) cognitive ability, (2) risk aversion, (3)

61. Spearman's coefficients show a non-significant moderate correlation.

academic record, (4) self-reported grades confidence and studying confidence and (5) gender (a dummy variable taking the value of 1 if the student is woman and 0 if men).

Table 30

OLS regressions for predictions and postdictions elicited in the midterm exams. Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	Predictions NBME	Postdictions NBME	Predictions BME	Postdictions BME
Cognitive Ability	0.0277 (0.0235)	0.0350 (0.0252)	0.0109 (0.0169)	0.0751*** (0.0187)
Risk aversion	-0.00881 (0.0969)	-0.0277 (0.104)	-0.0597 (0.0630)	-0.0766 (0.0713)
Academic Record	0.146 (0.312)	0.307 (0.335)	0.340 (0.230)	0.429* (0.253)
Grades Confidence	-0.0526 (0.0717)	0.0749 (0.0769)	0.0605 (0.0405)	0.00530 (0.0446)
Studying Confidence	0.0486 (0.0904)	-0.000721 (0.0969)	-0.0179 (0.0623)	0.0623 (0.0688)
Gender	0.154 (0.403)	-0.760* (0.432)	0.133 (0.273)	0.219 (0.304)
Constant	4.757** (2.307)	1.071 (2.473)	2.810* (1.548)	0.375 (1.702)
N	69	69	97	97
R-squared	0.030	0.160	0.082	0.217

The above models point out that when subjects' guesses are elicited under hypothetical incentives in a non-binding task, the potential driving factors do not have a statistically significant effect on the dependent variables. Regarding the BME, the variables affecting significantly students' postdictions are their cognitive ability and academic record. The first one has a positive effect, pointing out that subjects scoring high in the DAT-AR test postdict a

higher grade. The same relationship is found for students' academic record although the effect is weaker.

In second place, we analyse guesses generated in the third call of our experiment. OLS models are estimated in table 31 in order to explain BFE predictions/postdictions elicited using both hypothetical and real monetary incentives.

The first important result is related to cognitive ability. High cognitive ability subjects are motivated only when monetary incentives are real, inducing them to choose higher self-chosen goals. Moreover, cognitive ability does not play any role explaining postdictions independently of the type of incentives offered. In addition, subjects with better academic record only report higher predictions and postdictions when they are hypothetically incentivised. Regarding the importance of the type of incentives used to elicit students' guesses, we find that MI (a dummy variable indicating that subjects belong to the treatment group in which real monetary incentives are offered) is not significant: grades predicted and postdicted by MI subjects are, in average, similar to the corresponding ones by NMI subjects.

Table 31

OLS regressions for predictions and postdictions elicited in the final exam splitting the sample between monetarily incentivised and non-monetarily incentivised subjects. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

	Predictions all subjects	Predictions MI subjects	Predictions NMI subjects	Postdictions all subjects	Postdictions MI subjects	Postdictions NMI subjects
MI	-0.0775 (0.186)			-0.0707 (0.279)		
Cognitive Ability	0.0314** (0.0127)	0.0383** (0.0173)	0.0141 (0.0196)	-0.0105 (0.0191)	0.00537 (0.0219)	-0.0420 (0.0356)
Risk aversion	-0.0100 (0.0469)	-0.00854 (0.0626)	0.0101 (0.0809)	0.0340 (0.0703)	0.134* (0.0791)	-0.187 (0.147)
Academic Record	0.482*** (0.174)	0.220 (0.271)	0.727*** (0.227)	0.657** (0.261)	0.511 (0.342)	0.909** (0.414)
Grades Confidence	0.0314 (0.0304)	0.0238 (0.0398)	0.0389 (0.0529)	-0.00744 (0.0456)	-0.0375 (0.0503)	0.106 (0.0965)
Studying Confidence	-0.0310 (0.0473)	0.0316 (0.0752)	-0.0878 (0.0626)	0.0713 (0.0710)	0.160* (0.0950)	-0.0291 (0.114)
Gender	-0.187 (0.203)	-0.293 (0.280)	-0.0859 (0.304)	-0.390 (0.304)	-0.443 (0.353)	-0.497 (0.553)
Constant	2.095* (1.192)	2.727 (1.939)	1.412 (1.455)	0.609 (1.789)	-0.0702 (2.451)	-0.262 (2.651)
N	100	58	42	100	58	42
R-squared	0.234	0.185	0.380	0.114	0.148	0.203

