IMT School for Advanced Studies, Lucca
Lucca, Italy

Experimental Methods For Policy Analysis

PhD Program in System Science
Track in Economics Networks and Business Analytics
XXXVI Cycle

By

Niccolò Toccafondi

2025

The dissertation of Niccolò Toccafondi is approved.

PhD Program Coordinator: Massimo Riccaboni, IMT School for Advanced Studies Lucca

Advisor: Prof. Ennio Bilancini, IMT School For Advanced Studies Lucca

Co-Advisor: Prof. Sibilla Di Guida, IMT School For Advanced Studies Lucca

The dissertation of Niccolò Toccafondi has been reviewed by:

Trine Kjær, Professor at DaCHE - Danish Centre for Health Economics, Department of Public Health, University of Southern Denmark

Luca Polonio, Associate Professor, Depart. of Economics, Management and Statistics, University of Milano Bicocca

IMT School for Advanced Studies Lucca
2025

A Piero e Maria

Contents

List of Figures x

List of Tables xii

Acknowledgements xiv

Vita and Publications xvii

Abstract xx

1 Introduction 1

2 An online experiment for consumer protection policies 7

2.1 Introduction 7

2.2 Literature review 10

2.2.1 Virtual Currencies 10

2.2.2 Money Illusion 12

2.2.3 Non-Intuitive Exchange Rates 14

2.3 Experimental Design 15

2.3.1 The Decision Tasks phase 15

2.3.2 WTP Estimation 19

2.3.3 Experimental Procedure 19

2.4 Results 22

2.4.1 Descriptive and Statistical Analyses 23

2.4.2 Regression analysis 29

2.5 Discussion and conclusions 30

2.5.1 The Virtual Currency effect 32
2.5.2 The Money Illusion Effect 33
2.5.3 The non-intuitive exchange rates effect 34
2.5.4 Final remarks 35

3 An online experiment to reduce inappropriate demand of urgent healthcare services 36
3.1 Introduction 36
3.2 Literature review 40
3.2.1 Choosing for others 42
3.3 The model 43
3.3.1 Payoff Assumptions 46
3.4 Experimental Design 47
3.4.1 The task 50
3.4.2 Additional Questions 52
3.4.3 Hypothesis 54
3.4.4 Experimental Procedure 56
3.5 Results 60
3.5.1 Descriptive analysis 61
3.5.2 Result 1: Policy and Suggestion Effects 63
3.5.3 Result 2: More ER visits in the *for-others* task 69
3.5.4 Result 3: Preference for TH 70
3.6 Discussion 74
3.7 Conclusion 77

4 A field experiment for education policies 79
4.1 Introduction 79
4.2 CyberTrials 2023 and The Big Game 81
4.2.1 Lectures 82
4.2.2 Training Modules 83
4.3 Experimental Design 85
4.3.1 Online Surveys 86
4.3.2 Experimental procedure 87
4.4 Results 91
4.4.1 Econometric analysis 91

4.5 Discussion 94

5 Conclusion **98**

A Supplementary Material for Chapter 2 **101**
A.1 Supplementary Information 101
A.2 Sample Balance Tests 102
A.3 Shapiro – Wilk W test for normality and Levene's homogeneity of variance tests 102
A.4 Framing effect and Reference effect for ambiguous lotteries 103
A.5 Ambiguity Aversion 105
A.6 Robustness Check for Hypothesis 2 106
A.7 Additional Results 106
A.8 Additional tables 111
A.9 Additional Figure 112

B Supplementary Material for Chapter 3 **115**
B.1 Supplementary Information 115
B.2 Expected number of visits to the ER 116
B.3 Additional Questions - Results 117
B.4 Robustness check 119
B.5 Addtional Tables 120
B.6 Additional figures 127

C Supplementary Material for Chapter 4 **129**
C.1 Supplementary Information 129

List of Figures

1 Summary of this study structure. 16
2 Task's interface in condition T1 and lottery R6 18
3 Average WTP across conditions and lotteries in risky and ambiguous lotteries and 95% confidence intervals. 24
4 Average difference between WTP across conditions and lotteries and 95% confidence intervals. 27

5 Summary of the study structure in this study. 49
6 Payoff matrix shown to participants. 51
7 Average ER visits by condition, posterior and congestion. . 62
8 Average $ER_diff_{\mu,\omega}$ by condition t. 71

9 Self-reported measures of participation, interest, and perceiveddifficulty in lectures and training modules. 90
10 Average number of correct answers for each question and corresponding topic. 92
11 Marginal effects of *Participation to the lectures* and *Participation to the Activities* per question. 96

A.1 Average WTP across conditions and lotteries in risky and ambiguous lotteries. 112
A.2 Attention Checks 113
A.3 Additional Questions 114

B.1 Statistics on AQ1 and AQ2. 118

B.2 Answers to questions AQ3-AQ6. 119
B.3 Answers to AQ7. 120
B.4 Answers to AQ8. 120
B.5 Country of residence of the participants. 127
B.6 Condition TC 128
B.7 Condition TP 128
B.8 Condition TS 128

List of Tables

1	Risky and ambiguous lotteries in this study	17
2	Descriptive statistics for this study.	21
3	Summary statistics for the dummy variable Correct Conversion (Panel A). Test for less intuitive exchange rates (Panel B).	23
4	Cohen's d Effect Sizes and 95% CI	28
5	Random effect estimates of the average WTP	31
6	Marginal estimates across treatments for models in Table 5	31
7	Posteriors probabilities	52
8	Additional Questions	53
9	Symptoms, Clinical Profiles, and Severity	54
10	Summary Statistics.	59
11	Sample balance tests	60
12	Linear regressions on $ER_self_{i,t,\mu,\omega}$.	65
13	Linear regressions on $ER_self_{i,\mu,\omega}$ with baseline TP.	68
14	Wilcoxon signed rank test on $ER_diff_{\mu,\omega}$ by condition t.	70
15	Linear regressions on $ER_self_{i,t,\mu}$ for $t = \{TP, TS\}$ and $\omega = L$.	73
16	Linear regressions on $ER_self_{i,t,\mu}$ for $t = \{TP, TS\}$ and $\omega = L$.	73
17	Skills and Relative Descriptions	82

18 Summary Statistics for female students in the treatment group and in the second survey. 89
19 Poisson regressions on the number of correct answers. . . . 95

A.1 Sample balance tests. We report the p-values for the Kruskal-Wallis equality-of-population rank test and the p-values for each possible pairwise Wilcoxon Rank-Sum test. 103
A.2 Panel a) Shapiro – Wilk W test for normality. Panel b) Levene's homogeneity of variance tests. 104
A.3 Cohen's d Effect Sizes and 95% CI 108
A.4 Cohen's d Effect Sizes and 95% CI 109
A.5 Cohen's d Effect Sizes and 95% CI 110
A.6 Non-parametric Effect Sizes 111

B.1 Posterior probabilities μ which make a risk neutral agent indifferent between ER and TH 116
B.2 Posterior probabilities μ which make a risk averse agent with utility function $u(x) = \sqrt{x}$ indifferent between ER and TH. 117
B.3 Median probability per class of posterior and relative standard errors 121
B.4 Posterior Balance test 121
B.5 Heterogeneity Analysis. Linear regressions on $ER_self_{i,t,\mu,\omega}$ without covariates. 122
B.6 Heterogeneity Analysis. Linear regressions on $ER_self_{i,t,\mu,\omega}$ with covariates. 123
B.7 Linear regressions on $ER_self_{i,t,\mu,\omega}$ for intermediate levels of congestion. 124
B.8 Heterogeneity Analysis. Linear regression on $ER_self_{i,t,\mu,\omega}$ with TP as the baseline and TS as the regressor without covariates 125
B.9 Heterogeneity Analysis. Linear regression on $ER_self_{i,t,\mu,\omega}$ with TP as the baseline and TS as the regressor with covariates 126

Acknowledgements

I would like to thank my advisors, Ennio Bilancini and Sibilla Di Guida. Thanks to you, I could pursue a path that suited my aspirations and my needs. Your support and guidance have brought me great satisfaction.

I also sincerely thank Trine Kaer and Luca Polonio, for reviewing my doctoral thesis with speed and accuracy, and my colleagues Andrea Albertazzi, Leonardo Boncinelli, Tatiana Celadin, Roberto Di Paolo, Marzio Di Vece, Alice Dominici, Chiara Nardi, Erica Ordali, Folco Panizza, Massimo Riccaboni, and Francesco Serti for their support throughout these years.

I am truly grateful to Dorte Gyrd-Hansen, Benoît Carré, Rebecca Jones, Lucy Poole, Emily Johnson, and all the colleagues at DACHE for making me feel welcomed in their research centre and making my stay in Odense a wonderful (some might say whimsical) experience. I will not forget all the support you have given me and all the Monday morning breakfasts.

A heartfelt thank you goes to all the members of the BEE group, especially Bianca, Lina, Veronica, Ginevra, and Pablo. I could always count on you for honest and accurate feedback.

I also want to thank all the friends I met at IMT, Agnese, Alessio, Caterina, Fabio, Francesca, Giorgos, Lorenzo, Marco, Nick, Roberto, Silvia e Vittorio. I'm grateful for the many great moments with you, to have had (countless by now) breakfasts at Pecora Nera, and to have been able to share successes and failures with you. I also want to thank all my

friends with whom I share my passions. Those who know me know how important they are to me.

Finally, a special thank you goes to my family. To Alberto, Errica, and Giulia, you have always been by my side. To Irene, your love and kindness are what bring me joy.

-

Ringrazio i miei relatori Ennio Bilancini e Sibilla Di Guida per avermi dato la possibilità di seguire un percorso che fosse adatto alle mie aspirazioni e alle mie necessità. Il vostro sostegno e il vostro consiglio mi hanno portato preziose soddisfazioni.

Dedico poi sinceri ringraziamenti a Trine Kaer e Luca Polonio, per aver svolto il compito di revisori della tesi con rapidità e precisione, e ai miei colleghi Andrea Albertazzi, Leonardo Boncinelli, Tatiana Celadin, Roberto Di Paolo, Marzio Di Vece, Alice Dominici, Chiara Nardi, Folco Panizza, Erica Ordali, Massimo Riccaboni e Francesco Serti, per avermi aiutato a crescere professionalmente in questi anni.

Voglio inoltre esprimere la mia gratitudine per Dorte Gyrd-Hansen, Benoît Carré, Rebecca Jones, Lucy Poole, Emily Johnson e tutti i colleghi di DACHE. Mi avete accolto nel vostro centro di ricerca e mi avete offerto tante preziose opportunità. Non scorderò le colazioni del lunedì mattina.

Un ringraziamento speciale va a tutti i membri del gruppo BEE e in particolare a Bianca, Lina, Veronica, Ginevra e Pablo. Siete stati un punto di riferimento e un gruppo su cui fare affidamento nei momenti di incertezza.

Voglio poi ringraziare i tanti amici conosciuti a IMT, Agnese, Alessio, Caterina, Fabio, Francesca, Giorgos, Lorenzo, Marco, Nick, Roberto, Silvia e Vittorio. Sono felice di aver passato

tanti bei momenti insieme a voi, di aver fatto (ormai innumerevoli) colazioni alla Pecora Nera e di aver potuto condividere con voi successi e fallimenti. Voglio inoltre ringraziare tutti i miei amici, vecchi e nuovi, con cui riesco a coltivare le mie passioni. Chi mi conosce sa quanto siano importanti per me.

Infine un ringraziamento speciale va alla mia famiglia. Ad Alberto, Errica e Giulia, mi siete stati vicini in ogni occasione. Ad Irene, il tuo amore e la tua gentilezza sono ciò che mi fa stare bene.

Vita

June 27th, 1995	Born, Prato (PO) Italy
2020-	PhD Student in Economics, Networks and Business Analytics IMT School For Advanced Studies Lucca Lucca, Italia
2024	Visiting Ph.D. Student Danish Center of Health Economics (DACHE), University of Southern Denmark (SDU) Odense, Danmark
2020	M.Sc. Degree in Economics Final Mark: 110/110 magna cum laude Università degli studi di Firenze Firenze, Italia
2018	B.Sc. Degree in Economia e Commercio Final Mark: 107/110 Università degli studi di Firenze Firenze, Italia

Work Experiences

2023-2024	Health Economics Internship Fondazione Toscana Life Sciences (TLS) Siena, Italia

Publications

Publications

1. N. Toccafondi, E. Bilancini, L. Boncinelli, “Learning CyberSecurity with Story-Driven CTF Challenges: CyberTrials 2023,” in *Communications in Computer and Information Science*, vol. 2076, 2024, https://doi.org/10.1007/978-3-031-67351-1_21.

Working papers

1. N. Toccafondi, R. Di Paolo, S. Di Guida, "Virtual currencies in online gaming increase the willingness to pay for loot boxes: an experimental analysis", Status: *Submitted to Experimental Economics*.
2. N.Toccafondi, V. Pizziol, E. Santin, T. Antognozzi, R. Di Paolo, E. Bilancini “Delocalized events management: The experience of hobby stores hosting the Campfire event during LCG20.”

Work in progress

1. S. Di Guida, D. Gyrd-Hansen, E. Johnson, M. Riccaboni, M.Skovrider Aaskoven., N. Toccafondi, "The social effect of Nivolumab treatment as adjuvant therapy for resected melanoma stage 3 or higher: a retrospective study".
2. S. DiGuida, B. Sanesi, N. Toccafondi, "The impact of real-time congestion information on ER overcrowding: A theoretical and experimental study".
3. M. Di Vece, F. Serti, N. Toccafondi, "The effect of COVID-19 on blood donation of habitual donors: A Causal AI Investigation".

Presentations

1. N. Toccafondi, "What Drives Primary School Students to Play Fortnite?", Digital Games Research Association (DIGRA) Annual Conference, Krakow, Poland, 2022

2. N. Toccafondi, “The effect of Virtual Currencies and Money Illusion on the WTP for online transactions involving risk,” at Ninth Meeting of the Behavioral and Experimental Economics Network, University of Venezia, Venezia, Italia, 2023.
3. N.Toccafondi, "Learning Cybersecurity with Games: CyberTrials 2023", at 5th International Conference on Higher Education Learning Methodologies and Technologies Online, University of Foggia, 13-15 September 2023
4. N. Toccafondi, "Virtual currencies in online gaming increase the willingness to pay for loot boxes: an experimental analysis", at Young Economists of Tuscan Institutions (YETI) Meeting, 25^{th} January 2024
5. N. Toccafondi, "The impact of real-time congestion information on ER overcrowding: A theoretical and experimental study", at Thirteenth Meeting of the Behavioral and Experimental Economics Network, University of Bologna, Bologna, Italia, 2025.

Abstract

Traditional policy analysis typically address policy relevant questions using observational data and quasi experimental methodologies. There are cases, however, when observational data are either not available to researchers, not fit for causal interpretation, or non-existent. In these cases experimental methods provide a valid alternative to support the policy-making process in the design, implementation and evaluation phases. This dissertation contributes to the growing field of experimental policy analysis by examining three policy-relevant questions where observational data are private and not available to researchers (Chapter 2), are affected by a censoring bias and are not fit for causal interpretation (Chapter 3) and are non-existent (Chapter 4). In Chapter 2, I find that virtual currencies increase participants' willingness to pay for transactions involving risk and ambiguity, even if the exchange rate remains constant. Moreover, when the nominal value of a virtual currency is higher than its real value, participants reduce their willingness to pay and become more sensitive to risk and ambiguity. In Chapter 3, I examine the effects of providing real-time information on ER congestion. I find an increase of the total number of ER visits. This evidence suggests that when information is available the uncertainty of ER visits decreases, leading risk-averse individuals to choose the ER more frequently. In Chapter 4, I provide evidence supporting the importance of complementing theory-driven lectures with hands-on, gamified activities for computer science and cybersecurity education.

Chapter 1

Introduction

Traditional policy analysis typically address policy relevant questions using observational data and quasi experimental methodologies. The increasing availability of individual level data created the opportunity for researchers to adopt a multidisciplinary approach integrating data from different fields. For example, recent policy works examine the effect of innovative healthcare therapies on economic outcomes such as labour force participation and income (Daysal et al., 2024; Aaskoven, Kjær, and Gyrd-Hansen, 2022). Moreover, it created the opportunity for the development of new methodologies for causal inference (Callaway and Sant'Anna, 2021; Brand, Zhou, and Xie, 2023), available to researchers to make accurate predictions and robust causal analysis. There are cases, however, when observational data are either not available to researchers, not fit for causal interpretation, or simply non-existent. This can happen when data quality is scarce, incomplete or affected by measurement errors (Liu and Demosthenes, 2022). Moreover, ethical and legal constraints may limit data availability. Medical and labour records, for example, are protected by strict GDPR regulations to prevent privacy violations. Finally, results from traditional policy analysis can be context-specific, and causal effects may not be replicable or generalizable (Athey and Imbens, 2017).

In these cases, laboratory, field and online experiments, provide con-

trolled environments for testing economic models and causal relationships. For example, when studying models of strategic interactions between agents, researchers can manipulate group size, type of interactions and games' payoff structure (see Fehr and Gächter (2000) and Catola et al. (2023)). In studies that instead examine individual decision making processes in absence of interactions with other agents, experiments provide a method to control for the framing, the incentive schemes and task played (see Amos Tversky and Kahneman (1974) and Holt and Laury (2002)). In the context of policy analysis, experimental and behavioural analysis can provide support during the policy design phase, the implementation phase and the evaluation phase (Dupoux M. et al., 2025). In the design phase, behavioural and experimental approaches help identify the decision making processes that should be targeted to achieve desired policy outcomes. For example, the identification of the cognitive biases such as anchoring effects, endowment effects and loss aversion effects, is crucial to design effective policies. In the implementation phase, experiments are useful to gather preliminary evidence on policy effects. Finally, in the evaluation phase, experimental methods measure the effect of existing policies and can suggest potential modifications.

Two key challenges in the field of experimental methods for policy analysis, and more in general in the fields of experimental economics and social sciences, are replicability, which deals with the internal validity of a study, and generalizability, which instead concern the external validity of a study.

A study is replicable if "the results are sustained when using other data"(Duvendack, Palmer-Jones, and Reed, 2017). Broadly speaking, replicability refers to the property of being robust to: i) re-analysis using the same data and the same methods (pure replicability); ii) to the collection of similar data and use of similar estimation methods (statistical replication); iii) and to the use of "alternative theoretical or conceptual approaches" (scientific replications) (Duvendack, Palmer-Jones, and Reed, 2017). Replication studies can prevent controversial practices such as HARKing "hypothesizing after the results are known"(Kerr, 1998), "p-hacking" following which researchers test their hypothesis using dif-

ferent models until one of them produces a significant ($p<0.05$) result, and frauds. Hence, replication rates provide useful insights on the reliability of the results. However, in experimental economics replication study are not common practice and often fail in their intent. For example, Camerer et al. (2016) successfully replicated 11 out of 18 studies, with replicated effects being on average 66% of the original estimate. More generally, replication rates in economics be as low as 42% (Chang and Li, 2015). A practice to improve the transparency and the credibility of research studies is pre-registration (Rubin, 2020), following which researchers must clearly define the hypothesis under examination and the estimation methods used before collecting data or before the analysis is conducted. While pre-registrarion may help to reduce the publication bias, its impact on replication rates is still unclear.

A study has external validity if its results can be generalised to a real world scenario. Finding a balance between realism, which contribute to the external validity of the study, and simplification, which is necessary to model complex system and manipulate its components to examine causal relationships, depend on the endpoint of the study and must be decided by researchers in the design phase. Some researches argue that having an unrealistic setting may produce unrealistic data. Falk and Heckman (2009) argue, however, that laboratory, field - and online - experiments should be seen as complementaries. While, experimental studies help policymakers understand the direction of a policy effect, field studies provide more accurate estimates on the effect size. As an example, consider the literature on nudges, one of the most common behavioural interventions in policy research. Recent studies have shown that effect sizes in experimental settings tend to be larger than those in field studies, however, the direction of the effect remains consistent between the two type of works (Maier et al., 2022; Vlasceanu et al., 2024; Mertens et al., 2022; DellaVigna and Linos, 2022).

This dissertation contribute to the growing field of experimental policy analysis with three relevant policy work in which experimental methods are used to examine causal relationships when observational data are private and not available (chapter 2), suffer from a censoring bias and are

not fit for causal interpretation (chapter 3) or are non-existent (chapter 4). All three studies have been pre-registered and received the authorization to conduct the experiment from the Joint Ethical Committee of Scuola Normale Superiore, Sant'Anna School of Advanced Studies, and IMT School for Advanced Studies Lucca.

The first study is presented in Chapter 2. It provides an experimental analysis of virtual currencies as a form of payment for transactions of loot boxes. Loot boxes are digital containers of randomized rewards commonly found in video games and mobile apps (Zendle, Walasek, et al., 2021), and are especially common in free-to-play games. Recently, scholars examined the mechanisms behind loot box transactions as they may produce undesired effects on consumers. For example, these mechanism may distort the economic behaviour of the average consumer and thus be considered a form of unfair practice (Cerulli-harms et al., 2020; Zendle, Cairns, et al., 2020; Macey and Hamari, 2019). This study examines the role of virtual currencies, money illusion, and non-intuitive exchange rates on consumers' spending behaviour for online transactions involving risk and ambiguity. The experimental methodology was a necessary part of this study, as data on consumers' behaviour was not publicly available being private property of game studios. We find that virtual currencies under fixed exchange rate induce subjects to increase spending, contributing to the debate on the neutrality of virtual currencies and experimental currency units. Moreover, when the nominal value of a virtual currency is higher than its real value, subjects reduce their willingness to pay and become more sensitive to risk and ambiguity. Lastly, the use of non-intuitive exchange rates further reduces the willingness to pay of our subjects (non significantly though). Based on our findings, we recommend policymakers to support the use of real-world currencies in loot box transactions to improve transparency and consumers' understanding.

The second study, presented in Chapter 3, examines the effect of an existing policy that makes real-time ER congestion levels available on the websites of Italian hospitals. In particular, we examine the effect of this policy on the number of inappropriate ER visits which can be responsi-

ble for ER overcrowding, an urgent urgent issue in the medical sector. First, this study models the decision-making process of a patient choosing between visiting an emergency room or visiting territorial health facility. Then, using an incentivized discrete choice experiment we test the effect of the policy in a setting in which participants are given the probability to be in a critical health status, while information on ER congestion levels are manipulated in a between-subject design. The use of observational data for this study would have raised two issues. A retrospective study, in fact, would not have allowed to observe the behaviour of patients choosing alternative forms of care, introducing a censoring bias. Moreover, it would not have allowed to register patients' belief about their health status. Our experimental setting overcome both issues, recording the choice for both alternatives and also controlling for patients' belief. Our results show that real-time congestion data increase the overall number of visits. In particular, we observe more inappropriate visits in intermediate levels of congestion. This evidence suggests that when information is available the uncertainty of ER visits decreases, leading risk-averse individuals to choose the ER more frequently. We also test the effect of an alternative policy in which intermediate levels of congestion are equally probable. The effect on the total number of ER visits is not significantly different from 0, suggesting that the alternative policy may combine transparency improvements with desired outcomes. Our study also investigates differences in ER visits between a scenario where individuals make decisions for themselves and one where they decide on behalf of someone unable to do so, finding more frequent visits in the *for-others* scenario. Additionally, participants show a preference for TH treatments, which however is not driven by moral considerations towards the use of ER resources. We conclude that, to avoid problems of overcrowding, hospitals should carefully plan the information provided on their website.

The third study is presented in Chapter 4. It examines the relationship between students' engagement in CyberTrials 2023 educational activities and their educational outcomes. CyberTrials 2023 is a game-based educational program in computer science and cybersecurity. The program

consisted of eleven theory driven lectures and nine hands-on training modules, which made students take part in a series of CTF challenges organized following a narrative framework. Given the innovative nature of the program, a survey-based approach was necessary to conduct the intended analysis. In particular, two surveys were distributed, one before the program and one right after. However, following issues in the recruitment process, only the second survey was used for the analysis. The program successfully engaged students, as indicated by the low dropout rate and the self-reported measure of interest. Moreover, our results indicate the existence of a positive relationship between higher engagement in training modules and students' educational outcomes and a null relationship between higher engagement in lectures and students' educational outcomes. Thus, we conclude by highlighting the importance of complementing theory-driven lectures with hands-on, gamified activities for computer science and cybersecurity education.

Supplementary figures and tables, as well as relevant robustness checks and analysis for each study are provided in the Appendix.

Chapter 2

An online experiment for consumer protection policies

This essay is based on: Toccafondi, Di Paolo, and Di Guida (2025), Virtual Currencies in Online Gaming Increase the Willingness to Pay for Loot Boxes: An Experimental Analysis. DOI: https://doi.org/10.2139/ssrn.5146007

2.1 Introduction

The video game industry is nowadays one of the driving forces behind the entertainment industry. In 2020 alone, the video game industry generated revenues for 23.3 billion euros across the European Union (Coenen and Lesser, 2021). Interestingly, the majority of these revenues were collected through "in-game extras and additional downloadable content for games already purchased" (64%) rather than "full game downloads" (25%) (Coenen and Lesser, 2021). This composition of revenues can be attributed, in part, to the increasing popularity of free-to-play games. The concept behind free-to-play games has dramatically changed the gaming landscape. Instead of charging users upfront with a purchasing fee, these games typically generate revenues through in-app purchases, such

as virtual goods, power-ups, or cosmetic items. To increase profits, video game developers exploit a variety of marketing practices specifically designed for these types of transactions. A prevalent commercial practice in the realm of free-to-play games is the use of loot boxes, digital containers of randomised rewards (Zendle, Walasek, et al., 2021). Consumers can purchase a loot box by exchanging a predetermined monetary sum. In return, they are granted access to a chance-based assortment of in-game content. Usually, consumers are not aware of the individual probabilities associated with each possible outcome. Moreover, in some cases, consumers are not given information on the set of possible items that the loot box can contain.

Recently, loot boxes received attention from scholars for their potential negative effects on consumers (Cerulli-harms et al., 2020). This reward mechanism has been compared to gambling activities (Drummond and Sauer, 2018, Leahy (2022)), finding a positive correlation between loot box spending and gambling problems among adolescents (Zendle, Meyer, and Over, 2019) and gamers (Zendle and Cairns, 2018).

Researchers have identified some of the most common characteristics of loot boxes, which typically combine different features and marketing practices. For example, loot boxes are often sold in exchange for virtual currencies with non-intuitive exchange rates, with time-limited offers, or in bundles (Cerulli-harms et al. (2020), Zendle, Cairns, et al. (2020), Macey and Hamari (2019)). These practices could potentially distort the economic behaviour of the average consumer and be considered unfair according to the European marketing framework (European Parliament and The Council of the European Union, 2015). Indeed, some of the features of loot boxes have been addressed as potentially unfair (Cerulli-harms et al. (2020), King and P. H. Delfabbro (2018), King, P. H. Delfabbro, et al. (2019), P. Delfabbro and King (2020), Petrovskaya and Zendle (2021)). Thus, regulating loot boxes could be relevant to increase transparency and protect consumers from inefficient allocation of resources.

Among the many features that could be examined, we focus on the use of in-game virtual currencies, given their relevance from an economic perspective. Virtual currencies can be found in most online games. For

example, *Magic: The Gathering Arena* uses gems and gold coins, *League of Legends* uses blue essence and RIOT points and *Fortnite* V-Bucks. Virtual currencies are used by game developers to sell loot boxes as well as cosmetics, characters and cards in games. While virtual currencies can reduce transaction costs and increase efficiency, they can disguise the true price of in-game products and, consequently, distort consumers spending behaviours. Often the exchange rates between virtual and real-world currencies are not intuitive, confusing the true value of a loot box. In addition, game developers typically set exchange rates with nominal prices higher than the corresponding real prices. This strategy can lead to the so-called "money illusion effect" (or face value effect), which has been shown to be relevant for consumers' product evaluation and spending behaviour (Raghubir and Srivastava, 2002).

In our study, we examine the interplay of virtual currencies, money illusion and non-intuitive exchange rates on consumers' spending behaviour in an incentivised randomised controlled trial. To investigate consumers' spending behaviour, we measure participants' Willingness-To-Pay (WTP) for acquiring different loot boxes. In our experimental setup, loot boxes are presented as a series of risky and ambiguous lotteries, where the possibility of receiving a reward is based on a random component. Risky lotteries allow us to control how different probabilities of winning/losing affect players' behaviour in such a setting, while ambiguous lotteries mimic loot boxes where the information about the probability of receiving a reward (or losing the endowment) is generic, a common feature in free-to-play games.

Our first result shows that in presence of virtual currencies subjects increase their willingness to pay. This evidence challenges the neutrality of virtual currencies as we show that consumers' behaviour is influenced simply by the use of fictitious currencies, even in cases where the exchange rate is 1 to 1. Moreover, we found a negative effect of money illusion for intuitive and non-intuitive exchange rates. Indeed, when the nominal value is higher than the real value of a virtual currency, the willingness to pay of agents drops dramatically as they become more sensitive to risk and ambiguity. Lastly, we find no evidence that the use of

non-intuitive exchange rates has an impact on agents' willingness to pay compared to intuitive exchange rates of comparable size.

In conclusion, we recommend policymakers to support the use of real-world currencies in loot box transactions by video game developers to improve transparency and consumer understanding.

The rest of the paper is organised as follows. In Section 2.2 we present the literature related to virtual currencies, money illusion and non-intuitive exchange rates. In Section 2.3 we present the experimental design. In Section 2.4 we present descriptive, non-parametric and econometric results for our research questions. In Section 2.5 we discuss our results and conclude with policy implications.

2.2 Literature review

In this section, we review the literature related to the effects under examination. As we focus on three different effects, which are only partially addressed together in the literature, we divide the section in thematic subsections. First, we examine previous research discussing the impact of experimental currency units and virtual currencies on economic choices. Then, we focus on the money illusion effect and the face value effect. Lastly, we review the literature related to the effects of non-intuitive currency conversions.

2.2.1 Virtual Currencies

Virtual currencies are defined as "digital representations of value, not issued by a central bank, credit institution or e-money institution, which in some circumstances can be used as an alternative to money" (European Central Bank, 2015). Virtual currencies are used in economic systems to reduce transaction costs and increase transaction efficiency. In video games and mobile apps, virtual currencies help frame in-game transactions as part of the game itself, avoiding the interference of real money with the game experience (Holin Lin and C. T. Sun, 2011). From the user perspective, virtual currencies facilitate trading activities and final trans-

actions (Alghamdi and Beloff, 2015); however, they may be considered an "unfair practice" (Petrovskaya and Zendle, 2021). Neither the economic nor the game science literature examines the behavioural effects of virtual currencies on consumers' spending behaviours; hence, it is still unclear whether the use of virtual currencies for loot box transactions can be rightfully considered an unfair practice.

The economic literature significantly contributed to the debate on the neutrality of experimental currency units (ECU)[1], showing that, depending on the context, ECUs may not be neutral. For example, Mazar, Amir, and Ariely (2008) showed that paying participants in a search task with physical tokens increased their rate of dishonest behaviour with respect to cash payments. Instead, Reinstein and Riener (2012) focused their study on the tangibility effect, showing that donations in a dictator game were higher when money was virtual compared to cash. The token effect and the tangibility effect were also studied together by Shen and Takahashi (2013), which compared offers and rejections of an ultimatum game where subjects played with cash or virtual points with an exchange rate of 1:1. Offers were lower and rejections were more frequent in the virtual points condition. The authors suggest that a higher risk aversion may have led participants in the cash condition to avoid rejection by increasing their offers in the first stage of the game. Drichoutis, Lusk, and Nayga (2015) studied the effect of experimental currency units on second-price auction (SPA) bids with induced values in a virtual environment. No significant differences were found in the reported induced values across different exchange rates, concluding that in the particular framework of SPA, experimental currencies do not distort participants' behaviour.

More recently, Salminen, Järvelä, and Ravaja (2018) studied the effect of virtual currencies, exchange rates and endowments on the spending behaviour of participants in a series of free-to-play game scenarios. This is the study closest to tackling the questions of interest for our study. Results show no significant differences in purchasing intentions between euros and diamonds, an example of hard currency[2]. This study, however,

[1] We consider virtual experimental currency units as an example of virtual currencies

[2] Hard virtual currencies can be obtained only in exchange for a fixed amount of money

has two limitations. First, using authentic free-to-play game scenarios led to a limited sample size. Second, the task is non-incentivised.

The current literature shows that, depending on the context, virtual currencies may have unexpected behavioural effects on participants and may not be neutral. However, to the best of our knowledge, no studies focus on the WTP for transactions involving risk and ambiguity. We contribute to the debate on the neutrality of virtual currencies by examining their effect on consumers' spending behaviour for risky and ambiguous lotteries in a virtual environment and with equal exchange rates. This contribution is useful to assess whether the use of virtual currencies in games and mobile apps can be considered an unfair practice. Adapting previous results to the context of interest, we expect participants to show less risk aversion in the virtual currency condition as shown in Shen and Takahashi (2013).

Hypothesis 1. *We expect participants to report higher WTP for risky and ambiguous transactions in the presence of virtual currencies, compared to situations where real currencies of comparable value are involved.*

2.2.2 Money Illusion

The money illusion effect is defined as "the tendency to think in terms of nominal rather than real monetary values" (Shafir, Diamond, and A. Tversky, 1997). In the context of online gaming, the nominal value of an item is represented by the price shown to the consumer. For example, a loot box may cost 260 gems. The real value is instead represented by the price of the item in terms of real-world currency. In the above example, gems can be sold with an exchange rate of 177 to 1. Thus, the loot box costs 1.47 euros, which represents the the real value of the item.

The money illusion effect is a well-studied phenomenon in the economic literature. Money illusion has been shown to be relevant in explaining nominal price inertia after negative and positive nominal shocks (Fehr and Tyran, 2001), as well as the relationship between inflation and real prices in housing markets (Brunnermeier and Julliard, 2008) and

(Salminen, Järvelä, and Ravaja, 2018).

stock markets (Cohen, Polk, and Vuolteenaho (2005), Acker and Duck (2013)). Furthermore, in a strategic environment, it has been shown that money illusion consistently leads players in an n-player pricing game to choose a Pareto inefficient equilibrium (Fehr and Tyran, 2007). Recently, contrary to previous expectations, the money illusion effect has been shown to be relevant over time in strategic interactions (Morales and Fatas, 2021).

In the context of consumers' behaviours, the money illusion effect takes different names depending on the specific situation. Gamble et al. (2002) show that the money illusion, under the name of euro-illusion, distorted price perceptions for goods and services when the Italian government introduced the euro. The effect was also found to be relevant in the presence of a budget constraint for low-priced products (Gamble, 2006). The euro illusion has also been studied in the context of charitable giving, and it was found to significantly increase donations in Italian churches for the year 2002 (Cannon and Cipriani, 2006). Money illusion also applies to product valuations using foreign currencies, taking the name of face value effect. Raghubir and Srivastava (2002) show that consumers overspend when the nominal value of a product is lower than its real value, and they underspend when the nominal value is greater than the corresponding real value. More recently, the face value effect was found to be relevant in 80% of tourist transactions (Saayman et al., 2022), especially for less valuable goods.

We contribute to the money illusion and face value effect literature by extending these concepts to the context of virtual currencies. We introduce a methodological novelty by controlling for the virtual currency effect.

Hypothesis 2. *We expect the WTP for online transactions involving risk or ambiguity to decrease when the nominal value of transactions is higher than the corresponding real value.*

2.2.3 Non-Intuitive Exchange Rates

Video games and online platforms often present non-intuitive exchange rates between real-world and virtual currencies, making price estimation difficult and cognitively demanding. As a consequence, consumers may resort to quick but inaccurate currency conversion strategies (Patrick Lemaire and Mireille Lecacheur (2001), Imbo, Duverne, and Patrick Lemaire (2007)). Therefore, non-intuitive exchange rates may distort consumers' spending behaviour and potentially represent another unfair marketing practice.

In the literature on estimation, currency conversion is an example of numerical estimation that involves two numerical quantities, one of which is exact (Campbell, 2005). P. Lemaire, M. Lecacheur, and Ferréol-Barbey (2001) show that participants in a cognitively demanding currency conversion task use 6 different strategies, most commonly employing decomposition, transformation and direct calculation strategies. Patrick Lemaire and Mireille Lecacheur (2001) investigated, instead, the efficiency of two different currency conversion strategies, showing that their average accuracy and execution time are influenced by the amount to be converted and the age of participants. Subsequently, the study was replicated by Patrick Lemaire (2007) 5 years later, revealing fewer strategies employed, faster execution times and higher accuracy rates. The study concludes that participants have gradually opted for quicker and more accurate strategies. However, according to Lemaire and subsequent research (Patrick Lemaire and Mireille Lecacheur (2001), Patrick Lemaire (2007)), there is evidence of deviation from exact conversions. If we assume a similar behaviour for online transactions with non-intuitive exchange rates, consumers are likely to adopt efficient but inaccurate conversion strategies and consistently deviate from situations where numerical estimations are not necessary.

We contribute to the literature on numerical estimation by investigating the effect of non-intuitive exchange rates between real-world and virtual currencies on the WTP for online transactions that involve risk. We study a unique scenario of a non-intuitive exchange rate, where the con-

version is challenging and not further affected by money illusion bias. Previous studies have not extensively researched this area, so we adopt an exploratory approach, and we test whether the willingness to pay under non-intuitive exchange rates differs significantly from the one observed in lotteries with intuitive exchange rates of comparable size.

Hypothesis 3. *We expect that in lotteries with non-intuitive exchange rates, participants report a different WTP for risky and ambiguous transactions compared to the one reported in transactions involving intuitive exchange rates.*

2.3 Experimental Design

To test our hypotheses, we run an experiment composed of 5 different treatments, each including 2 within-subject conditions (see Figure 1 for a summary of the experimental structure). All treatments shared the same structure and differed only for the experimental currency used. Each treatment consisted of four phases: 1) Instructions, 2) Decision Tasks, 3) Additional Questions, 4) Payment.

2.3.1 The Decision Tasks phase

In the Decision Tasks phase, participants faced 16 lotteries, which we used to estimate their willingness to pay (WTP) in situations involving different levels of risk and uncertainty (see Table 1 for a summary of the lotteries). The lotteries were either "Risky" (11 lotteries) or "Ambiguous" (5 lotteries). Risky lotteries reported the possible outcomes of the lottery and the probabilities associated with each outcome, while ambiguous lotteries reported the possible outcomes but only generic information about the probabilities associated. Risky lotteries followed the structure of Holt and Laury (2002), giving us a complete representation of the change in WTP from lotteries with sure probability of gain to lotteries with sure probability of loss. These types of choices recreate the situation where loot boxes contain items of high and low values and consumers are informed about the probabilities to obtain them. Ambiguous

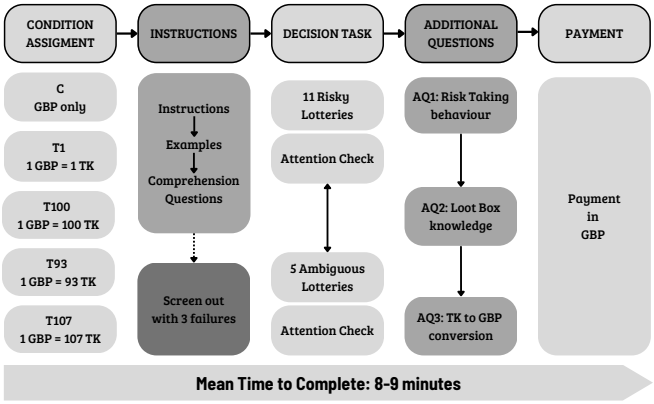


Figure 1: Summary of this study structure.

In the condition assignment phase, each participant is assigned to one condition. In the instructions phase, participants receive instructions on the task and are provided with an example to familiarise themselves with the task. Finally, they are given a set of comprehension questions for which a correct answer is necessary to take part in the study. In the decision task phase, participants are required to report their WTP for 16 different lotteries and complete 2 attention checks. In the additional question phase, participants are asked some additional questions. Additional question 3 is administered in conditions T1, T100, T93 and T107. In the payment phase, participants receive a participation fee of £0.4 and a bonus payment depending on the outcome of the lottery.

lotteries were designed to mimic settings commonly observed in free-to-play video games, where the probabilities of win/loss are unknown, or players are only informed that the chance to win or lose randomized content is either "common" or "rare". We then designed five ambiguous lotteries, two that emphasised the chance of winning (either "common" called A1, or "rare" called A2) and two that focused on the chance of losing (again "common" or "rare", called A3 and A4 respectively). Additionally, we devised a lottery where no information is provided about the chance of either outcomes (A5). By construction, lotteries A1 with A4, and A2 with A3 were equivalent (though the participants were not aware of this). Participants were randomly presented with two blocks of lotteries (risky and ambiguous). The order of the lotteries within each block was randomised too.

Table 1: Risky and ambiguous lotteries in this study

Risky			**Ambiguous**		
	% Win	% Lose		% Win	% Lose
R0	0	100	A1	Common	x
R1	10	90	A2	Rare	x
R2	20	80	A3	x	Common
R3	30	70	A4	x	Rare
R4	40	60	A5	Unknown	Unknown
R5	50	50			
R6	60	40			
R7	70	30			
R8	80	20			
R9	90	10			
R10	100	0			

The five between-subject treatments differed for the experimental currency used. The control treatment (henceforth C) was administered using a real-world currency which was well known to the participants: the British Pound Sterling (GBP). In treatment T1, participants used experimental *tokens*, whose exchange rate at the end of the experiment was 1:1, i.e., each token was worth 1 GBP. As the nominal value and the real value of the currency used in C and T1 is the same, we interpret any

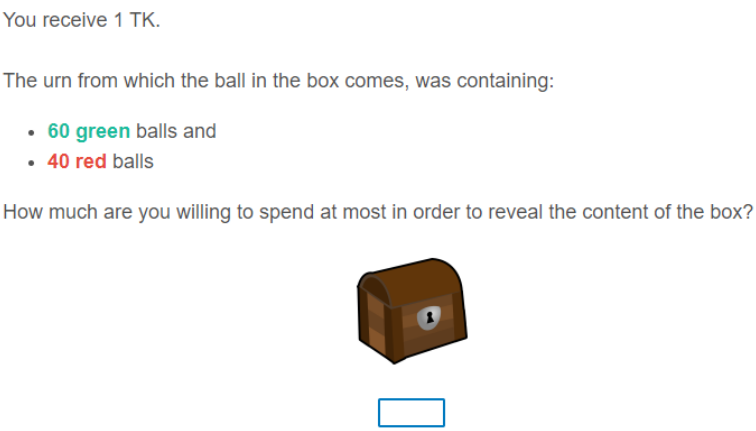


Figure 2: Task's interface in condition T1 and lottery R6

observed difference as the product of switching from real to virtual currency. Therefore, comparing the WTP in treatments C and T1 allows us to study the emergence of a "virtual currency effect". Treatment T100 was designed to test the money illusion bias. In this treatment, participants used virtual tokens whose exchange rate was 100:1, i.e., 100 tokens were worth 1 GBP. In such treatment, the nominal value of the currency is much higher than the real value. Therefore comparing treatments T100 and T1 allows us to test the money illusion effect in a setting where both currencies are virtual. Lastly, treatments T93 and T107 allow us to verify the stability of the money illusion effect, i.e., whether not only the distance between real and nominal value, but also the complexity of calculating the exchange rate play a role. In these last treatments, the exchange rate was 93:1 and 107:1, resulting in a comparable nominal value to that of T100. However, the mathematical calculation is more demanding due to the more intricate exchange rate. Therefore, any difference observed between T100 and T93&T107 can be attributed to a non-intuitive exchange rate effect.