4.3.2. Grades

In this section, determinants of students' grades for the three types of exams are analysed. Below we present OLS models explaining students' grades in midterm exams (where incentives offered are hypothetical) and in the BFE (where subjects randomly assigned to the MI group are incentivised with real money and those assigned to the NMI group are hypothetically incentivised).

Table 32

OLS regressions for grades obtained in the midterm and the final exam. The latter regression is split in monetarily incentivised and non-monetarily incentivised subjects. Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	NBME	BME	BFE	<i>BFE MI subjects</i>	<i>BFE NMI subjects</i>
MI			1.055*** (0.400)		
Cognitive Ability	-0.0153 (0.0266)	0.0745*** (0.0270)	0.0183 (0.0274)	-0.00796 (0.0369)	0.0665 (0.0413)
Risk aversion	-0.0406 (0.0994)	-0.0347 (0.102)	-0.0716 (0.101)	-0.125 (0.133)	0.0713 (0.171)
Academic Record	1.254*** (0.395)	1.648*** (0.369)	1.988*** (0.374)	2.390*** (0.577)	1.598*** (0.480)
Grades Confidence	-0.107 (0.0800)	-0.0656 (0.0645)	-0.0181 (0.0654)	-0.0270 (0.0847)	-0.0951 (0.112)
Studying Confidence	0.337*** (0.103)	0.190* (0.100)	0.307*** (0.102)	0.271* (0.160)	0.371*** (0.132)
Gender	-0.973** (0.422)	0.0725 (0.433)	-0.365 (0.436)	-0.974 (0.595)	0.664 (0.642)
Constant	-6.054** (2.676)	-8.851*** (2.507)	-13.74*** (2.563)	-12.35*** (4.128)	-11.69*** (3.076)
Observations	94	100	100	58	42
R-squared	0.254	0.283	0.403	0.374	0.465

The first three models point out that the students' academic record plays a crucial role forecasting grades obtained in any type of exam. The last two models indicate that this predictive power is not affected by the type of incentives offered. In addition, we obtain that subjects' confidence in their studying method affects positively to NBME and BFE grades, especially when subjects are NMI. Moreover, cognitive ability only affects positively the grade obtained in the BME. Furthermore, a gender effect is only found in the NBME pointing that women performed worse than men in this type of exam did. Regarding the model explaining BFE grades, MI is a crucial variable: MI subjects obtain, in average, 1.06 points more in the BFE grades than NMI subjects. Thus, real monetary incentives are effective improving significantly students' academic performance.

In order to reinforce our previous result highlighting the effectiveness of real monetary incentives increasing academic performance, we split the sample for each type of exam between NMI subjects and MI subjects attending our third call⁶². Figure 22 shows that ex-post MI students obtain lower grades than NMI students in both midterm exams, where monetary incentives are hypothetical. Nevertheless, when MI students (performing worst the midterm exams) are incentivised with real money they obtain significantly higher grades in the BFE than NMI students do. This fact evidences the importance of real monetary incentives in order to improve students' academic performance.

62. Although until our third call these groups were not created, we divide both midterm exam samples in ex-post treatment groups in order to test whether ex-post MI subjects obtained higher grades in the midterm exams than NMI subjects did.