2.3.2 WTP Estimation

We measured participants' WTP following the Becker DeGroot Marschak (BDM) procedure (G. M. Becker, DeGroot, and Marschak, 1964), as described in Eisenberger and Weber (1995). The BDM procedure was proven to be the most accurate predictor of the real WTP [3] (Miller et al., 2011). In total, participants had to play 16 rounds, one for each lottery. At the beginning of each round, participants received an endowment (£1 or the equivalent value in tokens (TK): TK1, TK100, TK93, TK107) and were then shown a box containing either a green or a red ball with a certain probability (Figure 2[4] shows the interface for the task). They were then instructed to report their maximum willingness to spend to reveal the box's content. This information represents their WTP. At the end of the study, we randomly selected one lottery and extracted its cost C from a Uniform distribution $U[0.01, 1]$. If $WTP < C$, the content of the box was not revealed, and the final payoff of the participant was equal to £ 1. If instead, $WTP \geq C$, the content of the box was revealed. Participants got £ $(2 - C)$ if a green ball was revealed (a win) and £ $(1 - C)$ if a red ball was revealed (a loss) [5]. Information about the outcome of the lottery, its cost and content were revealed at the end of the experiment.

2.3.3 Experimental Procedure

We conducted a pre-registered[6] online experiment. Participants were recruited through the platform Prolific (Palan and Schitter, 2018), and the

[3] Four different methods were compared: (i) Open-ended question; (ii) Choice-based conjoint analysis; (iii) BDM incentive-compatible mechanism; (iv) Incentive-aligned choice-based conjoint analysis

[4] The depiction of the box is copyrighted (Creative Commons CC0 License) and dedicated to the public domain by copyright holder https://commons.wikimedia.org/wiki/File:Chests_1.svg

[5] In Eisenberger and Weber (1995) the cost C was extracted from a Uniform distribution $U[0, 1]$, giving the chance to win up to £2 to participants bidding £0. A more realistic setting would not give a positive probability of winning a lottery to participants reporting a WTP equal to 0. By increasing the minimum cost to £0.01, we gave the possibility to win £1 additional only to participants reporting a strictly positive WTP. Furthermore, participants that did not wish to participate in the lottery could do so by choosing WTP=0

[6] `https://aspredicted.org/2J4_B91`

experimental interface was developed using Qualtrics (Qualtrics, 2005). The experiment was conducted online to recreate the setting in which virtual currencies are commonly used. Specifically, the Prolific platform has been chosen because it allows to have a representative sample of UK participants, which is fundamental to ensure the validity of the results in the control condition.

Upon accessing the study, participants were randomly assigned to one of the treatments (C, T1, T100, T93, T107). After reading the instructions, subjects answered four comprehension questions (Instructions phase). Participants who failed to answer all comprehension questions correctly within three attempts were excluded from the study. After successfully completing the comprehension questions, participants took part in the Decision Tasks, where the different lotteries were presented, one at a time. After the Decision Tasks, participants were asked three additional questions (Additional Questions phase). The first was to report their risk preference using the item proposed by Falk, A. Becker, et al. (2018) (AQ n.1), and the second tested their familiarity with the definition of a loot box (AQ n.2). These questions are useful controls to be included in the analysis (Figure A.3 shows AQ n.1, AQ n.2 and AQ n.3). As last step, we informed participants of their winnings, presenting the amount in the same currency used throughout the experiment. Participants in T1, T100, T93, and T107 were asked to convert their token-based winnings into GBP (AQ n.3). This task was used to further test the participants' level of understanding regarding exchange rates, and it did not influence their final payment. Once answered this last question, participants were paid (Payment phase).

We recruited 752 participants through the Prolific platform, among which 116 failed 3 consecutive times the comprehension checks and were not allowed to proceed. Additionally, 34 participants voluntarily withdrew from the study. Data were collected from 602 participants, which corresponds to the pre-registered sample size[7]. We recruited 153, 151

[7]The sample size was computed *a-priori* using the G-power 3.1 tool for repeated measurement and within-between interaction considering a significance level of 0.05, a power of 0.9 and an effect size of0.25. However, based on the fact that in our data normality assumption is rejected and different lotteries display different variances, a standard com-

and 149 participants in conditions C, T1 and T100, respectively, while 73 and 76 in conditions T93 and T107. Treatments T93 and T107 offer a joint sample of 149, comparable to the size of the other treatments. Data from treatments T93 and T107 will be mostly analyzed jointly, as together, the two treatments allow to test the impact of non-intuitive exchange rate without the possible confound of a "two digits" VS "three digits" effect. The sample included only UK participants to control for nationality-fixed effects and guarantee they were accustomed to the use and value of GBP. The sample was balanced in age, gender, risk preferences, percentage of students, loot box knowledge and time taken in the experiment across treatments. We report relevant sample balance tests in the Appendix Table A.1. Relevant summary statistics for the participants that satisfy our inclusion criteria are reported in Table 2. The experiment lasted around 9 minutes, and participants received a participation fee of £0.4, plus an average bonus payment of £1.13.

Table 2: Descriptive statistics for this study.

	Control	T1	T100	T93	T107	Total
Age	34.26	35.30	34.87	34.36	34.38	34.70
	0.66	0.69	0.67	0.89	0.86	0.08
Female	0.62	0.54	0.59	0.60	0.55	0.60
	0.04	0.04	0.04	0.06	0.06	0.02
Risk Preference	5.31	5.35	5.46	4.99	5.16	5.30
	0.18	0.18	0.21	0.28	0.27	0.10
Student	0.19	0.10	0.15	0.10	0.16	0.14
	0.03	0.02	0.03	0.03	0.04	0.01
Loot Box Knowledge	0.59	0.54	0.57	0.52	0.62	0.57
	0.04	0.04	0.04	0.06	0.06	0.02
Duration in Minutes	8.28	8.84	8.72	9.59	8.93	8.77
	0.29	0.32	0.30	0.79	0.40	0.17
Observations	153	151	149	73	76	602

parison of means was not feasible and we had to adjust the test ex-post

2.4 Results

In this section, we report the results of our statistical analyses. Section 2.4.1 presents descriptive and non-parametric analyses, and Section 2.4.2 presents the results of an econometric analysis. To allow for cross-treatment comparisons, participants' WTP is normalized by dividing it by the exchange rate, thus obtaining a continuous variable in the interval $[0, 1]$.

To verify that exchange rates in conditions T93 and T107 are indeed perceived as less intuitive, we test whether the rate of correct answers to the additional question n.3 differs across treatments. We create a dummy variable *Correct Conversion*, which takes value 1 if the participant reported a correct TK to GBP conversion and 0 otherwise. In Table 3, Panel A reports the average rates of correct answers and standard errors across conditions. Rates of correct answers in conditions T93 and T107 are significantly smaller with respect to conditions T1 and T100 (see Table 3 Panel B), confirming that participants perceive exchange rates in conditions T93 and T107 as significantly less intuitive. Results hold for a Bonferroni correction for multiple testing. No difference is found between conditions T1 and T100 nor between conditions T93 and T107, as expected.

To verify that our experimental design produces results in line with those in the literature, we compare the average WTP of equivalent lotteries between our study and the one of Eisenberger and Weber (1995). Specifically, we compare the average WTP of lottery R5 in condition C and the WTP(+) of Urn R in Eisenberger and Weber (1995). Similarly, we compare the average WTP of lottery A5 and the WTP(+) of Urn A. To make the comparison more intuitive, we divide by 10 the WTP(+) for Urn R and Urn A of Eisenberger and Weber (1995). We obtain similar levels of WTP between risky lotteries (0.390 for lottery R5 and 0.423 for Urn R) and ambiguous lotteries (0.351 for lottery A5 and 0.351 for Urn A), indicating that our experimental design is in line with the closest and most relevant experimental design in the literature.

Table 3: Summary statistics for the dummy variable Correct Conversion (Panel A). Test for less intuitive exchange rates (Panel B).

Panel A

	Control	T1	T100	T93	T107	Total
Average Values	-	0.91	0.90	0.56	0.71	0.82
S.E.	-	0.02	0.02	0.06	0.05	0.02

Panel B

	T1/T100	T1/T93	T1/T107	T100/T93	T100/T107	T93/T107
P-values	0.814	<0.001	<0.001	<0.001	<0.001	0.085

Panel A. We report average values and standard errors across conditions. Panel B. Pairwise comparison between treatments. We report the p-values for a series of Wilcoxon Rank-Sum tests.

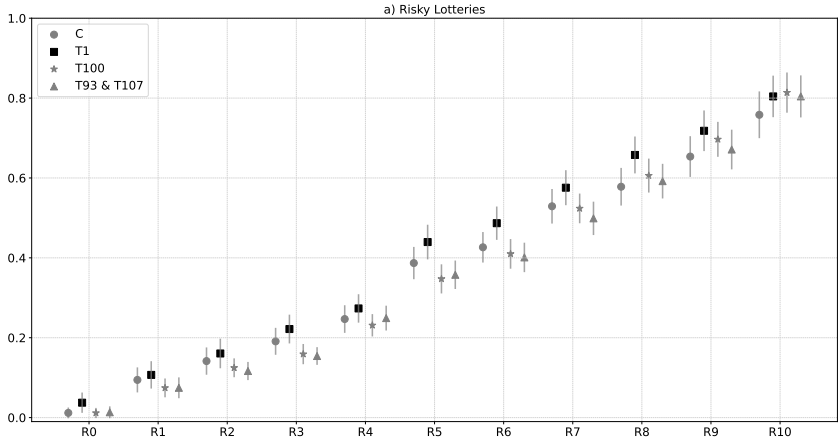
2.4.1 Descriptive and Statistical Analyses

Before focusing on our hypotheses, we examine the effect of loss aversion, framing effect and ambiguity aversion on the average WTP. Figure 3 and Figure 4 present the average WTP across conditions and lotteries together with the corresponding 95% confidence intervals. In the following analyses, conditions T93 and T107 will be merged and treated as a unique treatment with non-intuitive exchange rates.

Figure 3 panel a) shows that in risky lotteries, the marginal increase of the average WTP is linear. However, starting from lottery R5 the marginal growth rate is higher with respect to the one between lotteries R0 and R4. Moreover, between lotteries R4 and R5 there is a relevant discontinuity in the average WTP. The observed pattern may be attributed to higher levels of loss aversion in lotteries from R0 to R4[8]. When participants are offered the choice to bid on lotteries R0 to R4 (for which the probability of losing is higher than 50%), the WTP is reduced compared to R5 - R10 by two components: first, the average WTP reduces according

[8]Loss aversion is defined as the psychological propensity that losses loom larger than equal-sized gains relative to a reference point (Kahneman and Amos Tversky (1979), Kahneman, Knetsch, and Thaler (1991))

(a)



(b)

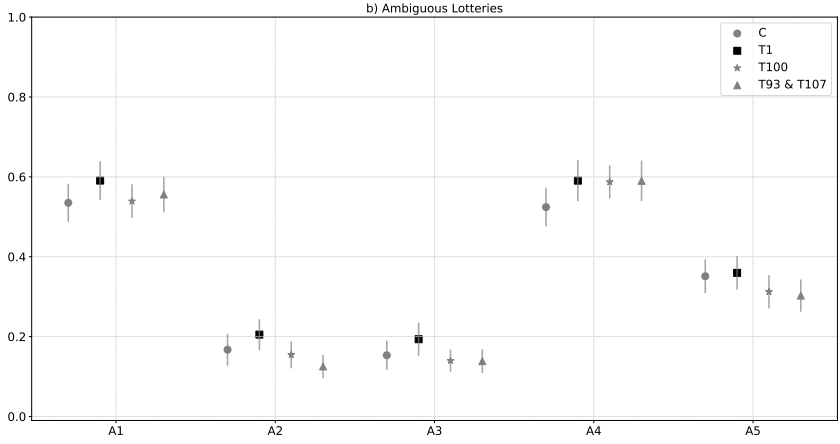


Figure 3: Average WTP across conditions and lotteries in risky and ambiguous lotteries and 95% confidence intervals.

to the linear trend; second, by an additional amount to satisfy loss aversion preferences and mitigate the possibility of paying for a lottery that's more likely to result in a loss. A similar effect has been documented in Gächter, Johnson, and Herrmann (2022), where 71% of participants exhibited preferences characterized by loss aversion in experiments involving risky choices. Figure A.1 visually depicts the trends.

With regard to ambiguous lotteries, it is shown in Section A.4 of the appendix that a framing effect influences the average WTP between lotteries A1 and A4. A null effect is instead found between lotteries A2 and A3.

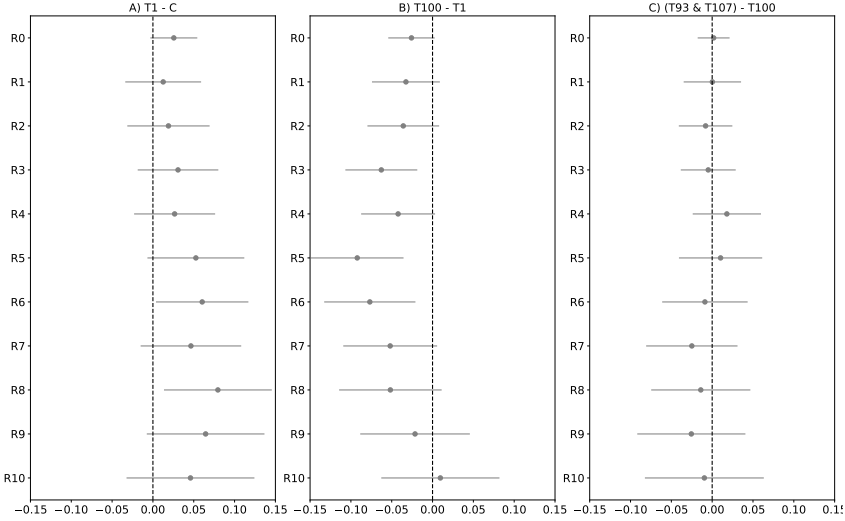
In Section A.5 of the appendix, we compare risky and ambiguous lotteries as participants may express ambiguity aversion (Ellsberg (1961), Abbey et al. (2019)). Participants show ambiguity aversion preferences, i.e. smaller levels of WTP for ambiguous lotteries, in lotteries A2, A3 and A5. Opposite effects are found for lottery A4, i.e. the lottery with a high probability of winning in the losing framing. Finally, no differences are found for lottery A1.

Our results so far are aligned with the literature, which allows us to move to the testing of our hypotheses. To test the impact of the different treatments on the average WTP, a standard comparison of means is not feasible for two reasons. First of all, the normality assumption for the average WTP across lotteries is consistently rejected (all p-values<0.001, see Section A.3 of the Appendix). Second, some lotteries display unequal variances, as confirmed by a series of Levene's tests of homogeneity of variances (see Appendix A.3). Following the work of Anvari, Verdeș, and Marchiori (2022), we use Cohen's d effect sizes and non-parametric methodologies based on the analysis of the estimated effect size. Figure 4 presents the difference in average WTP (and 95% CI) across conditions and lotteries. According to H1, we expect that the use of virtual currency should reduce subjects' risk aversion, i.e. to observe a higher WTP in treatment T1 than in treatment C. Indeed, in both panel A (risky lotteries) and panel D (ambiguous lotteries), the average WTP is systematically higher in T1 than in C. Table 4 shows the estimated Cohen's d effect sizes, that are all negative when comparing T1 and C. We test the hypothesis of

different WTP between conditions T1 and C. Under the null hypothesis of no virtual currency effect, we can model the distribution of positive and negative effect sizes as a Binomal $(16, 0.5)$. However, the probability of obtaining all positive effect sizes is equal to 0.00001. Hence, our results indicate a significant effect of the virtual currency on the average WTP of participants.

H2 addresses the Money Illusion effect. According to our hypotheses, we expect the WTP for online transactions involving risk or uncertainty to decrease when the transaction's nominal value is higher than the corresponding real value. The WTP in T100 should then be systematically lower than the WTP in T1. As shown in Figure 4 and Table 4, the average WTP is lower in T100 with respect to T1 in all lotteries except for lottery R10. As previously done for hypothesis 1, we model the distribution of positive and negative effect sizes as a Binomial $(16, 0.5)$. Under the null hypothesis of null money illusion effect, the probability that only 1 out of 16 lotteries exhibit a positive effect size is equal to 0.0002. Hence, our results indicate a significant effect of money illusion on the average WTP of participants. To further support such claim, we complete the same analysis comparing T93&T107 with T1, replicating the results (see Appendix A.6).

Finally, we examine H3, which tests the impact of non-intuitive exchange rates on the average WTP. We merge data from T93 and T107 and compare them against T100. As shown in Figure 4, the average WTP difference is approximately 0 in each lottery. Moreover, Table 4 reports very small effect sizes. As before, we test the null hypothesis of no effect for non-intuitive exchange rates. Table 4 shows 5 negative and 10 positive effect sizes. We exclude Lottery R1 as the corresponding effect size is equal to 0. Under the null hypothesis, the probability that 10 out of 15 lotteries exhibit a positive effect size, given an underlying Binomial distribution $(15, 0.5)$, is equal approximately to 0.0916, which does not allow us to reject the null hypothesis. Hence, our results do not provide support to the hypothesis that non-intuitive exchange rates might impact the WTP.



(a) (b)

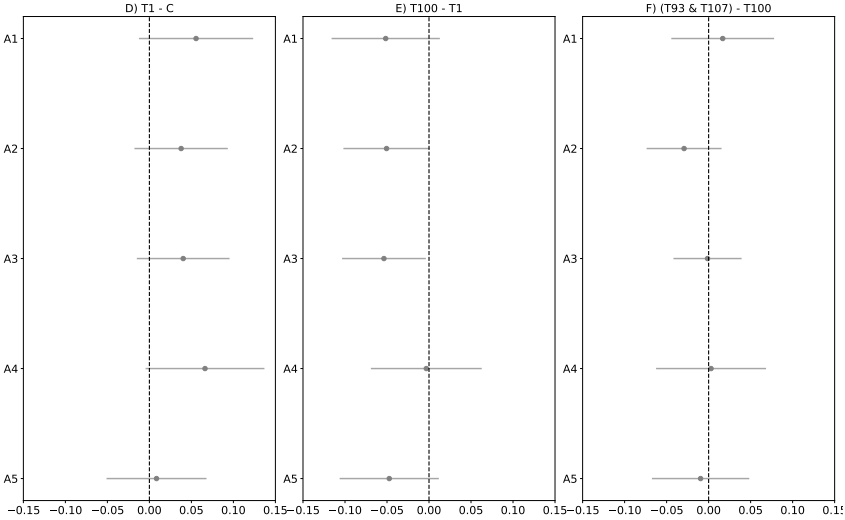


Figure 4: Average difference between WTP across conditions and lotteries and 95% confidence intervals.