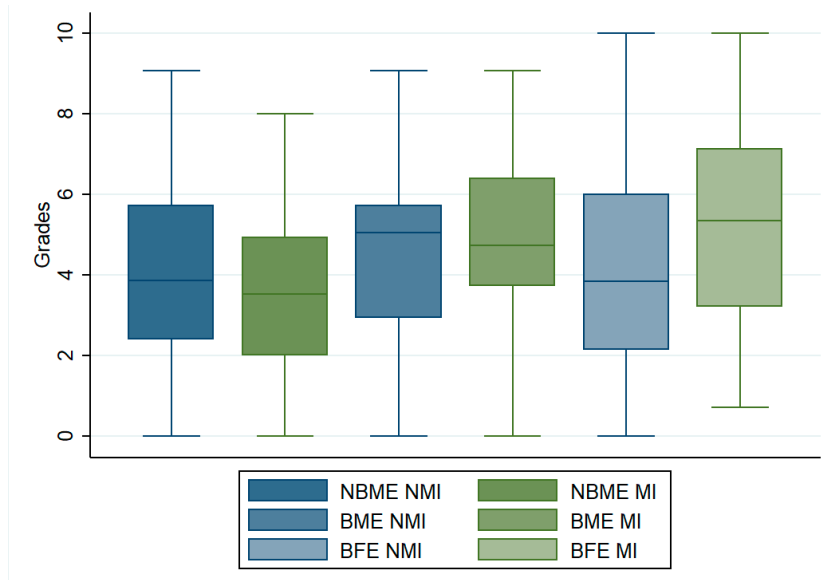


Fig. 22. Grades obtained in each exam for treatment group.

Finally, in order to determine the factors driving students' successful predictions/postdictions, we present two probit models in which the dependent variable is a dummy that takes the value of 1 if the student has exceeded his/her prediction/postdiction and 0 otherwise. From both models, we find that students' academic record affects positively to their success prospect: it is more likely that subjects with a better academic record exceed his/her prediction/postdiction than poor academic record students. More interesting is the role of the MI variable in the presented models: the fact of being incentivised with real money increases in 29.7% the probability to exceed the postdiction. For the case of predictions, the effect is weaker.

Table 33

Probit regressions for prediction success and postdiction success. Coefficients are marginal effects. Standard errors in parentheses
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	Prediction Success	Postdiction Success
MI	0.184* (0.101)	0.297*** (0.110)
Cognitive Ability	0.0076 (0.0069)	0.0093 (0.0749)
Risk aversion	-0.0085 (0.0253)	-0.0534* (0.0286)
Academic Record	0.254*** (0.094)	0.2434** (0.106)
Grades Confidence	0.0033 (0.0169)	0.0087 (0.0182)
Studying Confidence	0.0269 (0.0259)	0.0099 (0.0277)
Gender	0.0737 (0.311)	-0.0162 (0.118)
Observations	100	100

4.3.3. Overconfidence

In this section, four OLS models are estimated in order to shed light on the determinants of subjects' overconfidence. The explained variables in these models are POC and AOC for both midterm exams.

Table 34

OLS regressions for POC and AOC exhibited in the midterm exams. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

	POC NBME	AOC NBME	POC BME	AOC BME
Cognitive Ability	0.0374 (0.0431)	0.0447 (0.0390)	-0.0631** (0.0309)	-0.00213 (0.0281)
Risk aversion	-0.0352 (0.177)	-0.0540 (0.161)	-0.0900 (0.118)	-0.0930 (0.108)
Academic Record	-0.612 (0.572)	-0.452 (0.517)	-1.296*** (0.419)	-1.210*** (0.381)
Grades Confidence	0.0439 (0.131)	0.171 (0.119)	0.130* (0.0738)	0.0752 (0.0671)
Studying Confidence	-0.341** (0.166)	-0.390** (0.150)	-0.220* (0.114)	-0.144 (0.103)
Gender	1.439* (0.738)	0.525 (0.668)	0.304 (0.505)	0.309 (0.459)
Constant	9.105** (4.224)	5.419 (3.823)	11.90*** (2.818)	9.562*** (2.562)
N	69	69	95	95
R-squared	0.143	0.145	0.201	0.146

For the NBME, the only statistically significant variable is the studying confidence subscale, affecting negatively to POC and AOC: subjects with higher confidence in their studying methods have less overconfidence. For the BME, the variable affecting overconfidence is the academic record. Specifically, subjects with high university record show less overconfidence. This effect goes in line with cognitive ability, but only in the case of the POC.

Second, similarly to the case of midterm exams, we present OLS models based on subjects' overconfidence for the BFE. Additional models are reported in italics splitting the sample according to the type of incentives offered.

Table 35

OLS regressions for POC and AOC exhibited in the final exam splitting the sample in monetarily incentivised and non-monetarily incentivised subjects. Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	POC	POC MI subjects	POC NMI subjects	AOC	AOC MI subjects	AOC NMI subjects
MI	-1.122** (0.432)			-1.117** (0.427)		
Cognitive Ability	0.0117 (0.0299)	0.0440 (0.0406)	-0.0525 (0.0436)	-0.0302 (0.0296)	0.00974 (0.0395)	-0.109*** (0.0398)
Risk aversion	0.0703 (0.113)	0.131 (0.151)	-0.0609 (0.180)	0.114 (0.111)	0.282* (0.147)	-0.258 (0.165)
Academic Record	-1.508*** (0.403)	-2.173*** (0.622)	-0.873* (0.506)	-1.332*** (0.398)	-1.880*** (0.606)	-0.689 (0.463)
Grades Confidence	0.0513 (0.0708)	0.0547 (0.0923)	0.134 (0.118)	0.0124 (0.0699)	-0.00432 (0.0898)	0.202* (0.108)
Studying	-0.342*** (0.110)	-0.246 (0.174)	-0.459*** (0.139)	-0.239** (0.109)	-0.121 (0.169)	-0.400*** (0.127)
Gender	0.182 (0.471)	0.694 (0.644)	-0.752 (0.676)	-0.0204 (0.465)	0.553 (0.627)	-1.166* (0.618)
Constant	15.84*** (2.765)	15.08*** (4.454)	13.10*** (3.241)	14.35*** (2.732)	12.28*** (4.337)	11.42*** (2.963)
Observations	100	58	42	100	58	42
R-squared	0.306	0.280	0.362	0.261	0.271	0.418

In line with results 1 and 2, above regressions point out that, in order to estimate both POC and AOC, the type of incentives used (hypothetical or real) to elicit subjects' guesses, is very important. Specifically, MI subjects show a POC and AOC in average 1.13 points lower than overconfidence exhibited by students hypothetically incentivised. In conclusion, we can state that:

Result 3: *Controlling for potential driving factors, if we use real monetary incentives to elicit students' guesses, the overconfidence bias can be significantly reduced, even eliminated in the case that grades are calibrated after completing the exam.*

In addition, from previous OLS models we obtain that students' academic record affects negatively to both POC and AOC: subjects with better academic record show less overconfidence, especially when they are incentivised with real money. The same relationship is found regarding the studying subscale but only when incentives are hypothetical: subjects with more confidence in their studying methods show lower overconfidence. As opposite, cognitive ability, risk taking and gender do not play any role explaining POC and AOC.⁶³

Concerning empirical evidence on the existence of the Dunning-Kruger bias, we can summarise the following findings. Our results confirm this phenomenon using cognitive ability as the measure of potential skill only when incentives offered are hypothetical (students with higher scores in the DAT-AR test show potential overestimation in the BME and actual overestimation in the BFE). However, these effects disappear when incentives are real. Specifically, although students with higher cognitive abilities choose higher goals, we do not find any relationship between cognitive ability and potential overconfidence when predictions are elicited using real money. In addition,

63. With an exception: subjects with higher cognitive abilities show less AOC when they are hypothetically incentivised.

fluid intelligence of MI subjects is not related neither to their postdictions nor to their actual overconfidence.