Table 4: Cohen's d Effect Sizes and 95% CI

Risky Lotteries

	T1-C		T100-T1		T93&T107-T100	
	Effect Size	CI	Effect Size	CI	Effect Size	CI
R0	0.20	[-0.020, 0.380]	-0.21	[-0.370, 0.010]	0.02	[-0.190, 0.230]
R1	0.06	[-0.180, 0.290]	-0.18	[-0.390, 0.030]	0.00	[-0.230, 0.220]
R2	0.08	[-0.140, 0.310]	-0.18	[-0.380, 0.050]	-0.06	[-0.280, 0.170]
R3	0.14	[-0.080, 0.380]	-0.32	[-0.540, -0.090]	-0.03	[-0.230, 0.200]
R4	0.12	[-0.080, 0.340]	-0.21	[-0.420, 0.020]	0.10	[-0.130, 0.330]
R5	0.20	[-0.040, 0.420]	-0.37	[-0.590, -0.150]	0.05	[-0.200, 0.270]
R6	0.24	[0.020, 0.470]	-0.31	[-0.530, -0.080]	-0.04	[-0.270, 0.170]
R7	0.17	[-0.040, 0.380]	-0.20	[-0.440, 0.020]	-0.10	[-0.330, 0.120]
R8	0.27	[0.040, 0.510]	-0.19	[-0.420, 0.040]	-0.05	[-0.280, 0.180]
R9	0.20	[-0.030, 0.440]	-0.07	[-0.300, 0.170]	-0.09	[-0.330, 0.140]
R10	0.13	[-0.090, 0.350]	0.03	[-0.200, 0.250]	-0.03	[-0.250, 0.200]

Ambiguous Lotteries

	T1-C		T100-T1		T93&T107-T100	
	Effect Size	CI	Effect Size	CI	Effect Size	CI
A1	0.18	[-0.050, 0.410]	-0.18	[-0.400, 0.050]	0.06	[-0.180, 0.300]
A2	0.15	[-0.090, 0.370]	-0.22	[-0.460, -0.006]	-0.15	[-0.370, 0.080]
A3	0.16	[-0.060, 0.370]	-0.24	[-0.440, -0.003]	-0.01	[-0.250, 0.240]
A4	0.21	[-0.010, 0.450]	-0.01	[-0.230, 0.230]	0.01	[-0.220, 0.260]
A5	0.03	[-0.190, 0.250]	-0.18	[-0.410, 0.030]	-0.04	[-0.280, 0.170]

2.4.2 Regression analysis

To further examine the hypotheses, we complement the results obtained in Section 2.4.1 with an econometric analysis conducted using a linear random effect model with lottery-specific effects. The model, represented by Equation 2.1, is defined for individuals $i \in (1, ..., N)$, treatments $m \in (0, ..., 3)$ and lotteries $l \in (1, ..., 16)$.

$$WTP_{il} = \beta_1 + \sum_{m=0}^{3} \left(T_{i,m} \, \beta_m \right) + \mathbf{X}_i' \boldsymbol{\beta} + L_l + \alpha_i + \epsilon_{il} \tag{2.1}$$

In Equation 2.1 WTP_{il} is the dependent variable and represents the Willingness to Pay for individual i in lottery l; β_1 is a constant term; $T_{i,m}$ are dummy variables representing the m treatments and β_m are the associated marginal effects; $\mathbf{X}_i'$ is a (1xK) matrix containing the individual characteristics presented in Table 2 , with $\boldsymbol{\beta}$ being the characteristics' marginal effects; L_l are lottery fixed effects; α_i are the individual-specific random effects to control for individual-specific heterogeneity; ϵ_{il} is the idiosyncratic component. Econometric results of Equation 2.1 are reported in Table 5. The econometric analysis is conducted separately on risky lotteries, models (1) and (2), and ambiguous lotteries, models (3) and (4). In models (2) and (4), we introduce the covariates reported in Table 2. Namely, we consider the variables *Age*; *Female*, a dummy equal to 1 if the participant is female; *Student* a dummy equal to 1 if the participant is currently a student; *Risk preference*, which takes value between 0 and 10; *Loot box knowledge*, a dummy equal to 1 if the participant reports knowing what a lootbox is; and *Attention checks*, a dummy equal to 1 if the participant successfully completed both checks. Analyses in Table 5 take condition C as baseline to examine the treatment effects. In both models (1) and (3), the coefficient for condition T1 is positive and significant indicating that using a virtual currency instead of a real one increases participants' WTP. This supports hypothesis 1 and the emergence of a virtual currency effect. Introducing the individual characteristics decreases the estimated effect from 4.2% to 3.8% and 3.6%, respectively for risky and ambiguous lotteries, still supporting H1. None of the individual characteristics per se seems to have an important impact, with the

exception of risk preferences, according to which subjects with higher risk propensity have a higher WTP (2.3% and 3% more in risky and ambiguous lotteries). Coefficients for T100 and T93&T107 are both negative, indicating that WTP decreases when the transaction's nominal value is higher than the corresponding real value. To test the money illusion effect without the real-currency VS virtual-currency confound, we compare the post-estimation differences of T100 and T93&T107 with T1 (see Table 6). For all models (1 to 4) subjects in T100 have a WTP of around 4% lower than that observed in T1, while WTP in T93&T107 spans from 3.3% to 5% less, depending on the model. All these differences are significant. The inclusion of covariates (models (2) and (4)) does not modify the magnitude nor the significance of the difference between conditions. Overall, these results support Hypothesis 2, indicating that the money illusion effect is reducing participants' WTP for risky and ambiguous lotteries, both with intuitive and non-intuitive exchange rates.

Lastly, the role of non-intuitive exchange rates is examined by comparing the average WTP of conditions T100 and T93&T107. Models (1) and (3) in Table 6 indicate negative coefficients, while models (2) and (4) show instead positive coefficients. Overall, the effects are non significant and very small (less than 1%) suggesting an overall null effect of non-intuitive exchange rates. Similar considerations can be drawn comparing coefficients in row 1 and 2 of Table 6, where changes T100 and T93&T107 induce similar changes in WTP compared to T1.

2.5 Discussion and conclusions

Loot box transactions are very common in free-to-play games and mobile apps. Recently, however, loot boxes have received attention from scholars for their potential negative impact on consumers. We examine three practices common to most loot box transactions that could potentially distort the economic behaviour of the average consumer. In accordance with the European marketing framework, these practices could be considered unfair and should receive particular attention from policymakers. An online randomized controlled trial is conducted to examine

Table 5: Random effect estimates of the average WTP

	Risky		**Ambiguous**	
	(1)	(2)	(3)	(4)
T1	0.042**(0.018)	0.038**(0.017)	0.042*(0.022)	0.036*(0.020)
T100	-0.002 (0.017)	-0.006 (0.016)	0.000(0.021)	-0.006 (0.020)
T93&T107	-0.008(0.018)	-0.004(0.017)	-0.004(0.022)	0.003(0.021)
Attention checks		0.021*(0.011)		0.012(0.014)
Age		0.001(0.001)		0.001(0.001)
Female		0.019(0.012)		0.043***(0.014)
Student		-0.020(0.017)		-0.030(0.020)
Risk Preference		0.023***(0.003)		0.030***(0.003)
Loot Box knowledge		-0.001(0.012)		-0.006(0.015)
Constant	0.365***(0.013)	0.200***(0.039)	0.346***(0.016)	0.120**(0.047)
N	6622	6589	3010	2995

Table 6: Marginal estimates across treatments for models in Table 5

	Risky		**Ambiguous**	
	(1)	(2)	(3)	(4)
Conditions	WTP	WTP	WTP	WTP
T100 - T1	-0.044***(0.017)	-0.044***(0.015)	-0.041**(0.020)	-0.042**(0.018)
T93&T107 - T1	-0.050***(0.017)	-0.042**(0.016)	-0.045**(0.020)	-0.033*(0.019)
T93&T107 - T100	-0.006 (0.016)	0.002(0.016)	-0.004(0.020)	0.008(0.019)

the influence of virtual currencies, money illusion, and non-intuitive exchange rates on consumers' spending behaviour. Participants' WTP is measured across a series of risky and ambiguous lotteries (representing different loot boxes), and in four between-subjects conditions useful to examine each effect separately.

2.5.1 The Virtual Currency effect

In our experimental setting, participants show a significantly higher WTP when using virtual currencies compared to using GBP, even though the exchange rate is 1:1. This phenomenon is observed both in risky and ambiguous lotteries. The reasons behind this effect are yet to be explored. In accordance with Cerulli-harms et al. (2020), the average WTP could increase as a consequence of smaller levels of pain of paying, which is defined as the immediate sense of displeasure associated with paying the price of a transaction (Prelec and Loewenstein (1998), Mazar, Plassmann, and Robitaille (2016)). Recently, it has been shown that credit cards can mitigate the pain of paying effect, increasing consumers' expenditure (Park, Lee, and Thomas (2021), Huang, Lim, and Lin (2021)). A similar effect may influence participants in our study. Lesser known and intangible currencies can deceive the opportunity cost of a transaction, decreasing participants' pain of paying and increasing the overall spending. Our result contributes to the literature discussing the neutrality of virtual currencies. We show that the spending behaviour of the average consumer can be affected by the currency itself, as exchange rates between conditions are fixed by design. However, we hesitate to conclude that using virtual currencies for transactions represents an unfair marketing practice. According to the European Parliament and The Council of the European Union (2015), an unfair practice must also be "contrary to the requirements of professional diligence", an assessment which is beyond the scope of this analysis. Policy-makers should, however, put particular attention on transactions involving risk and virtual currencies to counteract inefficient allocation of resources. Future research on the matter should examine the exact causes of this phenomenon. In partic-

ular, it is relevant to understand how virtual currencies modify the idea of transactions' opportunity cost and the relationship between the use of virtual currencies and the pain of paying effect. Furthermore, considering the widespread adoption of virtual currencies in both economic and psychological experiments, future research should explore the relevance of the effect in different tasks, such as investment, resource allocation, or market dynamics.

2.5.2 The Money Illusion Effect

According to our results, when the nominal value of a transaction is higher than its real value, consumers decrease their WTP as a consequence of higher levels of risk aversion. This result contributes to the literature on money illusion and face value effect by showing their validity for transactions involving virtual currencies. Moreover, while previous studies in the context of consumer behaviour compared situations with different exchange rates and different currencies, our experiment varies the exchange rate maintaining the currency fixed. This reveals that the face-value effect can be the result of the sole difference in exchange rates. Furthermore, our experiment shows evidence of money illusion effect between two intuitive exchange rates, namely conditions T1 and T100. This evidence suggests that the money illusion effect can be independent of computational skills. The results holds when comparing T1 with treatments where the exchange rate was computationally more complex to manage (T93&T107). Future research should examine the face-value effects for different exchange rates and, in particular, it should examine the face-value effect for virtual currencies on exchange rates smaller than 1, for which previous studies found overspending behaviours (Raghubir and Srivastava, 2002).

It is worthy to discuss the observed interaction of virtual currencies and money-illusion effects, comparing the average WTP in condition C with that of conditions T100 and T93&T107. As shown in Section A.7 of the Appendix, on average, the WTP between condition C and conditions T100 and T93&T107 cannot be interpreted as significantly dif-

ferent. It could be concluded that the virtual currency effect is counterbalanced by the money illusion effect or, alternatively, that participants bid correctly by thinking that 1 TK equals 1 penny. However, even though Table A.4 shows small and not statistically significant effect sizes, it should be noted that comparing conditions T100 with C, and conditions T93&T107 and C, the corresponding effect sizes are consistently negative for lotteries with a small probability of winning, i.e. lotteries from R1 to R7 and lotteries A2, A3 and A5. On the other hand, lotteries with a higher chance of winning, i.e. lotteries R8, R9, R10, A1 and A4, show a systematically positive effect size. This pattern may suggest an effect of loss aversion when the nominal value of a currency is higher than the real value. Future research should examine the validity of these claims in detail in the context of virtual currencies.

2.5.3 The non-intuitive exchange rates effect

An open question is whether non-intuitive exchange rates influence risk attitudes and consumers' WTP. Our results suggest that this might not be the case. In our setting, the average WTP does not significantly differ between conditions T100 and T93&T107, nor between C and T93&T107. A possible interpretation is that participants use effective conversion strategies. However, accurate conversion strategies can be ruled out based on the findings presented in Section 2.4.1, which indicate that participants were unable to convert the amounts they acquired during the task precisely. However, caution should be exercised, as the observed null effect of non-intuitive exchange rates could be attributed to the experimental methodology we employed. Specifically, exchange rates in conditions T93 and T107 were intentionally chosen to be non-intuitive and to exclude any potential face-value effects. The similarities in exchange rates between conditions T93 and T107 with that of condition T100 may have caused participants to employ fast but inaccurate conversion strategies, which may have gone undetected by the experiment's measures. Future research should examine more in detail how non-intuitive exchange rates could affect consumers' WTP. This topic is particularly relevant as

the majority of transactions in video games and mobile apps involve non-intuitive exchange rates.

2.5.4 Final remarks

The video game industry generated revenues for 23.3 billion euros across the European Union, in 2020 alone. In this study, we examined the role of virtual currencies, money illusion and non-intuitive exchange rates on consumers' spending behaviour for online transactions involving risk and ambiguity. One of our goals was to contribute to the recent literature that explores whether the use of virtual currencies could be considered an unfair practice. Based on our results, we recommend that policymakers further investigate the impact of virtual currencies and high exchange rates. Indeed, virtual currencies have the potential to decrease the transparency of transactions and induce consumers to allocate resources in undesired ways.

Chapter 3

An online experiment to reduce inappropriate demand of urgent healthcare services

3.1 Introduction

A relevant issue in the medical sector is emergency room (ER) overcrowding. This phenomenon is widely documented worldwide (M. L. V. Carret, A. C. G. Fassa, and Domingues, 2009; Morley et al., 2018; Pines and Griffey, 2015) and has been shown to be related to an increase in medical staff errors and patient waiting times, thus increasing patient dissatisfaction and patient mortality (Morley et al., 2018). Moreover, ER overcrowding increases the risk of violent episodes towards medical personnel (Morley et al., 2018).

ER overcrowding can occur due to a shortage of medical resources or excessive demand for medical services. Supply shortages can be caused by limited supply of medical personnel, limited access to primary care, delays in conveying laboratory results, and an inefficient use of the healthcare systems resources (Morley et al., 2018; Estey et al., 2003; Van Der Lin-

den et al., 2017; Uscher-Pines et al., 2013). An excessive demand instead, can occur following an unexpected exogenous shock, such as a natural disaster, or when the demand itself is inappropriate (Su et al., 2021). In 2022 inappropriate visits on the Italian territory represented 22% of the total visits to the ER (Randazzo, 2024), for a total of 3.8mln visits[1]. In addition to negative externalities on ER visitors waiting times, inappropriate visits can also weaken the service continuity of patients with their primary care physicians, which is especially important for chronically ill patients (Ahuja, Alvarez, and Staats, 2020). Thus, reducing inappropriate visits is crucial to create an efficient and resilient healthcare system.

Given the relevance of the problem, different policies have been proposed to decrease the number of inappropriate ER visits, such as copayment fees, educational campaigns, tele-medicine and GP-referral schemes, (S. R. Morgan et al., 2013; Flores-Mateo et al., 2012; Van den Heede and Van de Voorde, 2016). However, such policies raise organizational and fairness issues.

Since recently, a selection of ER in Italy, have begun to publish online and in real-time the level of congestion of ER departments. The aim of this policy is to improve the transparency of the healthcare system[2], helping patients to choose where to be treated. However, its effect on inappropriate and non-urgent visits is still unclear. According to behavioral theory, information on ER congestion may reduce uncertainty on waiting times, thus reducing the payoff uncertainty of ER visits. Hence, we might expect risk averse patients to visit the ER more often with respect to the case in which waiting times are not known. This is a particularly

[1]In the Italian framework, the following 5 conditions must be satisfied at once for a visit to be considered "inappropriate" (Randazzo, 2024): i) the visit must be identified as a code 4 or 5 on the CTAS scale, a validated triage tool that classifies patients from 1 to 5, with 1 being the most urgent and 5 the least urgent (Afilalo et al., 2004), ii) the visit must be voluntary or referred by the GP, iii) the ER must be visited during daytime hours, iv) the visit must result in a discharge to home or sanitary facility, v) visits following physical traumas are excluded. The orginal definition is “Accessi in codice bianco e verde alla visita medica, con l’esclusione dei traumi; giunti in PS in modo autonomo o inviati dal medico di famiglia, nei giorni feriali e festivi e in orari diurni, con dimissione al domicilio o a strutture ambulatoriali”

[2]In this interview, Simone Magazzini, Emergency department director Azienda USL Toscana Centro, introduces the policy and its aims

relevant scenario, as it has been shown that following a health shock individual levels of risk aversion may increase (Decker and Schmitz, 2016).

In this study, we examine the effect of real-time congestion data on the total number of ER visits, with a focus on inappropriate and appropriate ER visits. First, we model the decision process of a person who, being uncertain of their health status, must decide between visiting the ER, possibly overcrowded, or some territorial health facility. In our model, congestion levels are defined as the ratio between the number of people visiting the ER and the capacity of the ER. Four levels are introduced: Low, Medium, High and Over-Congested. We, then test the predictions of our model using a discrete choice experiment with congestion, social externalities and monetary incentivization. Specifically, we compare the number of ER visits across three conditions in a between-subject design. In condition TC (Treatment Control), congestion levels are unknown. In condition TP (Treatment Policy), which represent the policy adopted by Italian hospitals, congestion levels are fully known. Finally, in condition TS (Treatment Suggestion), we investigate a policy suggestion in which congestion Medium and High are equally probable. The task is repeated using a within-subject design in eight different scenarios. In particular, we introduce five risky posteriors, where the probability to be in a critical health status is known, and three ambiguous posteriors, where the probability to be in a critical health status is "likely", "unlikely" or "unknown".

Our first result suggest that visits to the ER are more frequent in condition TP compared to condition TC. Specifically, we find more inappropriate visits in congestion Low and Medium, a similar number of visits in congestion High, and fewer appropriate visits in congestion Over-congested. These results, coherently with the model's predictions, suggest that when information is available the uncertainty of ER visits decreases, leading risk-averse individuals to choose the ER more frequently. When testing the alternative policy (condition TS), we find similar conclusions for congestion Low and Over-Congested compared to condition TP. However, for intermediate levels of congestion, the total number of visits is similar to the one in condition TC. From a policy

perspective, our policy suggestion combines improvements in the transparency of the healthcare system and few inappropriate visits in intermediate levels of congestion. On the other hand, condition TP may increase the number of inappropriate visits. Therefore, we suggest that, to avoid overcrowding, hospitals should carefully plan the information provided on their website.

Our study also test whether there are differences in ER visits when the decision is taken on behalf of a person unable to decide for themselves. To investigate these differences, the decision task is repeated using the strategy method and telling participants to imagine a situation in which they must choose the health treatment for someone else. We find an overall higher number of visits to the ER in the *for-others* task. In particular, significant differences are found when the uncertainty on the health status is high. We argue that the observed changes in risk attitudes can be attributed to the additional framing condition. Moreover, we argue that decision maker is more risk averse towards the health status of the other person, which prompts them to behave more cautiously and visit the ER more frequently. From a policy perspective, these results are useful to understand the demand of ER visits, and more in general of healthcare services, in those situations in which the decision maker is for example a family member.

Finally, our study investigates the presence of moral consideration towards the use of hospitals' resources. Despite a clear and substantial preference for TH treatments we found little to no correlation with moral considerations towards the inappropriate use of hospital resources. This suggests the need of future research to understand the drivers of the observed preference for TH treatments.

The rest of the work is organized as follows. In Section 3.2 we examine the literature on the demand of ER visits and the policies adopted to reduce inappropriate visits. Moreover, we examine the literature of self-others differences in risky choices. In Section 3.3 we introduce the utility maximization problem. In Section 3.4 we introduce the experimental design and the hypothesis of the study. In Section 3.5 we present descriptive and econometric analysis of the hypothesis. In Section 3.6 we

discuss our results and the relative policy implications.

3.2 Literature review

Following a health shock, a person must often choose between visiting the emergency room (ER) for urgent care, or visit the general practitioner, the closest drug store or a primary care facility, receiving non-urgent medical attention in what we call territorial health facilities (TH). When a non-urgent patient visits an ER, such a visit can be considered inappropriate. Inappropriate visits can contribute significantly to ER overcrowding, which can ultimately increase mortality rates, decrease patient satisfaction and increase violent episodes towards medical staff (Morley et al., 2018). In addition, inappropriate visits can also weaken the continuity of care of patients with their general practitioners (GPs), which is especially important for chronically ill patients (Ahuja, Alvarez, and Staats, 2020). Thus, reducing inappropriate visits is especially important to create a more efficient and resilient healthcare system.

The typical characteristics of an ER, such as the possibility of obtaining more laboratory tests, diagnostic tests and specialist visits (Botelho et al., 2019; M. L. Carret, A. G. Fassa, and Kawachi, 2007) can make patients to prefer ERs over territorial health facitilities (M. L. Carret, A. G. Fassa, and Kawachi, 2007; Coleman, 2001). This preference imbalance may contribute to higher rates of inappropriate visits.

With the aim to reduce such preferences gap and limit inappropriate visits to the ER, many policies were proposed in the literature. Financial incentive schemes such as copayment fees have been one of the main policies used to shape patients' preferences. Selby, Fireman, and Swain (1996) show that copayment fees can reduce the overall number of visits without reducing visits in the "*always an emergency*" category. This policy has also been proven to be robust to age, sex and geographical differences (Selby, Fireman, and Swain, 1996; Lowe, Fu, and Gallia, 2010; Jung et al., 2014; Anantharaman, 2008; D. J. Becker et al., 2013). Despite their success, copayment fees have often been modified to take into consideration patients' socio-economic status and health conditions, thus improving

their fairness. This creates, however, a trade-off with the effectiveness of the scheme, as the cost of visiting the ER will be null for some groups of patients.

In the attempt to reduce inappropriate visits, the literature also studied how patients acquire and process information before the visit happens. For example, patients may overestimate the severity of their symptoms (Lega and Mengoni, 2008; Botelho et al., 2019). Inappropriate visits can also occur in those situations in which patients know their real health status, but are unaware of the correct treatment to adopt. Again, Botelho et al. (2019) shows that after informing participants on the actual severity of their condition, among those that made a inaccurate assessment, only half switched service.

To improve the acquisition and process of information, educational campaigns (Rector, Venus, and Laine, 1999; Scott et al., 2004; Bird et al., 2007; Caplan et al., 2004; DeSalvo et al., 2000; Anantharaman, 2008), telemedicine, telephone triage and consultation (Darnell et al., 1985; Lattimer, 2000; Peter Vedsted, Morten Bondo Christe, 2001; McKinstry et al., 2002; Richards, 2002; Lattimer et al., 1998; S. Sun, Lu, and Rui, 2020), and GP referral schemes (Su et al., 2021) have been proposed. Evidence for educational campaigns is mixed depending on the mode of communication, with studies showing a reduction of inappropriate visits and studies that instead show no significant differences (S. R. Morgan et al., 2013; Flores-Mateo et al., 2012; Van den Heede and Van de Voorde, 2016). Tele-medicine can reduce patients' length of stay at the hospital (S. Sun, Lu, and Rui, 2020) and decrease hospital costs(Lattimer, 2000). On the other hand, it can increase out of hours and emergency attendance in primary care facilities (McKinstry et al., 2002). Finally, under particular conditions, GP referral schemes are shown to be effective in reducing non-urgent visits to emergency departments (Su et al., 2021).

In recent years, a number of Italian hospitals began to publish online and, in real-time, the level of congestion of their ER departments. The characteristics of the policy are region-specific and sometimes specific at the hospital level. This is a result of the high decentralization of the Italian healthcare system (Ricciardi and Tarricone, 2021). The most

common application reports the number of patients waiting to be treated and currently being treated by triage-code. Moreover, in some cases, it is reported the ratio between occupied beds and the total capacity of the hospital or the expected waiting time by triage code. The policy under consideration has many potential advantages over traditional policies. For example, assuming equal access to the website, it is not affected by fairness issues. Moreover, the overall cost to be realized may be lower with respect to educational campaigns and tele-medicine triage, and its effect are potentially immediate. This policy has been initially applied to increase the transparency of the healthcare system[3] expecting a decrease in the number of inappropriate visits. However, its effects are still unclear. According to the behavioral theory, information on ER congestion may reduce uncertainty on waiting times thus reducing the payoff uncertainty of ER visits. Hence, we might expect risk averse patients to visit the ER more often with respect to the case in which waiting times are not known.