In consequence, we can state that:

Result 4: *Using cognitive ability as subjects' measure of potential skill, for both potential and actual overconfidence, we do not find evidence of the Dunning-Kruger effect when students' guesses are elicited using real monetary incentives.*

Using students' academic record as measure of actual skill, we obtain that real monetary incentives reinforces the presence of the Dunning-Kruger bias, shown in a weaker way under hypothetical incentives in the BFE. Therefore, we conclude that:

Result 5: *Subjects with a better academic record obtain higher grades in all type of exams and show lower both potential and actual overconfidence when students' guesses are elicited using real monetary incentives.*

Combining the two previous results, we obtain the following corollary:

Corollary: *Using real monetary incentives to elicit students' guesses, we find that subjects with higher actual (potential) skills are (not) less overconfident than less actual (potential) skilled individuals.*

In the matter of the potential influence of subjects' risk preferences on their guesses and overconfidence, our empirical evidence shows that:

Result 6: *Subjects' risk aversion does not play a significant role in explaining how students choose their own goals or postdict their academic performance. Moreover, there is no relationship between risk aversion and overconfidence independently of the type of incentives used to elicit students' guesses.*

With regard to the influence of the self-reported academic confidence on behavioural data generated, we find that:

Result 7: *When guesses are elicited by means of hypothetical monetary incentives, although students reporting a higher confidence in their studying methods do not predict or postdict higher grades, they perform better in general, exhibiting a lower overconfidence. Moreover, their self-reported grades' confidence does not correlate with their guesses, grades or overconfidence in any case.*

5. Conclusions

Our paper is the first empirical research that introduces real monetary incentives and self-chosen academic goals in order to study students' overconfidence, meaning their tendency to overestimate their achievements. We have analysed the presence of two types of overestimation bias: overestimation of own potential performance and overestimation of own actual performance, depending on the type of monetary incentives (hypothetical or real) used to elicit students' guesses. In addition, students' potential and actual skills, risk preferences and self-reported academic confidence were measured in order to control for potential driving factors of subjects' overconfidence. Taking into account these individual characteristics, we obtain that students' overestimation of their potential achievements is significantly reduced when real money is used to elicit students' predictions. This effect is stronger in the case of subjects' actual overestimation of their grades, causing the bias elimination. Our findings suggest that overconfidence may not be a psychological bias given that it is generally conceived as a stable tendency that persist even in the presence of a subject's consciousness of it. From this point of view, incentives should not be able to affect subjects' overconfidence as we observe in our results.

However, this incentive effect is not as clear-cut in the case of the Dunning-Kruger cognitive bias. In fact, real monetary incentives enhance the presence of the Dunning-Kruger phenomenon when we use students'

academic record to rank them. Specifically, we find that, especially under real incentives, students with a better academic record overestimate less both their potential and actual academic performance in our Microeconomic course. On the contrary, using real money to elicit students' guesses weaken this cognitive bias when we adopt subjects' cognitive ability to measure them.

References

- Agnihotri, R. (1986). Manual for Agnihotri's self confidence inventory. (A. (ASCI) National Psychological Corporation, Ed.).
- Andersson, O., Holm, H. J., Tyran, J. R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or Noise? *Journal of the European Economic Association*, *14*(5), 1129–1154.
- Andrews, K., Swanson, J., & Kugler, P. (2007). Grade Expectations. *Journal of Economics & Economic Education Research*, *8*(2), 3–18.
- Armstrong, M. J. (2013). A Preliminary Study of Grade Forecasting by Students. *Decision Sciences Journal of Innovative Education*, *11*(2), 193–210.
- Bazerman, M. H., & Moore, D. A. (2009). Judgment in Managerial Decision Making. John Wiley (Vol. 44). Wiley.
- Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is “behavioral”? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, *11*(6), 1231–1255.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). Manual for the Differential Aptitude Tests. (Psychological Corporation, Ed.). New York.

- Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior and Organization*, 146, 1–15.
- Białek, M., & Domurat, A. (2018). Cognitive abilities, analytic cognitive style and overconfidence: a commentary on Duttte (2016). *Bulletin of Economic Research*, 70(1), 119–125.
- Blackwell, C. C. (2010). Rational expectations in the classroom: a learning activity. *Journal for Economic Educators*, 10(2), 1–6.
- Blavatsky, P. R. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty*, 38(1), 39–49.
- Bruine de Bruin, W. B., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, 92, 342–350.
- Burns, D. J. (2007). An examination of the accuracy of students' expected grades. *Academy of Educational Leadership Journal*, 11(3), 45–58.
- Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5(3), 182–191.

- Caplan, D., Mortenson, K. G., & Lester, M. (2018). Can incentives mitigate student overconfidence at grade forecasts? *Accounting Education, 27(1)*, 27–47.
- Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14(1)*, 77–90.
- Clark, D., Gill, D., Prowse, V., & Rush, M. (2017). Using Goals to Motivate College Students: Theory and Evidence from Field Experiments. National Bureau of Economic Research.
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4(1)*, 20–33.
- Colom, R., Escorial, S., Shih, P. C., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42(8)*, 1503–1514.
- Cordero, A., & Corral, S. (2006). DAT - 5 Tests de Aptitudes Diferenciales. (TEA Ediciones, Ed.). Madrid.
- Croson, R. (2005). The Method of Experimental Economics. *International Negotiation, 10(1)*, 131–148.

- Crosen, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 448–474.
- Dhall, S., & Thukral, P. (2009). Intelligence as Related to Self-confidence and Academic Achievement of School Students. *Journal of All India Association for Educational Research*, 21(2), 80–83.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review*, 100(3), 1238–1260.
- Dunning, D. (2011). The Dunning–Kruger Effect: On Being Ignorant of One’s Own Ignorance. *Advances in Experimental Social Psychology*, 44, 247–296.
- Duttle, K. (2016). Cognitive skills and confidence: interrelations with overestimation, overplacement and overprecision. *Bulletin of Economic Research*, 68(S1), 42–55.
- Falk, A., & Knell, M. (2004). Choosing the Joneses: Endogenous Goals and Reference Standards. *Scandinavian Journal of Economics*, 106(3), 417–435.
- Feld, J., Sauermann, J., & de Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral and Experimental Economics*, 68, 18–24.

- Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: the role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*(1), 1–19.
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25–42.
- Fryer, J. W., & Elliot, A. J. (2008). Self-regulation of achievement goal pursuit, motivation and self-regulated learning. In *Motivation and self-regulated learning: theory, research, and applications*. Lawrence Erlbaum Associates.
- Galarza, F., & Bejarano, H. (2016). Can cognitive skills and risk aversion explain inconsistent choices?: an experiment. Repositorio de la Universidad del Pacífico - UP. Universidad del Pacífico. Centro de Investigación.
- Glaser, M., Nöth, M., & Weber, M. (2004). Behavioral finance. In Wiley-Blackwell (Ed.), *Blackwell handbook of judgement and decision making* (pp. 527–546). Oxford.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*(1), 84–93.
- Goerg, S. J., & Kube, S. (2012). Goals (Th)at Work – Goals, Monetary Incentives, and Workers' Performance. *SSRN Electronic Journal*, 19.