3.2.1 Choosing for others

A relevant situation in the context of emergency room visits is represented by the decision taken on behalf of someone else. This is the case in which the other person is for example a child, an elderly person or he/she is momentarily incapacitated. We examine how does the number of visits change when the decision to visit the ER is taken for someone else.

Our study contributes to the literature on self-others differences in risky decision making scenarios by exploring a new application in the context ER visits. Prior research in healthcare has explored related contexts, such as end-of-life decisions by surrogates (Shalowitz, Garrett-Mayer, and Wendler, 2006) and healthcare treatment decisions by professionals, often finding higher risk aversion in the *for-other* scenario (Stone, Yoon-Sun, et al., 2013; Garcia-Retamero and Galesic, 2012; Zikmund-Fisher et al., 2006). Our findings also offer practical insights for specific settings,

[3]Interview to Simone Magazzini, Emergency department director Azienda USL Toscana Centro

such as pediatric ERs, where decisions are frequently made on behalf of others.

To address this topic, we briefly review the literature on decision-making *for-others*, particularly in situations where the decision-maker has no monetary incentives. Early studies suggest no self-others differences in the context of gambles (Teger and Kogan, 1975). A similar result is found in hypothetical scenarios involving risk, regardless of whether feedback on the decision-maker's actions was provided (Stone, Yates, and Caruthers, 2002).

The literature, however, also present studies in which risk aversion increases or decreases in the *for-other* decision. Some studies show decreased risk aversion, such as in multiple price lists (MPL) (Chakravarty et al., 2011) and gambling tasks, where loss aversion is lower (Polman, 2012) Conversely, in investment decisions, individuals exhibit more cautious behavior when deciding for others (Montinari and Rancan, 2013; Eriksen, Kvaløy, and Luzuriaga, 2020). Eriksen, Kvaløy, and Luzuriaga (2020) concludes that risk preferences depend on contextual factors, such as accountability, regret and incentives. Batteux, Ferguson, and Tunney (2019) draw similar conclusions following a meta analysis of studies on self-others differences in risky decision making.

3.3 The model

In this section, we model the decision process of a person who, being uncertain of their health status, must decide between visiting the ER, possibly overcrowded, or some territorial health facility.

Consider one agent characterized by a fixed and unobservable health status $\theta \in \{s, c\}$. The agent can be either in a *stable* or *critical* health condition. The agent has a prior belief about her health status $\mu_0(\theta)$. The prior belief represents the probability assigned by the agent to a given health status prior to any information.

$$\mu_0 : \{s, c\} \to [0, 1]$$

The form of the prior belief over the health state, $\mu_0(\theta)$, depends on the

subjective characteristics of the agent, such as age, sex, chronic illness, and possibly related past experiences.

When the agent receives any information, such as a symptom, they update their prior belief, forming a posterior belief $\mu(\theta)$ about their health status .

$$\mu : \{s, c\} \rightarrow [0, 1]$$

For the sake of this study, we assume the health status to be constant, meaning that it doesn't deteriorate or improve until adequately treated.

Evidence shows that agents, after a health shock, adopt more risk averse choices (Decker and Schmitz, 2016), which may lead to an overestimation of the probability of negative outcomes. For example, when observing a symptom that is associated with a very small probability of critical health status, they could update their belief about their health status in a pessimistically biased way. To capture this bias, we add a behavioral parameter $b > 1$, which multiplies the probability of being in critical health status after the belief update. This parameter captures a specific risk aversion attitude towards health-related risky choices.

Following the belief update, the agent chooses an action $a \in \{ER, TH\}$ to maximize their expected utility based on their biased posterior beliefs. The agent's utility function is state-specific, but since the states are unobservable, the decision will be made to maximize the expected utility.

The action involves going to the Emergency Room (ER) or other Territorial Health services (TH), like a general practitioner or pharmacy. While the quality and the cost of the treatment are reasonably predictable in the case of territorial services (the person can book an appointment with the General Practitioner and receive assistance), when choosing to go to the ER, the agent faces the second level of uncertainty. On top of the uncertainty about her health status, the agent is uncertain about the number of people in line for the *ER* service. Formally, TH is an action that produces a certain outcome, while choosing ER results in opting for a lottery over possible outcomes. The agent is unaware of the actual congestion status $\omega \in \{L, M, H, O\}$ of the ER. She has a prior belief π_0 that assigns a probability to each possible congestion level, specifically Low, Medium, High, and Over-congested. As for the health status, the

agent updates her prior belief after observing new information, forming a posterior π. This second level of uncertainty enters the conjecture of the agent only if they consider the possibility of going to the ER. When the agent chooses ER, they face a probability distribution over the possible levels of crowdedness of the emergency department.[4] It is important to notice that the bias we considered for the uncertainty concerning health status b is not relevant for this second level of uncertainty. Meaning that the belief about the level of crowdedness of the ER is based on the prior and the additional information received by the person, without additional pessimistic bias due to the fear of being in a critical health status.

Summarizing, the total utility of the agent is the sum of three components that depend on the action selected.

The first positive component, $m_{\theta,\omega}(a)$, is the health benefit received thanks to the action a. By assumption, the health benefit received when ER is selected is higher than when TH is selected. This comes from the fact that the hospital treatment is more complete and offers exams and consultancies unavailable elsewhere.

$$m_\theta(ER) > m_\theta(TH)$$

The second component represents the economic opportunity cost, which summarizes the monetary expense of accessing the health service and the time spent to obtain it.

$$t_{\theta,\omega}(a)$$

The third component is the ethical cost, which represents the negative feeling coming from the consumption of a common good that has the property of rivalry.

We consider the agent to be ethical, meaning that they suffer from contributing to the congestion of common scarce resources. In particular, we consider the ethical cost $e\mathbb{1}_{a=ER}$ to be paid if the ER is chosen.

The formalization follows. The utility of choosing $a = ER$, in the

[4]Case of continuous space: $Ec_S(ER) = \int c(\tau)f(\tau)d\tau$, where $f(\tau)$ is the probability density function of the random variable τ.

health state θ and congestion state ω, is

$$u_{\theta,\omega}(ER) = \underbrace{m_{\theta,\omega}(ER)}_{\textit{medical cure}} - \underbrace{t_{\theta,\omega}(ER)}_{\textit{time cost}} - \underbrace{e_{\theta,\omega}}_{\textit{ethical cost}} \tag{3.1}$$

The utility of choosing $a = TH$, in the health state θ and congestion state ω, is

$$u_{\theta,\omega}(TH) = \underbrace{m_{\theta}(TH)}_{\textit{medical cure}} - \underbrace{t_{\theta}(TH)}_{\textit{time cost}} \tag{3.2}$$

The maximization problem of the agent with biased belief $b \cdot \mu$ on her health state and unbiased belief π on the congestion state who can choose to go either to the Emergency Room or to the Territorial Health system is the following.

$$\max_{a \in \{ER,TH\}} \sum_{\omega \in \{L,M,H,O\}} \pi(\omega)[b \cdot \mu(c) u_{c\omega}(a) + (1 - b \cdot \mu(c)) u_{s\omega}(a)] \tag{3.3}$$

From this representation appears clear how the two levels of uncertainty faced by the person are treated differently, and together contribute to the perception of her total utility.

3.3.1 Payoff Assumptions

In what follows, we formalize the assumptions introduced with the literature review in Section 3.2 and the model in Section 3.3.

Assumption 1 A patient of type $\theta = c$ must receive urgent healthcare treatment. Therefore, an agent in a critical health status gets higher utility from visiting the ER: $u_c(ER) > u_c(TH)$. Moreover, we assume that territorial health services can offer first aid procedures to patients in a critical health status but it is best suited for non-urgent patients. Hence, we assume: $0 < u_c(TH) < u_s(Th)$.

Assumption 2 The utility of visiting the ER is decreasing with higher congestion levels for agents of type $\theta = s$ and type $\theta = c$, hence: $u_{s,L}(ER) > u_{s,M}(ER) > u_{s,H}(ER) > u_{s,C}(ER)$ and also $u_{c,L}(ER) > u_{c,M}(ER) >$

$u_{c,H}(ER) > u_{c,C}(ER)$. Following this assumption, higher congestion levels not only decrease the utility of agents of type $\theta = s$ by increasing their waiting times (and ultimately decreasing their satisfaction), but they also decrease the utility of agents of type $\theta = c$ by decreasing the quality of care provided at the ER.

Assumption 2bis. The utility of visiting the TH is not affected by congestion levels, hence $u_c(TH)$ and $u_s(TH)$ are constants.

Assumption 3 Whenever the congestion level is Low, the ER offers the best available healthcare treatment for patients of any type. Hence: $u_{s,L}(ER) > u_s(TH)$ and $u_{c,L}(ER) > u_c(TH)$.

Assumption 4 Financial costs, such as copayment fees, decrease the utility of agents of type $\theta = s$ which visits the ER, hence: $u_{c,\omega}(ER) > u_{s,\omega}(ER)$.

We use these relationships to build a coherent framework for our experimental design.

3.4 Experimental Design

The main outcome of this study is the decision to visit the Emergency Room or the Territorial Health. This decision is examined using an online incentivized discrete choice experiment with congestion and social externalities. The online

We conduct an experiment composed of three between-subject conditions. Specifically, we compare the number of visits to the ER in a condition in which congestion levels are unknown (condition *TC* - Treatment Control), a condition in which congestion levels are fully known (condition *TP* - Treatment Policy) and a condition in which congestion levels are partially known (condition *TS* - Treatment Suggestion). Congestion levels are defined as the ratio between the number of people visiting the ER and the capacity of the ER. We introduce four different congestion

levels, i) Low, ii) Medium, iii) High and iv) Over-Congested. To represent different scenarios, we introduce 8 different posteriors probabilities, which represent the probability to be in a critical health status. In condition *TC* congestion levels are all equally probable, thus participants express one preference per posterior for a total of 8 choices. In condition *TP* congestion levels are known and participants express four preference per posterior, one for each congestion level, for a total of $8 \cdot 4 = 32$ choices. In condition *TS* congestion levels Medium and High are paired and equally probable, thus participants express three preference per posterior for a total of $8 \cdot 3 = 24$ choices.

To represent the situation in which the decision to visit the ER is taken on behalf of a younger or an older relative, we repeat the decision task using the strategy method, thus without incentives, and telling participants to imagine a situation in which they must choose the health treatment for someone else[5].

The concept of social externalities is introduced by framing the study. Participants are told to imagine a situation in which they are experiencing some health symptoms. Moreover, they are told to consider information like the quality of the treatment provided at the ER and how they are influenced by the congestion level of the ER[6]. The decision to frame the study to introduce social externalities, follows the work of Böhm, Betsch, and Korn (2016), which show that in the context of vaccines a framed condition may elicit moral considerations, increasing vaccination uptake.

Figure 5 shows a visual summary of the experimental procedure.

[5]Additional framing: "You should imagine a situation where someone you know - like a child or an elderly person - is experiencing some health symptoms and need medical attention. This person however, cannot take a decision for themselves, because they are either too young, too old, temporarily incapacitated or unable to make an independent decision. Your must decide on their behalf whether they should visit the Emergency Room the Territorial Health. You will receive information about their health status presented in the same way as the one you received before"

[6]The complete set of instructions is available in the online supplementary material.

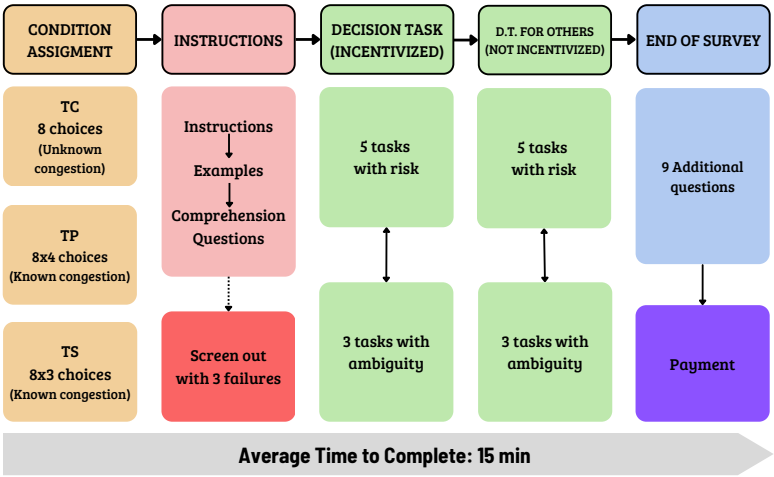


Figure 5: Summary of the study structure in this study.

Note. In the "condition assignment" phase, each participant is assigned either to condition TC, TP or TS. In the "instructions" phase, participants receive instructions on the task and are provided with an example to get used with the task. Finally, they are given a set of comprehension questions for which a correct answer is necessary to take part in the study. In the "decision task" phase, participants express their preference for visiting the ER for any posterior. In the "decision task for others" phase, the previous phase is repeated under the new framing. In the "additional question" phase, participants are asked nine additional questions. In the payment phase, participants receive a participation fee and a bonus payment depending on the outcome of the task.

3.4.1 The task

The ER is modeled as a good with the properties of *rivalry* and *non-excludability*. The property of *non-excludability* allows any person with health symptoms to receive medical attention at the ER. This is a property shared by many emergency healthcare systems in Europe. To represent this property, in our study for any posterior there will be the possibility to visit the ER. We also introduce the property of *rivalry*, assuming that the ER has limited resources and it cannot provide adequate healthcare treatment to the whole population of patients at once. This concept is introduced with congestion levels modeled in Assumption 2 in Section 3.3. The alternative to the ER is the decision to visit any source of primary or territorial health (TH). These include visits to general practitioners, primary care facilities, and drug stores. We assume healthcare services in the TH to be delivered through a booking system or in a timely way, hence the payoffs are not affected by different congestion levels.

Figure 6 is the payoff matrix shown to participants. Each row represents a different congestion levels (from low to congested). The first two columns are the bonus payments that participants can receive if they choose to visit the ER. The last two columns are the bonus payments that participants can receive if they choose to visit the TH. Individual returns, are chosen to represent the trade off between ER's health benefits, financial costs and waiting costs reviewed in Section 3.2 and modeled in Equation 3.1 and Equation 3.2. In particular, the relationships between individual returns are chosen following the assumption in Section 3.3.1.

The choice to keep in our experimental setup both the framing and the monetary incentives is motivated by the following reasons. Framing the study was necessary to examine whether participants shows evidence of a preference for TH treatments. Monetary incentives, on the other hand, allows to create a set of predictions and compare it with the observed behavior of participants in our study. These information can be used to make inference on participants' risk preference and attitudes towards ER use.

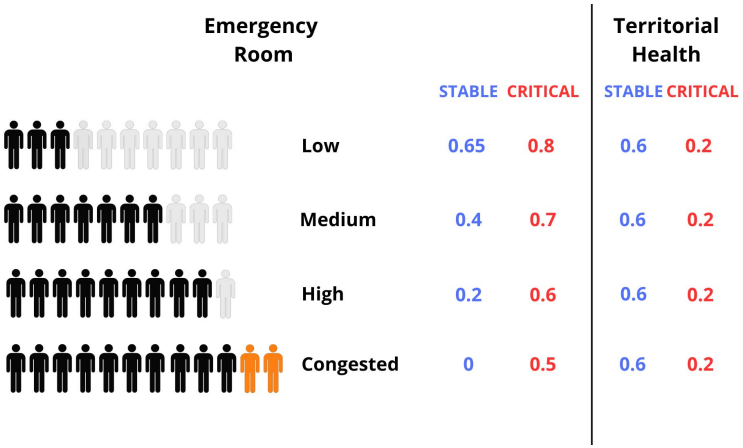


Figure 6: Payoff matrix shown to participants.

Note. In condition *TC*, participants express 1 preference for every posterior. In condition *TP*, participants express 4 preferences for every posterior, i.e. 1 for any congestion level and any posterior. In condition *TS*, participants express 3 preferences for every posterior, as the Average and High congestion levels are considered equally probable. Payoffs are in GBP.

Table 7: Posteriors probabilities

Risky		**Ambiguous**	
R0	[0 , 19]	A1	Likely
R20	[20, 39]	A2	Unlikely
R40	[40, 59]	A3	Unknown
R60	[60, 79]		
R80	[80,100]		

Note: Each posterior represent the chance of being of type $\theta = c$, multiplied by 100 for readability. In decisions from R0 to R80, participants are shown one probability drawn at random for each range of probability. Probabilities are known as in risky posteriors. In decision from A1 to A3 probabilities are not known. These posteriors are equivalent to $\mu = 80\%$, $\mu = 20\%$ and $\mu = 50\%$, respectively

To represent different scenarios, we introduce different probabilities of being in a critical health status μ, as shown in Table 7. First, we introduce a series of posteriors in which the probability is known, namely posteriors from R0 to R80. Participants are shown one posterior for each interval. At the beginning of the experiment each posterior is drawn at random from a Uniform distribution with extremes corresponding to each interval. Hence, each participant likely experienced a different set of posteriors. This method allows to have a complete and homogeneous representation of the decision to visit the ER over the posteriors from 0 to 100 (Appendix B.4). In addition, to represent the uncertainty that characterizes health statuses, posterior probabilities are also represented using three different ambiguous terms, "Unlikely", "Unknown" and "Likely". As previously done in Toccafondi, Di Paolo, and Di Guida (2025), these posteriors are equivalent to $\mu = 20\%$, $\mu = 50\%$ and $\mu = 80\%$ respectively, but unknown to participants.

3.4.2 Additional Questions

After the task is completed, participants are given nine additional questions to fulfill. Table 8 shows the text presented to participants for each

Table 8: Additional Questions

Item	Text
AQ1	You are given a series of clinical profiles. For each clinical profile we ask you to indicate what is its severity in your opinion. Severity equal to 1 corresponds to a life-threatening situation, while severity equal to 5 corresponds to a minor situation which will unlikely worsen if the treatment is delayed.
AQ2	You are given a series of clinical profiles. For each clinical profile we ask you to indicate what would you choose between visiting the ER or the TH.
AQ3	Have you ever considered how your visit to the ER might impact the the resources of the hospital and staff workload?
AQ4	Have you ever decided NOT to visit the ER because you were concerned about using resources that could be better allocated to someone in greater need?
AQ5	Have you ever chosen NOT to visit the ER because you were concerned about the risk of infection from other patients?
AQ6	Have you ever chosen NOT to visit the ER because you were concerned about the risk of infecting other patients?
AQ7	How often have you visited the Emergency room in the last 5 years?
AQ8	What is the quality of the alternatives to the emergency room in your area?
AQ9	Please tell us, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "extremely willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale.

additional question[7].

Following the work of Botelho et al. (2019), participants are given a set of symptoms corresponding to an unknwon clinical profile. In AQ1, for each set of symptoms shown in Table 9 participants are asked to indicate the corresponding severity on a scale from 1 (most urgent) to 5 (least urgent). In AQ2, for each set of symptoms, participants must indicate what would they choose between visiting the ER and the TH. These questions evaluates participants' ability to indicate symptom's severity and their ability to choose the appropriate health treatment. It is useful to correlate this information with risk aversion levels and the preference for using the ER.

Questions from AQ3 to AQ6 examine participants attitude towards

[7] All of the additional questions had the "Prefer not to say" alternative available.

Table 9: Symptoms, Clinical Profiles, and Severity

Set	Symptoms Presented	Clinical Profile and Severity
1	Diarrhea, history of occasional blood in feces, abdominal pain	Diarrhea (Code: 3 - ER)
2	Muscle pain, shivering, runny nose, cough	Flu (Code: 4 - TH)
3	Urinary problems, pain and burning sensation when urinating, blood in urine	Urinary infection (Code: 4 - TH)
4	Insect bite, severe pain, shortness of breath	Insect bite/anaphylactic shock (Code: 2 - ER)
5	Back pain, no trauma, no fever, walking difficulty	Back pain (Code: 4 - TH)

ER use. These questions are on a 5-point Likert scale[8]. Then, questions AQ7 and AQ8 examine participants' past experience with ER visits. These are also measured using a 5-point Likert scale[9]. Finally, participants risk preference is measured using the item proposed by Falk, A. Becker, et al. (2018).

A note on AQ4. In the pilot study, we intended to capture participants' attitudes toward the use of hospital resources using AQ3. However, we realized afterwards that the relationship between this attitude and ER visits was not evident. To address this, we introduced AQ4, which better captures this aspect. As a result, AQ4 is available only for participant recruited in the main study.

The results of the additional question are presented in Section B.3.

3.4.3 Hypothesis

We use this section to summarize the objectives of this study and present the set of hypothesis registered in the pre-registration phase of the work.

[8]Scale used: Never, Rarely, Sometimes, Most of the times and Always

[9]Scale of AQ7: Never(0 times), Rarely(1 or 2 times), Sometimes (between 3 and 5 times), Frequently (between 6 and 10 times) and Very Frequently (more than 10 times). Scale of AQ8: Terrible, Poor, Average, Good, Excellent

H1. Policy effect. *As a consequence of the lower payoff uncertainty in conditions TP and TS, the total number of ER visits is expected not to be lower with respect to those in condition TC. Furthermore, we expect more frequent ER visits in congestion Low and Medium, and less frequent ER visits in congestion High and Over-Congested.*

The first claim in this hypothesis is that the payoff uncertainty is lower in condition TP with respect to condition TC. This can be noticed following an analysis of Equation 3.3. Agents in condition TC has two source of uncertainty: $\pi(\omega)$, which is the uncertainty on the congestion, and $\mu(c)$, which is the uncertainty on the health status. However, by construction, agents in condition TP choose to visit the ER is a situation in which there is no uncertainty on the congestion of the ER. Hence, our first claim. A similar reasoning can be made to prove that the uncertainty in condition TS is higher with respect to condition TP, but lower than in condition TC. The second claim is that, as a consequence of the lower uncertainty in TP and TS, ER visits are expected not to be lower with respect to those in condition TC. This claim can be proved following the prediction of a RN (risk neutral) and a RA (risk averse) agents provided in Appendix B.2. First, consider intermediate levels of congestion. We argue that, in these cases, the healthcare system can benefit the most from a reduction of inappropriate visits. In fact, in congestion level Low the additional burden of a patient is low, and in congestion level Over-Congested non-urgent patients avoid visiting the ER. In congestion Medium and High, however, the ER might suddenly experience overcrowding. Thus inappropriate visits should be as low as possible. In these congestion, it can be observed that ER visits in conditions TP and TS are expected to be more frequent with respect to those in condition TC for a risk averse agent. Moreover, notice that the expected behavior of a RN agent is similar across conditions TC, TP (considering the average) and TS. This contributes to the internal validity of our experimental design. In fact, it can be reasonably assumed that a risk neutral agent should be indifferent to the policy. Finally, our last claim regarding the behavior in each congestion, follows again from the analysis of the behavior of a risk averse agent provided in Appendix B.2.

H2. Decision for others. *The framing effect in the decision task for someone else is expected to induce higher levels of risk aversion, thus increasing visits to the ER in the for-other condition.*

The first claim of this second hypothesis is that the framing effect can induce higher levels of risk aversion. In particular, despite the literature offers mixed evidence on this topic, we expect that in the *for-others* task, agents show higher concern towards the health status of the other person. The second claim is that, following higher risk aversion, ER visits will be more frequent. In particular, we expect the agent to behave more cautiously, choosing to visit the ER more frequently in the *for-other* condition, which yields higher payoffs for a patient in a critical health status (Assumption 1 in Section 3.3). We argue that this effect can be the consequence of reputational effects or loss aversion effects. However, this study is not designed to examine the source of an eventual difference and it is left for future studies.

H3. Ethical cost. *The framing induces a preference for Territorial Health treatments, which is explained with moral considerations about the use of hospital resources and the concern of infecting other patients when visiting the ER.*

The first claim in this hypothesis is that framing induces a preference for Territorial Health treatments. In our setup, visiting the ER in congestion Low and in conditions TP and TS is payoff-dominant with respect to TH. In particular, the higher the posterior μ and the higher is the benefit to visit the ER. However, we expect a series of factors to induce patients to avoid inappropriate visits, choosing to visit TH over ER. To address the source of this preference, we study the relationship of these choices with AQ4 and AQ6, which examine participants' moral consideration towards the use of ER resources.

3.4.4 Experimental Procedure

We conduct a pre-registered online experiment (OSF Pre-Registration Link). Participants are recruited through the Prolific platform (Palan and Schitter, 2018), and the experimental interface is developed using

Qualtrics (Qualtrics, 2005). The experiment is conducted online to recreate the setting in which the policy develops. The experimental setup also allows to observe the behaviour of participants choosing the alternative forms of care, avoiding a censoring bias that could have affected a retrospective study on this topic.

Eligible participants for our study satisfy the following conditions: i) are resident in Austria, Belgium, Bulgaria, Czech Republic, Croatia, Estonia, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Poland, Portugal, Romania, Slovakia, Slovenia or Spain, ii) are at least 18 years old and at most 80 years old, iii) with an approval rate on the prolific platform of at least 50% and iv) fluent in english. Criterion i) includes countries that are part of the European Union and in which emergency rooms have the property of non-excludability. This condition allows to have a sample of participants familiar with the healthcare system described in the experiment. Figure B.5 shows the country of residence of the participants in our study. Criterion ii) considers participants relevant to the topics of our study. Criterion iii) considers participants familiar with Prolific procedures and experiments. Criterion iv) prevents the introduction of translation biases, which could be introduced by conducting the experiment in many different languages.

When participants enter the study, they are randomly assigned to one of three conditions, TC, TP or TS (Condition Assignment phase in Figure 5). After giving the consent to the consent form and reading the instructions, participants are given an example of the task to learn how it works. The example is also an attention check. Then participants have three attempts to correctly answer three comprehension questions all at once. Participants who fail are asked to return their submission (Instructions phase in Figure 5).

After the comprehension questions phase, participants are brought to the *for-self* decision Task phase (see Figure 5). Here they are presented, randomly, a page including risky tasks and a page including ambiguous tasks. All of the risky and ambiguous tasks are also randomly presented respectively in the same page. Figures B.6, B.7 and B.8 show a series of screenshots of the experimental interface in the three conditions. This

part of the experiment is incentivised. Each page contains an attention check in which participants must either click all the available boxes or none of them.

Following the *for-self* task, participants enters the *for-others* task phase, in which they must repeat the task with a new framing. The choice is in fact for "someone you know - like a child or an elderly person - (who) is experiencing some health symptoms and need medical attention" (Decision Task For-Others in Figure 5). There are also two additional differences with repsect to the *for-self* task. First, the bonus payment is removed to avoid the incentive to interfere with the framing of this task. Second, there are no attention checks to reduce the time taken to complete the study.

After the completion of these tasks, participants are given nine additional questions to answer (Section 3.4.2).

At the end of the experiment, a standard payment procedure using a randomised incentive system is used. One posterior, one congestion level and one decision among those in the "Decision Task" phase are chosen at random. The participant type is computed and individual payoffs, showed in Figure 6, are granted accordingly. On average particpants took 16.5 minutes to complete the experiment and were paid an average of 1.44 pounds for the show-up fee and 0.54 pounds for the bonus payment.

A total of 443 participants have been recruited through the Prolific platform. Among these, 47 showed evidence of multiple submissions and were excluded from the study. Among the remaining, 100 participants either voluntarily withdrew from the experiment or failed three consecutive times the comprehension checks and were asked to return the submission. Thus, we collected data on 296 participants (53 in the pilot study and 243 in the full study), which correspond to the pre-registered sample size. Table 10 shows relevant summary statistics for the collected sample[10]. Variable with the question mark are dummy variables. Most of the recruited participants are around 33 years old, male and employed. The young age explains the low percentage of participants with

[10]Covariates were available for 294 participants.

Table 10: Summary Statistics.

	TC	TP	TS	Total
Age	34.32	32.02	32.13	32.84
	1.18	0.85	0.80	0.56
Female?	0.42	0.35	0.37	0.38
	0.05	0.05	0.05	0.03
Student?	0.35	0.20	0.32	0.29
	0.05	0.04	0.05	0.03
Children?	0.31	0.20	0.17	0.23
	0.05	0.04	0.04	0.02
Caring Responsabilities?	0.21	0.18	0.23	0.21
	0.04	0.04	0.04	0.02
Chronic Illness?	0.08	0.04	0.09	0.07
	0.03	0.02	0.03	0.01
Employed?	0.92	0.92	0.91	0.92
	0.03	0.03	0.03	0.02
Risk Preference	4.55	4.93	5.29	4.92
	0.24	0.24	0.20	0.13
N. Attention checks	2.90	1.87	2.86	2.55
	0.04	0.04	0.04	0.04
Time taken in minutes	14.45	17.40	15.61	15.79
	0.61	0.69	0.61	0.37
Observations	100	97	99	296

Note. Average values and the relative standard errors are reported. Variables with the question mark are dummy variables. The maximum number of Attention checks is 3.

Table 11: Sample balance tests

	TC/TP	TC/TS	TP/TS	K-Wallis
Age	0.401	0.537	0.740	0.671
Female?	0.400	0.573	0.779	0.593
Student?	0.062*	0.744	0.124	0.04**
Children?	0.167	0.092*	0.770	0.046**
Caring Responsibilities?	0.674	0.786	0.490	0.611
Chronic Illness?	0.638	0.894	0.548	0.365
Employed?	0.976	0.894	0.919	0.959
Risk Preference	0.320	0.039**	0.360	0.129
N. Attention checks	<0.001***	0.403	<0.001***	<0.001***
Time taken in minutes	<0.001***	0.077*	0.052*	0.001***

Note. This table report a series of sample balance tests. In columns TC/TP, TC/TS and TP/TS, we report a series of Wilcoxon rank-sum tests for unmatched data. In column K-Wallis we report the Kruskal–Wallis equality-of-populations rank tests across conditions TC, TP and TS. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01

chronic conditions (7%), children (23%) and informal caring responsibilities (21%). The variables *Children?* and *Caring Responsabilities?* are used in the analysis of Hypothesis 2. Table 11 reports a series of sample balance tests, showing that the sample is balanced in variables *Age*, *Female?*, *Caring Responsabilities?*, *Chronic Illness?*, *Employed?*, *Risk Preferences*. Variables *Student?* and *Children?* are balanced considering p-values equal to 0.005, which is a Bonferroni corrected p-value with 10 comparisons. Table 11 shows a significative lower number of passed attention checks in condition TP and higher time taken to complete the study in conditions TP and TS with respect to the remaining conditions.

3.5 Results

To gather useful insights from participants' behavior in our study we first provide a descriptive analysis of our results. Then, we examine each hypothesis separately using non-parametric tests and regression analysis. For our analysis we consider the choice to visit the ER for par-

ticipant i, condition t, posterior μ and congestion level ω represented by $ER_self_{i,t,\mu,\omega}$. The variable takes value 1 if the participant chose to visit the ER and 0 otherwise. We collected a total of 8 observations for participants in condition TC, 32 for participants in condition TP and 24 for participants in condition TS. Throughout the following analysis risky posteriors and ambiguous posteriors will be treated separately.

3.5.1 Descriptive analysis

Figure 7 shows participants behavior in each combination of conditions, posteriors and congestion levels. In particular, it shows the variable in Equation 3.4, which takes values between 0 and 1, and represents the probability to observe a visit at the ER in condition t, posterior μ and congestion level ω.

$$ER_self_{t,\mu,\omega} = \sum_{i=1}^{n_t} \frac{ER_self_{i,t,\mu,\omega}}{n_t} \tag{3.4}$$

In Equation 3.4 n_t is the number of observations in condition t for an agent i. Higher values of the variable on the y-axis can be interpreted as more frequent visits to the ER. The analysis of this variable allows to make a series of relevant considerations on participants' behavior in our study.

First, it can be observed in Figure 7 that the probability to visit the ER is increasing with the probability to be in a critical health condition for both risky and ambiguous posteriors. This is consistent with Assumption 1 in Section 3.3.1. Considering only condition TC, one can also observe that the marginal increase between posteriors R20 and R40 is greater than the marginal increase registered for posteriors R40 and R60. This is again consistent with our framework. As shown in Appendix B.2 risk neutral and risk averse agents switch from TH treatments to ER treatments in R40. The same effect is not found among ambiguous posteriors, where the marginal increases are similar between the three posteriors. This could be the result of participants' ambiguity aversion, which decreases ER visits for unknown probabilities of being in a critical health condition. Figure 7 also shows that the probability to visit

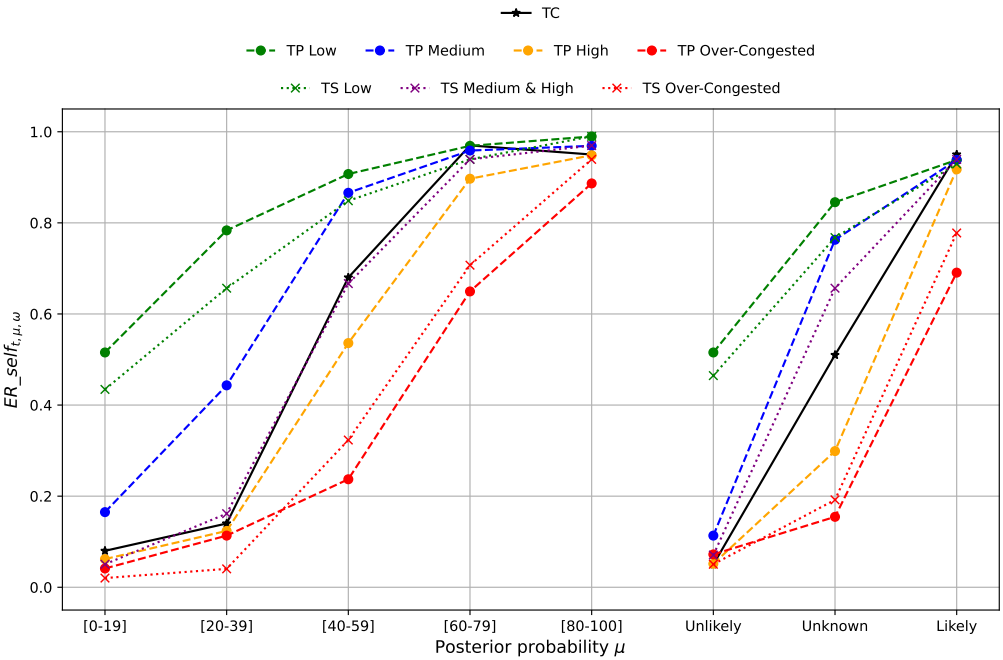


Figure 7: Average ER visits by condition, posterior and congestion.

the ER is decreasing with higher levels of congestion. This is consistent with Assumption 2. Moreover, one can observe that for posteriors R0 and R20 the values in condition TC (in black) are below 0.20 and similar to the values in TS-Medium&High (in purple), TP-High (in orange) and TP-Over-Congested (in red). However, for posteriors R40 and R60, the values in condition TC are above 0.60 and higher than those in TP-Over-Congested and TS-Over-Congested. Moreover, one can observe that the probability to visit the ER in posteriors R60, R80 and A-Likely and conditions TP and TS are around 0.65, 0.90 and 0.7 respectively, lower than the respective values in condition TC. This shows that, in terms of desired outcomes, condition TC may be more effective than conditions TP and TS. Participants in condition TC show in fact few inappropriate visits and many appropriate visits. Comparing conditions TP and TS, it can also be observed that the values in TS-Medium&High are similar to those in TP-High and smaller than those in TP-Medium. This shows that our suggested policy may combine the effectiveness of the control condition and the transparency of the policy under examination.

Finally, contrary to economic expectations, it can be observed that participants in TP-Low and TS-Low exhibit a preference for TH, likely induced by the framing. The average number of visits in these situation is in fact consistently different from 1. We examine this effect in details in Section 3.5.4.

In what follows we examine the hypothesis of our study.

3.5.2 Result 1: Policy and Suggestion Effects

The effect of TP and TS on TC is estimated using the linear random effect model in Equation 3.5.

$$ER_self_{i,t,\mu,\omega} = \beta_1 + \beta_2 TP_i + \beta_3 TS_i + \mathbf{X}_i^{'}\beta + \alpha_i + \alpha_\mu + \alpha_\omega + \epsilon_{i,t,\mu,\omega} \quad (3.5)$$

In Equation 3.5 the variable $ER_self_{i,t,\mu,\omega}$ is the dependent variable; β_1 is a constant which represent the average number of ER visits in our study; β_2 is the overall effect of condition TP with respect to condition

TC; β_3 is the overall effect of condition TS with respect to condition TC; TP_i and TS_i are dummy variables indicating the condition assignment for individual i; $\mathbf{X}_i^{'}$ is a matrix of individual-specific covariates and β the corresponding marginal effects; α_i is the random effect component, i.e. the average number of visits to the ER of individual i; α_μ are posterior specific fixed-effects, i.e. the average number of visits observed for posterior μ; α_ω are fixed effects for the each congestion level, i.e. the average number of visits to the ER for each congestion level ω; $\epsilon_{i,t,\mu,\omega}$ is the error component, clustered at the individual level. Table 12 reports the results of the regression analysis. Models (1) and (2) are estimated on risky posteriors while models (3) and (4) are estimated on ambiguous posteriors. Also note that, models (2) and (4) are estimated with the inclusion of the covariates in Table 10 and additional questions 7[11] and 8[12].

Table 12 shows an overall positive and significant effect of the condition TP for risky posteriors. This effect is robust to the inclusion of the covariates in Table 10 and it increases from 3.9% to 10.1%. If instead ambiguous posteriors are considered, the effect is again positive but not significant. On the other hand, Table 12 shows an overall positive but not significant effect of the condition TS for risky posteriors in models (1) and (2). Considering ambiguous posteriors, the effect is again positive but not significant in model (3) but positive and significant in model (4) with the inclusion of the covariates.

To gather further insights on the effect of conditions TP and TS on condition TC, Table B.5 examine these effect for any given combination of posteriors and congestion levels, without the inclusion of the covariates and using standard errors robust to heteroskedasticity[13]. Considering the comparison between TP and TC, Table B.5 shows significant and positive effects for TP-Low and TP-Medium, and for posteriors smaller than 59% (R0,R20, R40, A-Unlikely and A-Unknown). On the other hand, in TP-High there are a majority of null effects with the exception of the coefficients for R40, R60 and A-Unknwon which are negative and significant.

[11] How often have you visited the Emergency room in the last 5 years?
[12] What is the quality of the alternatives to the emergency room in your area?
[13] Only 1 observations per id is available with this analysis.

Table 12: Linear regressions on $ER_self_{i,t,\mu,\omega}$.

	Risky		Ambiguous	
	(1)	(2)	(3)	(4)
TP	0.039** (0.019)	0.101*** (0.032)	0.022 (0.024)	0.055 (0.041)
TS	0.010 (0.019)	0.017 (0.019)	0.040 (0.025)	0.050** (0.025)
Age		0.001 (0.001)		-0.000 (0.001)
Female?		0.003 (0.017)		-0.021 (0.020)
Student?		0.012 (0.014)		-0.003 (0.017)
Children?		0.009 (0.022)		0.042 (0.030)
Caring Responsabilities?		0.014 (0.015)		-0.019 (0.015)
Chronic Illness?		0.048 (0.030)		0.067*** (0.022)
Employed?		-0.051* (0.030)		0.031 (0.031)
Risk Preferences		0.001 (0.003)		-0.007* (0.004)
Attention Checks		0.049** (0.023)		0.024 (0.030)
Previous experiences		0.015* (0.009)		0.001 (0.010)
Quality of alternatives		0.013* (0.008)		-0.005 (0.010)
Constant	0.564*** (0.014)	0.329*** (0.086)	0.503*** (0.019)	0.465*** (0.111)
N	5920	5880	3552	3528

Note. The baseline treatment is condition TC. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the individual level and reported in parentheses. Estimates are computed in the long format; hence, there are 32 observations per participant, one for each combination of posterior and congestion level. Models (1) and (2) consider risky posteriors. Models (3) and (4) consider ambiguous posteriors. Models (1) and (3) are estimated without further controls. Models (2) and (4) are estimated with the reported controls.

Finally, TP-Over-Congested shows negative and highly significant estimates in R40, R60, A-Unlikely and A-Likely. These results reflect what has been observed in Figure 7: the policy increases the number of inappropriate visits in TP-Low and TP-Medium, it produces fewer visits in TP-High when the uncertainty on the health status is high, and reduces appropriate visits in TP-Over-Congested. These results are robust to the inclusion of the covariates, as only marginal differences can be found between Table B.5 and Table B.6. The only exception is represented by TP-High, which in Table B.6 shows a majority of null effects.

Now, consider the comparison between TS and TC. Table B.5 shows again a majority of positive effects in TS-Low and a majority of negative effects in TS-Over-Congested. For intermediate levels of congestion however there are no significant effects, indicating that participants in TS-M&H behave as participants in TC. Our results are robust to the inclusion of the covariates, as only marginal differences can be found between Table B.5 and Table B.6.

To conclude the analysis of the effect of TP and TS on TC, Table B.7 presents the estimates of a linear regression which considers only congestion levels Medium and High. The effect of TP is again positive in all the models, but significant only in model (2). It can also be noted that the effect is similar to that of Table 12. This is further evidence that for intermediate levels of congestion, conditions TP may increase the number of ER visits. The effect of TS is instead positive and significant only for ambiguous posteriors, in models (3) and (4).

Result 1. Consistently with H1, the policy represented in TP increase the overall number of visits with respect to condition TC. On the other hand, the TS alternative do not. In particular we find: i) more inappropriate visits in TP-Low, TP-Medium and TS-Low; ii) fewer visits in TP-High when the uncertainty on the health status is high (even though it is not robust to the inclusion of covariates); iii) a null effect in TS-Medium&High; and iv) less appropriate visits in TP-Congested and TS-Over-Congested.

We finally examine the difference between conditions TP and TS. The overall effects are presented in Table 13. This regression shows a negative and significant effect in model (2) which consider risky posteriors and the inclusion of the covariates. This result indicates that the overall number of ER visits is smaller in TS, if the covariates are considered.

To examine the source of such difference, Table B.8 shows the estimates of a series of linear regressions where ER visits in conditions TP and TS are compared matching the respective congestion levels. Table B.9 shows similar estimates with the inclusion of the covariates in the analysis. This approach is useful also to check whether the effect for congestion levels Low and Over-Congested are null, being the choices equivalent between the two conditions. Figure 7 showed different values of ER visits between the two Low congestion levels (green lines). As expected, model (1) in Table B.8 shows negative effects. While the overall value is significant in congestion Low, the only significant effect is found for R20. Moreover, the overall effect in model (1) is significant only at $p<0.1$ if the covariates are included. Instead, model (4) always shows non-significant effects. Therefore, we conclude that there are not significant differences between TP and TS in congestion levels Low and Over-Congested. We finally compare the intermediate levels of congestion. Our results suggest a negative and significant effect of TS with respect to TP in congestion level Medium, as shown from the values in R0 R20 and R40. Contrary to our expectations, the same effect is not found for A-Unlikely. For congestion level High, the overall effect appears to be not different from 0. Posterior A-Unknwon is the only exception, as it shows a positive and highly significant value. The overall effect for congestion High in Table B.8 appears to be significant, however it is not robust to the inclusion of covariates.