- Grimes, P. W. (2002). The Overconfident Principles of Economics Student: An Examination of a Metacognitive Skill. *The Journal of Economic Education*, 33(1), 15–30.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3(2), 101–121.
- Herranz-Zarzoso, N., & Sabater-Grande, G. (2018). Monetary incentives and self-chosen goals in academic performance: *An experimental study*. *International Review of Economics Education*, 27, 34–44.
- Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644–1655.
- Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, 110(2), 97–100.
- Hossain, B. (2013). Can Business Students Forecast Their Own Grade? *American Journal Of Business Education*, 6(1), 85–92.
- Hossain, B., & Tsigaris, P. (2015). Are grade expectations rational? A classroom experiment. *Education Economics*, 23(2), 199–212.

- Jensen, P. A., & Moore, R. (2008). Students' Behaviors, Grades & Perceptions in an Introductory Biology Course. *The American Biology Teacher*, 70(8), 483–487.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York.
- Khachikian, C. S., Guillaume, D. W., & Pham, T. K. (2011). Changes in student effort and grade expectation in the course of a term. *European Journal of Engineering Education*, 36(6), 595–605.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1), 23–43.
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral Economics of Education: Progress and Possibilities. *Handbook of the Economics of Education*, 5, 1–74.
- Magnus, J. R., & Peresetsky, A. A. (2018). Grade Expectations: Rationality and Overconfidence. *Frontiers in Psychology*, 8, 2346.
- Maki, R. H. (1998). Test Predictions Over Text Material. In Erlbaum (Ed.), *Metacognition in Educational Theory and Practice* (pp. 117–144). Mahwah, NJ: Routledge.

- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.
- Meloy, M. G., Russo, J. E., & Miller, E. G. (2006). Monetary Incentives and Mood. *Journal of Marketing Research, 43*(2), 267–275.
- Michailova, J., Mačiulis, A., & Tvaronavičienė, M. (2017). Overconfidence, risk aversion and individual financial decisions in experimental asset markets. *Economic Research-Ekonomska Istraživanja, 30*(1), 1119–1131.
- Moore, D. A., & Healy, P. J. (2008). The Trouble With Overconfidence. *Psychological Review, 115*(2), 502–517.
- Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass, 11*(8), e12331.
- Murad, Z., Sefton, M., & Starmer, C. (2016). How do risk attitudes affect measured confidence? *Journal of Risk and Uncertainty, 52*(1), 21–46.
- Murstein, B. I. (1965). The Relationship of Grade Expectations and Grades Believed to be Deserved to Actual Grades Received. *The Journal of Experimental Education, 33*(4), 357–362.

Nicholson, L., Putwain, D., Connors, L., & Hornby-Atkinson, P. (2013). The key to successful achievement as an undergraduate student: confidence and realistic expectations? *Studies in Higher Education, 38*(2), 285–298.

Nowell, C., & Alston, R. M. (2007). I Thought I Got an A! Overconfidence Across the Economics Curriculum. *The Journal of Economic Education, 38*(2), 131–142.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization, 72*(1), 147–152.

Russo, J. E., & Schoemaker, P. H. (1992). Managing overconfidence. *Sloan Management Review, 33*, 7–17.

Saenz, G. D., Geraci, L., Miller, T. M., & Tirso, R. (2017). Metacognition in the classroom: The association between students' exam predictions and their desired grades. *Consciousness and Cognition, 51*, 125–139.

Sander, P., De La Fuente Arias, J., Stevenson, K., & Jones, T. (2011). A Validation of the Academic Behavioural Confidence Scale with Spanish Psychology Students. *Psychology Learning & Teaching, 10*(1), 11–24.

Sander, P., & Sanders, L. (2009). Measuring academic behavioural confidence: the ABC scale revisited. *Studies in Higher Education, 34*(1), 19–35.

- Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition, 44*(7), 1127–1137.
- Stanovich, K. E., & West, R. F. (1998). Individual Differences in Rational Thought. *Journal of Experimental Psychology: General, 127*(2), 161–188.
- Sturges, D., Maurer, T. W., Allen, D., Gatch, D. B., & Shankar, P. (2016). Academic performance in human anatomy and physiology classes: a 2-yr study of academic motivation and grade expectation. *Advances in Physiology Education, 40*(1), 26–31.
- Svanum, S., & Bigatti, S. (2006). Grade Expectations: Informed or Uninformed Optimism, or Both? *Teaching of Psychology, 33*(1), 14–18.
- Tandon, R. K. (1971). *Manual of Group Test of General Mental Ability*. (K.G.K. College, Ed.).
- van Lent, M., & Souverijn, M. (2017). *Goal Setting and Raising the Bar: A Field Experiment (No. 2017–001/VII)*. Amsterdam and Rotterdam: Tinbergen Institute.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572.

Wolfe, R. N., & Grosch, J. W. (1990). Personality Correlates of Confidence in One's Decisions. *Journal of Personality*, 58(3), 515–534.

General discussion and conclusions

This thesis is based on the experimental economics methodology and specially, in some issues of major importance: incentives, risk and cognitive biases like framing and overconfidence.

The first chapter is the first study which introduces self-chosen academic goals in a field experiment including two different payment mechanisms based on monetary incentives. Particularly, the effectiveness of these monetary incentives has been tested in absolute and relative academic performance through a piece-rate and a rank-order tournament system respectively. The novelty of this study is that subjects fixed their own self-chosen goals compatible with monetary incentives. The results of this experiment point out that monetary incentives in this setting of self-chosen goals are effective regardless of the payment mechanism used, piece rate or rank-order tournament. Specifically, monetary incentives are effective improving academic performance of students. Given that this is an issue of mayor importance in everyday life, the results of this investigation can be used by policymakers in order to increase the level of the education system. Although policymakers may see the implementation of an incentive's program difficult in terms of financial resources, the results obtained indicate that academic performance increases no matter if the payment mechanism used is a rank-order tournament (low-cost program) or a piece-rate (high-cost program).

After that, the second chapter analyses how two risk tasks highly used in the literature can be affected by a cognitive bias called framing. In particular, the tasks analysed are a lottery vs. lottery and a lottery vs. certainty methods. There is not a clear consensus in the literature about framing effects and for this reason, this study analyses this bias in two risk tasks but taking into account other potential factors that may influence the conclusions obtained. These factors are imprecision and the proper work the Random Lottery Incentive Mechanism (RLI). Specifically, the framing effects analysed include changes in the structure of the task based on modifications in the number and/or the range of options controlling for imprecision and the functioning of RLI. Regarding the latter, this study has observed that there are subjects who violate the main principle of this mechanism because they do not take each decision separately as it is expected and for this reason, this mechanism does not provide truthful incentives to them. After taking into account all the underlying factors, the results indicate that the lottery vs. lottery method is robust to changes in the structure of the task whereas the lottery vs. certainty method is not. These results entail a warning for experimental economists given that some framing effects existent in the literature can be due to a malfunctioning in RLI that is the mechanism mostly used in experiments.

The third chapter analyses another persistent bias documented in the literature, the overconfidence bias. This study is the first one which uses real monetary incentives and self-chosen goals in order to investigate the aforementioned bias. Specifically, subjects predict their grade before the exam

(potential performance) and postdict their grade once the exam has finished (actual performance) several times. In this setting, this bias is analysed under two types of incentives: hypothetical and real monetary incentives. The results obtained point that the use of real monetary incentives significantly reduces the overestimation bias of potential achievements and completely eliminates it in the case of actual overestimation.

Altogether, this thesis sheds new light on issues of high importance not only in the experimental economics field but also in the academic field. The results stress the importance of real monetary incentives. First, they are effective increasing academic performance which can be seen as a path to follow in order to improve the level of the education system. Second, some misleading results can be obtained in experiments when subjects do not fulfil the hypotheses that some payment mechanisms consistent with monetary incentives have. Last but not least, monetary incentives also reduce and eliminate biases highly documented in the literature..