Result 1bis. Comparing conditions TP and TS, we find: i) similar visits in congestion Low and Over-Congested, contributing to the internal validity of our study; ii) for low probabilities of being in a critical health condition, ER visits in TS are similar to TP-High and significantly smaller than TP-Medium

Table 13: Linear regressions on $ER_self_{i,\mu,\omega}$ with baseline TP.

	Risky		Ambiguous	
	(1)	(2)	(3)	(4)
TS	-0.029 (0.019)	-0.113*** (0.035)	0.018 (0.022)	-0.018 (0.045)
Age		0.001 (0.001)		0.000 (0.002)
Female?		0.006 (0.020)		-0.020 (0.023)
Student?		-0.000 (0.017)		-0.022 (0.018)
Children?		0.003 (0.028)		0.056 (0.037)
Caring Responsabilities?		0.003 (0.012)		-0.003 (0.015)
Chronic Illness?		0.062** (0.031)		0.057** (0.025)
Employed?		-0.047 (0.033)		-0.014 (0.038)
Risk Preferences		0.007 (0.004)		-0.005 (0.005)
Attention Checks		0.078*** (0.025)		0.038 (0.036)
Previous experiences		0.020* (0.011)		0.002 (0.012)
Quality of alternatives		0.013 (0.009)		-0.018 (0.012)
Constant	0.603*** (0.014)	0.354*** (0.072)	0.525*** (0.015)	0.543*** (0.096)
N	3920	3880	2352	2328

Note. The baseline treatment is condition TP. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the individual level and reported in parentheses. Estimates are computed in the long format, hence, there are 32 observations per participant, one for each combination of posterior and congestion level. Models (1) and (2) consider risky posteriors. Models (3) and (4) consider ambiguous posteriors. Models (1) and (3) are estimated without further controls. Models (2) and (4) are estimated with the reported controls.

3.5.3 Result 2: More ER visits in the *for-others* task

To conduct the analysis of hypothesis 2, we measure the difference in ER visits between the *for-self* and the *for-others* tasks obtaining the variable in Equation 3.6. Given the within-subject nature of this variable, it considers individual differences in condition assignment, posteriors probabilities, congestion levels and any other individual characteristic.

$$ER_diff_{i,t,\mu,\omega} = ER_others_{i,t,\mu,\omega} - ER_self_{i,t,\mu,\omega} \tag{3.6}$$

Figure 8 shows the values of the variable $ER_diff_{t,\mu,\omega} = E_i[ER_diff_{i,t,\mu,\omega}]$ for each condition t, posterior μ and congestion level ω[14]. One can observe that the majority of the values are positive or very close to 0. This is a preliminary evidence that participants chose to visit the ER more often in the *for-others* task. Moreover, it can be observed that the values in posteriors where uncertainty about the health status is high, i.e. R20, R40 and A-Unknown, are higher than the respective values where the uncertainty is low.

Table 14 reports a series Wilcoxon signed rank test for paired data which examine the significance of the differences in Figure 8. A test is conducted: i) for any combination of conditions, posteriors an congestion levels; ii) grouping by posteriors; iii) grouping by conditions and congestion levels; iv) and overall. First, it can be observed that all of the significant differences are registered among positive values. Moreover most of the significant differences are registered for posteriors R20, R40 and A-Unknwon. Grouping by conditions and congestion levels, we find positive and significant differences for posteriors R20, R40, A-Unlikely and A-Unknwon. Grouping by posteriors, we observe that participants visited more often the ER in condition TC, TS and TP-High. Finally, we observe that the overall effect is positive and significant.

[14]By definition, the variable $ER_diff_{i,t,\mu,\omega}$ has the same length of both $ER_others_{i,t,\mu,\omega}$ and $ER_self_{i,t,\mu,\omega}$. Each observation is equal to: i) 1 if the participant chose ER in the *for-other* condition and TH in the *for-self* condition, ii) 0 if the participant chose similarly in the two conditions, iii) -1 if the participant chose ER in the *for-self* condition and TH in the *for-other* condition.

Result 2. Consistently with hypothesis 2, ER visits in the *for-others* condition are more frequent. In particular, the effect is significant for posteriors with a high uncertainty about the health status.

Table 14: Wilcoxon signed rank test on $ER_diff_{\mu,\omega}$ by condition t.

t_ω	R0	R20	R40	R60	R80	Unlikely	Unknwon	Likely	Total
TC	0.857	0.049**	0.006***	0.847	0.702	0.576	0.003***	0.850	0.001***
TP_L	0.872	1.000	0.708	0.845	0.841	0.185	0.723	0.703	0.565
TP_M	0.463	0.515	0.211	0.698	0.692	0.471	0.494	0.703	0.176
TP_H	0.850	0.164	0.200	0.262	0.556	0.850	0.005***	0.856	0.008***
TP_OC	0.698	0.585	0.017**	0.611	0.571	0.703	0.103	0.872	0.06*
TS_L	0.436	0.246	0.283	0.852	1.000	0.037**	0.017**	0.341	0.002***
TS_M&H	0.575	0.028**	0.036**	0.852	0.846	0.392	0.223	0.711	0.004***
TS_C	0.350	0.207	0.161	0.064*	0.706	0.575	0.148	0.246	0.001***
Total	0.755	<0.001***	<0.001***	0.129	0.286	0.011**	<0.001***	0.202	<0.001***

Note. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Column totals examine the effect across condition considering each posterior separately. Row totals examine the effect across posteriors considering each condition and congestion level separately. The overall effect considers all the available observations.

3.5.4 Result 3: Preference for TH

In this section, we examine whether the framing induces a preference for TH treatments and if such preference correlates with moral considerations about the use of hospitals' resources.

For this analysis, ER visits in conditions TP and TS and congestion level Low are considered. These choices are considered because standard economic theory suggest that participants in these scenarios should visit the ER regardless of their posterior and their risk preference.

As evident from Figure 7, a relevant fraction of participants choose to visit the TH in TP-Low and TS-Low. This intuition is confirmed by a Wilcoxon signed rank test which shows that the observed effect is significant at $\alpha < 0.001$. Therefore, our fist conclusion is that the framing is

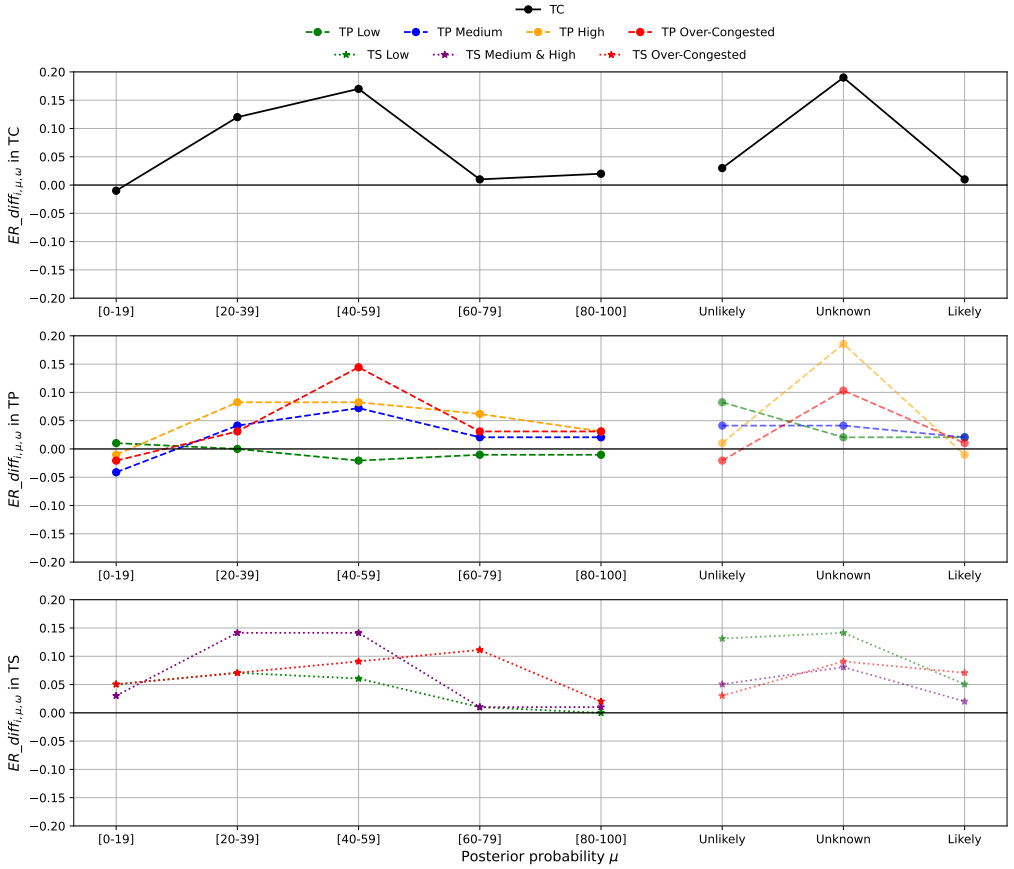


Figure 8: Average $ER_diff_{\mu,\omega}$ by condition t.

inducing a preference for TH.

To examine the source of this preference, we examine the correlation between participants choices in these scenarios and participants' preferences in AQ4[15]. Question AQ4 is chosen because it captures participants concern about inappropriate use of ER resources. A negative correlation between higher concern for the use of hospital resources and ER visits, would indicate that moral considerations are partly responsible for the preference of visiting the TH. This relationship is examined using Equation 3.7, which is computed for $t = \{TP, TS\}$ and $\omega = L$.

$$ER_self_{i,t,\mu} = \beta_1 + \alpha_t + \alpha_\mu + \sum_{j=2}^{6} \beta_j AQ4_j + \mathbf{X}_i^{'}\beta + \epsilon_{i,t,\mu} \tag{3.7}$$

In Equation 3.7, β_1 is a constant; α_t are treatment specific fixed-effects; α_μ are posterior specific fixed-effects; $AQ4_j$ is a dummy variable for the $j-th$ level of Additional Question 4 with β_j being the corresponding marginal effect; $\mathbf{X}_i^{'}\beta$ is a matrix of covariates and the corresponding marginal effects; $\epsilon_{i,t,\mu}$ is the error component, clustered at the individual level. Table 15 shows the results of this analysis. As expected the marginal effects, with respect to the baseline "Never", are negative in most of the cases. These coefficients are not significant with the exception of *AQ4:Always* in model (3), which however becomes not significant using Bonferroni corrected significance levels. Thus, our results are inconclusive.

Table 16 shows instead the results with AQ6[16]. Results are again inconclusive, as all the estimates are not significant.

Result 3. We conclude that participants show a preference for TH, however it could not be explained by moral considerations about inappropriate use of hospital resources or the risk of infecting other patients. Thus, hypothesis 3 has been partially verified.

[15]Have you ever decided NOT to visit the ER because you were concerned about using resources that could be better allocated to someone in greater need?

[16]Have you ever chosen NOT to visit the ER because you were concerned about the risk of infecting other patients?

Table 15: Linear regressions on $ER_self_{i,t,\mu}$ for $t = \{TP, TS\}$ and $\omega = L$.

	Risky		Ambiguous	
	(1)	(2)	(3)	(4)
AQ4:Rarely	-0.057 (0.045)	-0.057 (0.045)	-0.092 (0.067)	-0.093 (0.067)
AQ4:Sometimes	0.039 (0.043)	0.051 (0.043)	0.012 (0.056)	0.019 (0.057)
AQ4:Most of the time	-0.022 (0.058)	-0.016 (0.070)	-0.025 (0.082)	0.015 (0.093)
AQ4:Always	-0.060 (0.053)	-0.051 (0.059)	-0.138* (0.080)	-0.123 (0.084)
Constant	0.818*** (0.027)	0.503*** (0.155)	0.771*** (0.041)	0.699** (0.271)
N	810	800	486	480
Covariates	No	Yes	No	Yes

Note. Baseline *AQ4:Never*. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors (in parenthesis) are clustered at the individual level. Fixed effects for the treatment and for the posteriors are included.

Table 16: Linear regressions on $ER_self_{i,t,\mu}$ for $t = \{TP, TS\}$ and $\omega = L$.

	Risky		Ambiguous	
	(1)	(2)	(3)	(4)
AQ6: Rarely	0.033 (0.035)	0.032 (0.035)	0.059 (0.046)	0.068 (0.049)
AQ6: Sometimes	0.013 (0.040)	0.019 (0.040)	0.069 (0.052)	0.084 (0.053)
AQ6: Most of the time	0.005 (0.073)	0.001 (0.075)	-0.025 (0.108)	-0.044 (0.106)
AQ6: Always	-0.093 (0.194)	-0.064 (0.179)	-0.049 (0.222)	0.019 (0.205)
Constant	0.792*** (0.023)	0.595*** (0.146)	0.715*** (0.030)	0.828*** (0.215)
N	980	970	588	582
Covariates	No	Yes	No	Yes

Note. Baseline *AQ6:Never*. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors (in parenthesis) are clustered at the individual level. Fixed effects for the treatment and for the posteriors are included.

3.6 Discussion

A relevant issue in the medical sector is emergency room (ER) overcrowding. Non-urgent and inappropriate visits to the ER are among the causes of this issue (Su et al., 2021). Recently, in the attempt to reduce non-urgent and inappropriate visits to the ER some Italian hospitals have adopted a policy that makes real-time ER congestion levels available on their websites. In this work we examine the effect of such policy on patients' decision to visit the ER.

We model the decision process as a utility maximization problem in which an agent shows some health symptoms and must choose between visiting the ER or some territorial health service (TH), like a general practitioner, a primary care facility or the closest pharmacy. The utility functions of the two alternatives are composed by a factor representing the medical benefits of being treated, a factor representing the opportunity cost of time and money spent to be treated and, possibly, an ethical cost which represent the moral consideration about the use of ER resources. This model is informed by the existing literature on the topic of ER visits (Uscher-Pines et al., 2013; Morley et al., 2018; Botelho et al., 2019; M. L. Carret, A. G. Fassa, and Kawachi, 2007; Coleman, 2001; M. L. V. Carret, A. C. G. Fassa, and Domingues, 2009).

The effect of the policy is empirically examined with an incentivized discrete choice experiment with congestion and social externalities in which the ER has the properties of non-excludability (the ER can be always visited) and rivalry (resources are limited and as a consequence higher congestion reduces the quality of service and increases waiting times). Participants are assigned to a condition in which congestion levels are unknown (condition TC) or a condition in which congestion levels are fully known (condition TP). We also introduce a condition in which intermediate congestion levels are equally probable (condition TS), to suggest an alternative use of the policy. During each task, participants are given the probability to be in a critical health condition and then they are instructed to choose between visiting the ER or the TH.

Result 1. Consistently with Hypothesis 1, condition TP shows more frequent visits to the ER compared to condition TC in which congestion is unknown. Coherently with our predictions, this result indicates that real-time information on congestion decrease the payoff uncertainty of ER visits. Specifically, real-time information increase the number of inappropriate visits in congestion Medium, thus potentially contributing to the problem of overcrowding. A null effect is instead found in congestion High. Our analysis als shows that inappropriate visits increase in congestion Low and appropriate visits decrease in congestion Over-Congested. One may argue, however, that transparency improvements of the ER in congestion Low and Over-Congested can help patients to autonomously redirect to less crowded ERs, decreasing the average congestion over the territory.

The analysis of the alternative policy TS suggests no significant difference in the total number of visits with respect to the condition in which congestion levels are unknown (condition TC). In conditions TS-Low and TS-Over-Congested, we observe a similar behavior to equivalent choices in TP-Low and TP-Over-Congested, contributing to the internal validity of our experimental design. In intermediate levels of congestion, TS-M&H, participants behave as in condition TC and TP-High. From a policy perspective, our suggestion offers a valid alternative to condition TP, as participants in TS-M&H show few inappropriate visits and similar number of appropriate visits with respect to condition TC.

Result 2. Our study also test whether there are self-others differences in ER visits. We find an overall higher number of visits to the ER in the *for-others* task. In particular, significant differences are found in R20, R40 and A-Unknwon, when the uncertainty on the health status is high, and in conditions TC, TS and TP-High. These results allow for some considerations on the effect of the *for-others* framing on participants' risk aversion.

In Section 3.3 we introduced a distinction between the uncertainty towards the health status and the uncertainty towards the congestion of the ER. Here, we argue that the differences in ER visits between the *for-*

others and the *for-self* tasks are driven by differences in the perception of the health status. Consider, in particular, differences in TP-High. This is a particular condition where there is no uncertainty on the congestion of the ER and the average payoffs between ER and TH are equivalent. Following standard economic theory, an agent should behave similarly in the two tasks. However, Figure 8 shows a majority of differences greater than 0 and Table 14 shows an overall significant effect. One possible explanation of this result is that when an agent is uncertain about the health status of others, they behave more cautiously, considering the posterior probability as higher, and thus visiting the ER more often. This interpretation is consistent with our framework, as the average payoffs at the ER for critical patients are higher than the respective values for the TH. In our model, this effect is captured by the parameter b.

Our second remark is that any change in risk attitude is to be attributed to the additional framing condition, excluding an effect of the different incentivization strategy. We argue that the use of the strategy method instead of a random incentive system in the *for-others* task, would not have changed the final payoff of the participant in this task, which would have been null in both cases. Therefore, while me might expect differences in the intensity of the effect, we do not expect differences in its direction.

From a policy perspective, these results are useful to understand the demand of ER visits, and more in general of healthcare services, in those situations in which the decision maker is a non-professional and possibly related with the patient. For example, this result may contribute to understand the demand of ER visits in pediatric hospitals.

Result 3. Finally our study investigates the presence of moral consideration towards the use of hospitals' resources. Despite a clear and substantial preference for TH treatments, as shown in Figure 7, we found little to no correlation with moral considerations in AQ4 (see Table 15). This result highlights the importance of future research to understand the drivers of ER visits different from medical benefits, monetary costs and waiting costs that we included in our model. For example, Garrafa

et al. (2020) found a reduction in both appropriate and inappropriate ER visits following the COVID-19 pandemic. The authors suggest this decline may be due to the ongoing restrictions and public appeals to avoid ER visits during the lockdown period. However, it could also originate from the perception that ERs were no longer safe environments.

Additional questions. The analysis of AQ1 and AQ2 suggest that inappropriate ER visits may be more frequent when the uncertainty on the health status is high. When, instead, the symptoms indicate a non-urgent condition, participant tend to correctly assign the appropriate severity and show a low ER uptake. Instead, for diarrhea and urinary infection symptoms, the high variability in the ER uptake and overestimation of symptoms' severity indicate a high uncertainty on the nature of the condition and the correct treatment to adopt. From a policy perspective, reducing uncertainty on the health status with behavioral intervention may help reducing inappropriate visits.

3.7 Conclusion

Our study suggest that real-time information on ER congestion may increase inappropriate visits compared to the common situation in which congestion are unknown. We propose an alternative policy in which the effect on inappropriate visits is instead null for intermediate levels of congestion. These findings have important policy implications. Both policies under analysis improve healthcare system transparency relative to the common situation in which congestion levels are unknown. Greater transparency can help patients make more informed choices, such as visiting less crowded ERs when TH services are inadequate or when uncertainty about their health status is high. On the other hand, information on ER congestion reduces the payoff uncertainty of ER visits, for example reducing the uncertainty about waiting times. As a result, inappropriate visits may increase, and appropriate visits may decrease, as shown in the case of condition TP. Thus, to prevent overcrowding, hospitals must carefully design the way congestion information is presented

on their websites.

Additionally, we find that individuals are more likely to visit the ER when making the decisions for others. This insight has important implications for pediatric hospitals and other healthcare settings where decisions to visit the ER is made by caregivers rather than patients themselves. Understanding changes in attitudes toward risk aversion can help shape policies aimed at managing ER demand more effectively.

Finally, despite observing a clear and substantial preference for TH treatments, our results indicate that this behavior is not driven by moral considerations towards the use of hospital resources. This suggests that policy campaigns based on moral arguments may be ineffective in reducing inappropriate ER visits. Instead, our findings highlight the need to investigate additional drivers of ER visits, as existing literature does not fully explain patient decision-making.

Finally, future research should propose and study the effects of policies that combine transparency with behavioral nudges to encourage appropriate ER use.

Chapter 4

A field experiment for education policies

This essay is based on: Toccafondi N., Bilancini E., Boncinelli L. (2024), Learning CyberSecurity with Story-Driven CTF Challenges: CyberTrials 2023, Communications in Computer and Information Science, vol 2076. Springer, Cham. *doi* : $10.1007/978-3-031-67351-1_21$

4.1 Introduction

The gender gap prevailing in the science, technology, engineering, and mathematics (STEM) domain has been extensively documented (Beede et al., 2011; Wang and Degol, 2017). The problem is particularly relevant in computer science and in cybersecurity fields where women represent only 25% of the workforce (Osborne and S. Morgan, 2022). To address this issue, the Cybersecurity Lab of CINI, in collaboration with the IMT School for Advanced Studies Lucca and the GAME Science Research Center, designed CyberTrials 2023[1]. This game-based educational program is tailored to counteract the gender gap by equipping Italian high school female students with knowledge and skills in computer science and cybersecurity while also providing them with relatable female

[1] https://www.cybertrials.it/

role models.

Education has been proven to play a pivotal role in addressing the gender gap problem. The provision of role models in high schools is shown to have a positive impact on student enrollment in STEM fields (Breda et al., 2020). Moreover, targeted courses have been shown to heighten girls' interest in STEM disciplines (Bystydzienski, Eisenhart, and Bruning, 2015; Basiglio, Del Boca, and Pronzato, 2023; Stern, Reid, and Bancroft, 2015). In the realm of cybersecurity education, it is common to propose gamification based approaches, like cybersecurity themed escape rooms (C. J. Cornel, Rowe, and C. M. Cornel, 2017) and capture the flag (CTF) challenges, which are one of the most prevalent approaches. For example, Leune and Petrilli (2017) simulate cybersecurity breach scenarios using CTF challenges, showing the validity of this method to increase students' engagement, skills and confidence in their own abilities. McDaniel, Talvi, and Hay (2016) deconstruct fundamental cybersecurity concepts through individual CTF challenges, showing that students were motivated to delve deeper into cybersecurity concepts at the end of the course.

CyberTrials 2023 was designed drawing inspiration from the field of cybersecurity education. It consisted of 11 lectures and 9 gamified training modules and it employed two distinct gamification strategies to engage students. First, a point and leaderboard system was introduced using CTF challenges. This system was useful to monitor the progress of each participant and reward the best students with the possibility to attend a final challenge held in Turin on September 7th, 2023. Secondly, CTF challenges were organized following an innovative, narrative-driven framework.

The combination of these two gamification elements represents a novelty in the literature of cybersecurity education. We employ an online survey to examine the relationship between higher engagement in innovative educational activities and students' educational outcomes. Our results suggest that higher engagement with training modules is positively correlated with students' educational outcomes. A null relationship is instead found for higher engagement with lectures. The analysis

is then repeated for each question separately, finding that for some specific topic, lectures proved to be more effective than training modules. This research not only gives educators and practitioners a comprehensive overview of CyberTrials 2023, but it also examines the relationship between higher engagement with lectures and training modules and educational outcomes across 7 distinct cybersecurity domains.

The rest of the work is organized as follows. In Section 4.2, we examine CyberTrials 2023 in detail. In Section 4.3, we examine the methods for our analysis, and we provide the results in Section 4.4. Finally, we discuss the results and conclude our work in Section 4.5.

4.2 CyberTrials 2023 and The Big Game

CyberTrials is a game-based educational program designed by the Cybersecurity Lab of CINI, the IMT School for Advanced Studies Lucca and the GAME Science Research Center. CyberTrials is part of the "Big Game" project by the Cybersecurity Lab of CINI, aimed at promoting awareness, knowledge, interest and skills regarding cybersecurity in high school and university students. Specifically, CyberTrials aims to provide female high school students with entry-level knowledge in computer science and cybersecurity. Female high school students who voluntarily enrolled in the program attended 11 online lectures and 9 gamified training modules. Educational activities have been conducted from January 30th to April 15th. The students who participated in the program received a certificate of participation that counted as a formative credit for the Italian program of "Percorsi per le Competenze Trasversali e l'Orientamento" (PCTO). In addition, the best students won the chance to attend a final challenge on September 7th, 2023, in Turin. In what follows, we provide an exhaustive description of the educational activities, namely the lectures and the training modules.

Topic	Description
Networks	TCP/IP networks behaviour, connecting to a remote service via netcat/telnet, ports enumeration via Nmap.
Web	The HTTP protocol, DNS and address translation, DNS querying via dig.
OSInt	Advanced usage of search engines, social media intelligence, internet archive, metadata and ExifTool.
Threat modeling	Attack trees, risk evaluation metrics and risk assessment strategies.
Social engineering	Reconnaissance of phishing attacks, HTTP requests logging.
Computer forensics	Grep-based data search in the Linux filesystem, Perl-compatible regular expressions, text extraction via strings, Linux terminal history.
Cryptography and steganography	Text encoding mechanisms, base64 encoding, rot cipher, media file encoding and information hiding techniques.
Self-management skills	Ability to cope with competition and stress management.
Legal and ethical aspects	National and European legal framework, ethical hacking and responsible disclosure.
Team building and communication skills	Effort management, work organization, active listening, effective results presentation.

Table 17: Skills and Relative Descriptions

4.2.1 Lectures

Students who voluntarily enrolled in the program attended a series of 11 weekly lectures. The lectures were delivered online by computer science and cybersecurity experts from both academia and industry. In an attempt to provide relatable role models to the students, the experts were specifically chosen to attain gender parity. The first lecture was an introductory lecture. Subsequently, 7 topics were selected to teach basic computer science and cybersecurity skills. In addition, 3 *soft skills* modules were included to inform students on the program's legal aspects and help them meet the program's challenges, which required teamwork skills and an adaptive mindset. Selected topics are described in Table 17 as in Costa et al. (2023).

Students could obtain the certificate of attendance by attending the lectures as they were delivered online or by watching the video recordings provided on the Moodle platform.

As a preliminary measure of the engagement with the program's lectures, we examine the dropout level of the program. Among 941 enrolled students in CyberTrials 2023, 557 attended the first lecture, and 340 attended the last lecture (Costa et al., 2023). Therefore, the drop-out rate is 64% if we consider the total number of enrolled students and 40% if we consider the students who attended the first lecture.

4.2.2 Training Modules

The CyberTrials program provided at the end of each lecture a series of 9 training modules designed to complement the theoretical topics examined during the lectures (Costa et al., 2023). During the first module, students were divided into groups of 4 people, with each group representing a team. During the second module, students were introduced to i) the Discord platform; ii) the CTfd platform; and iii) an instance of Webtop, "a full-fledged Linux environment that teams access via browser-based remote desktop sessions" (Costa et al., 2023). In the successive modules, students were asked to perform a series of challenges on the topics of the corresponding lecture. As an example, provided in Costa et al. (2023), in one of the challenges students were asked to modify the author information of a specific image using tools such as *exiftool*, replacing it with the name of the forger.

The attendance to the training modules was not required to obtain the certificate of participation for the CyberTrials program. Participation and engagement were then promoted using game-based learning techniques (Nesti, 2017). A game-based learning approach has proven to be effective in a variety of educational programs targeting school students where the GAME Science Research Center was involved: promoting sustainable water use (Bilancini, Boncinelli, and Di Paolo, 2023; Di Paolo and Pizziol, 2024), introducing protein biosynthesis (Chiarello et al., n.d.), promoting financial literacy (Bisanti et al., 2022), fostering basic mathematics (Bertoni and Maffia, 2024; Maffia and Silva, 2022), introducing astrophysics (Inchingolo et al., 2023), promoting complementary and general skills (Inchingolo et al., 2023; Mattiassi, Ghirarduzzi, and

Bacaro, 2023). In the field of Cybersecurity, the game-based learning approach can assume different shapes, such as Defense Tower Game and Card game (Jin et al., 2018) and CTF challenges (Švábenský et al., 2021).

In Cybertrials, two game-based features were implemented. First, using the CTFd platform, a point and leaderboard system was introduced. The leaderboard system is known to increase engagement in educational activities (Swacha, 2021) and, in particular, in those related to computer science (Zahedi et al., 2021). Second, all the activities were motivated by a storyline, which unfolded with each module. This resulted in story-driven activities taking the form of a gamified capture-the-flag competition. Although student-tailored strategies are still under discussion (Oliveira et al., 2023), gamification seems to be gender neutral (Zahedi et al., 2021). In what follows, we examine each level of gamification separately.

Capture-the-flag (CTF).

During each module, each team was rewarded with a *flag* for each task they solved. A flag is represented by an alphanumeric code, and it granted a predetermined number of points on the CTFd platform. Each training module consisted of 3 CTF challenges with an increasing level of difficulty. Challenges in each module, as in some popular escape room formats, were sequential. Students had to complete the first challenge to access the second, and similarly, completing the second challenge was necessary to access the third. At the end of the competition, the 20 teams with the most points won the chance to compete in a final challenge held in Turin on September 7th, 2023. CTF challenges have been extensively used in the Computer Science and CyberSecurity field (Martorana, 2022) , and in particular for education purposes (Leune and Petrilli, 2017; McDaniel, Talvi, and Hay, 2016; Švábenský et al., 2021). CTF challenges are popular because they can increase students' overall level of engagement and allow organizers to keep track of participants' attendance and progress. They have been used especially to convey technical knowledge in the domain of data, component and software security (Švábenský et al., 2021).

Story-driven activities.

The popularity of the CTF challenges led to the development of various versions and adaptations of this particular gamification strategy. For example, points can be assigned dynamically on the total number of participants who solved the task. Moreover, participants can compete over a fixed amount of flags. CyberTrials adopted a particular variant of the challenge called *Quest* CTF, according to which challenges were organized following a storyline (Costa et al., 2023). In the story, participants are "white hat" hackers helping an artist against an antagonist who illicitly claimed authorship of some artworks and stole some of her paintings. Story-driven activities are a popular gamification element, and it is known to increase students' motivation (Kalogiannakis, Papadakis, and Zourmpakis, 2021; Bonde et al., 2014; Sánchez-Martín, Cañada-Cañada, and Dávila-Acedo, 2017; Pesare et al., 2016).

A preliminary analysis suggests that gamification elements successfully motivated students to participate to the training modules as the attendance was stable during the weeks of the program. Among the 941 enrolled students, 440 attended the first module, and 379 attended the last module (Costa et al., 2023). Therefore, the drop-out rate is 60% if we consider the total number of enrolled students and 14% if we consider the students who attended the first module.

4.3 Experimental Design

We designed a quasi-experiment to investigate the impact of the program on students' educational outcomes and the relationship between higher attendance to the gamified activities and students' educational outcomes. At the beginning (February) and conclusion (April) of the CyberTrials program's training activities, students were requested to complete an online survey. Both surveys had a similar structure. The treatment group in our analysis comprised students who voluntarily enrolled in the CyberTrials program. A control group, consisting of classmates who did not participate in the program, was also surveyed using the

same questionnaires. The dependent variable in our study is the cumulative number of correct responses to questions assessing competencies in computer science and cybersecurity. Below, we elaborate on the evaluation process.

4.3.1 Online Surveys

The initial survey comprised 35 questions distributed across four distinct sections. In contrast, the second survey consisted of 42 questions, with 34 questions identical to those in the first survey, 7 additional questions, and 1 unique question. Let's delve into the structure of the surveys.

To begin, we included 9 demographic questions aimed at collecting information about the students' backgrounds. Utilizing the responses to these demographic questions, we generated an anonymous time-invariant code. This code served as a means to link students between the first and second surveys seamlessly.

Within the demographic section, there was one specific question that differed between the two surveys. In the first survey, students were asked to indicate if they had participated in CyberTrials 2022, while in the second survey, students were asked to indicate their participation in CyberTrials 2023. This distinction allowed us to track changes over time and assess the impact of the program across different years.

In the subsequent section, we introduced 6 questions aimed at assessing students' level of interest in computer science and cybersecurity. This served as a measure of their motivation regarding these subjects. We included this aspect in our analysis because, due to the non-random allocation of students into the treatment and control groups, motivation could potentially introduce bias into the results concerning educational outcomes. Specifically, we anticipate that highly motivated students are more likely to enroll in the program and, on average, achieve better educational outcomes compared to their peers.

Following this, we incorporated 11 questions designed to evaluate students' digital competencies in computer science and cybersecurity. Each question in this section had three possible answers, with only one

being correct. These surveys were aimed at assessing the competencies and skills acquired during lectures and training modules. The cumulative number of correct answers in this section serves as the dependent variable for our analysis.

Lastly, we asked students 8 questions about their gaming habits. It's worth noting that this information will not be utilized in our analysis, as it is intended for different research purposes.

In the second survey, an extra section was included specifically for students in the treatment group. This supplementary section consisted of seven questions employing a 4-point Likert scale. The use of this scale was advantageous for our survey design as it prevented students from selecting a neutral choice and also reduced the estimated time required to complete the survey. These questions were intended to gauge students' opinions regarding the CyberTrials Project. We collected self-reported data on their attendance, perceived difficulty, and level of interest concerning both the lectures and training modules. Additionally, students were given the opportunity to provide suggestions for future editions of the program.

We incorporated two attention checks into the surveys to control for the participants' attentiveness.

4.3.2 Experimental procedure

The survey was developed using Qualtrics (Qualtrics, 2005). It was initially administered one week before the first lecture and remained open for a duration of two weeks.[2] The second survey was conducted immediately following the conclusion of the last training module and was open for two weeks as well. In total, we gathered 467 responses for the first survey and 177 responses for the second survey, resulting in a combined total of 644 responses.

To perform a causal assessment of the CyberTrials program's impact on students' educational outcomes, it is essential to consider students

[2]It's worth noting that the first lecture, which was a presentation of the program, was assumed to have a null effect on educational outcomes.

who completed both the first and second surveys. However, this condition was met by only 3 students in the control group and 43 female students in the treatment group. Consequently, conducting a causal analysis with such a limited sample would lack sufficient statistical power to yield significant results.

As a result, we opted to explore the relationship between increased attendance in gamified activities and students' educational outcomes. We narrowed our sample of observations to the second wave of surveys, focusing exclusively on students in the treatment group. Our final dataset comprises a total of 100 female students. Table 18 provides a summary of the statistics for students who meet our inclusion criteria.

It's noteworthy that the majority of students are enrolled in technical institutes (57%), which typically impart fundamental knowledge in computer science. Consequently, the average self-reported interest levels in computer science and cybersecurity are approximately rated as 3 on the Likert scale we employed. This corresponds to "Good" on the scale, which ranges from 1 (Poor) to 4 (Excellent). As expected, students in our sample display significant motivation towards the topics covered in the CyberTrials program. Additionally, we observed that 15% of students reported experiencing some form of cyber-attack. Lastly, the majority of students reside in Northern Italy. Specifically, 21% of students live in Piemonte (North-West Italy), 14% in Veneto (North-Eastern Italy), and 11% in Abruzzo (Central Italy).

We will now analyze students' attitudes towards educational activities. Figure 9 presents self-reported measures of participation, interest, and perceived difficulty regarding both lectures and training modules for the students in our sample. We remark that a substantial majority of students found both lectures (95%) and training modules (95%) to be interesting. This indicates a high level of engagement with the educational content(Costa et al., 2023). Also, the data reveals that a significant percentage of students perceived both lectures (85%) and training modules (94%) as challenging. Notably, while attendance and interest levels are quite similar between lectures and training modules, training modules were perceived as more difficult compared to lectures.

Table 18: Summary Statistics for female students in the treatment group and in the second survey.

	Mean	S.E.	Min	Max
Number of correct answers	5.82	0.22	1	10
Age	16.84	0.10	14	19
Family members	3.86	0.08	2	6
Victim of a cyber attack (Dummy)	0.15	0.04	0	1
Participated in CyberTrials 22 (Dummy)	0.05	0.02	0	1
Mothers' education				
High School	0.47	0.05	0	1
Bachelor	0.08	0.03	0	1
Master	0.24	0.04	0	1
Ph.D.	0.02	0.01	0	1
None of the above	0.19	0.04	0	1
School				
Liceo Classico	0.01	0.01	0	1
Liceo Scientifico	0.13	0.03	0	1
Liceo Scientifico – opzione Scienze applicate	0.28	0.05	0	1
Liceo scienze umane – opzione economico sociale	0.01	0.01	0	1
Istituti tecnici	0.57	0.05	0	1
Interest for Computer Science	3.12	0.08	1	4
Interest for Cyber Security	2.96	0.08	1	4
Pre-existing Knowledge in cyber-security	2.48	0.07	1	4
Attention Checks (Dummy)	0.45	0.05	0	1
Observations	100			

Note. We report average values, relative standard errors, and minimum and maximum observed values for each variable.

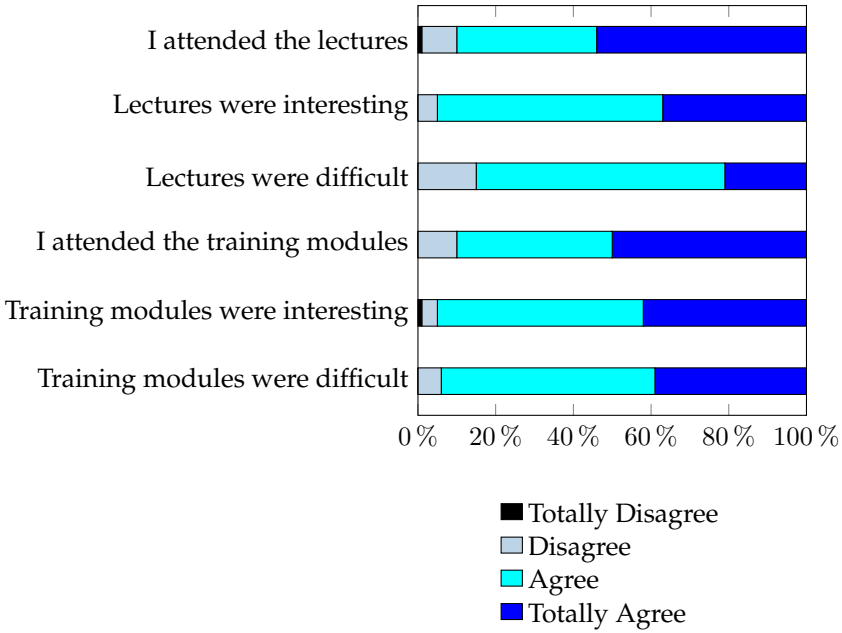


Figure 9: Self-reported measures of participation, interest, and perceived-difficulty in lectures and training modules.

Note. Female students in the treatment group and in the second survey are considered. We report the relative values for each measure.

4.4 Results

In this section, we will initially examine the dependent variable for our study and subsequently present the econometric findings.

The dependent variable in our analysis corresponds to the count of correct answers provided by each student. As shown in Table 18, the average number of correct responses is 5.82. Additionally, it's worth noting that students achieved a minimum of 1 correct answer and a maximum of 10 out of 11 correct answers. Figure 10 presents the average number of correct answers for each individual question. Notably, students performed well on questions Q3 to Q5 and Q7. However, for the remaining questions, the majority of students struggled to provide correct responses.

4.4.1 Econometric analysis

The dependent variable in our study can be conceptualized as a count variable. Accordingly, we employ a Poisson distribution to model it. The model is defined in Equation 4.1, with i being the generic student.

$$ln(E(C_i|\mathbf{X}_i)) = \beta_1 + \beta_2 L_i + \beta_3 TA_i + \mathbf{X}_i'\beta + \epsilon_i \tag{4.1}$$

In Equation 4.1, C_i represents the number of correct answers per student; β_1 is a constant term; L_i and TA_i represent the level of attendance of students to the lectures and to the training activities respectively; $\mathbf{X}_i'$ is a matrix containing student-specific features detailed in Table 18; ϵ_i is the idiosyncratic error term.

Table 19 presents the outcomes of our analysis through several models. Model (1) associates the self-assessed measure of attendance to both lectures and training activities with the dependent variable. Model (2) builds upon Model (1) by additionally controlling for the self-assessed measure of interest in computer science and cybersecurity. It also considers covariates such as age, family size, past experience with cyberattacks, participation in CyberTrials 2022, mothers' education level, students' school, students' region of residency, and a self-assessed measure

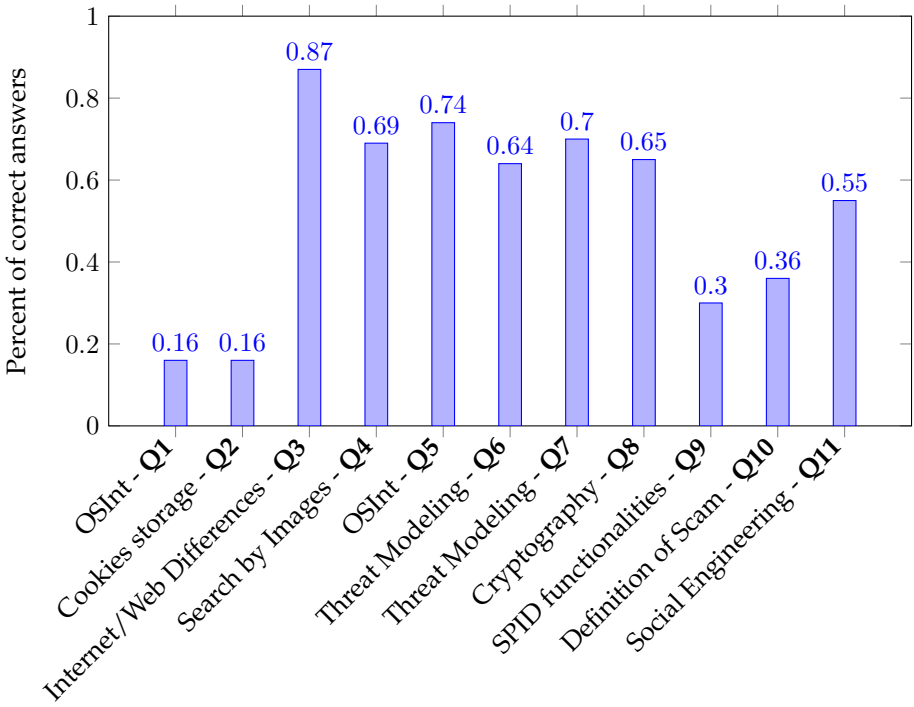


Figure 10: Average number of correct answers for each question and corresponding topic.

of prior knowledge in computer science. Model (3) introduces the variable *Attention Checks*, which equals 1 if the student successfully passed all the survey's attention checks. Model (4) extends Model (3) by incorporating interactions between the *Attention Checks* variable and the measures of participation in lectures and training activities. These models aim to explore the relationships and potential factors influencing students' performance in the CyberTrials program. The inclusion of various covariates allows us to assess the impact of attendance, interest, and other factors on the number of correct answers provided by students.

The variable *Participation to the Activities* is positive and statistically significant in model (1), suggesting that an increase in participation in the activities is positively correlated with the number of correct answers to the questionnaire. The variable *Participation to the Lectures* is instead always negative and not statistically significant. The results are robust to the inclusion in model (2) of students' motivation towards the topic of the program and further controls. As an additional robustness check, in model (3) we control for the variable *Attention Checks*. The variable is not statistically significant, indicating no relevant differences between students who passed all the attention checks and those who did not. Finally, in model (4), we check if students who made the effort to complete the survey correctly and participated in the training activities or the lectures were those who gave the higher number of correct answers. We do not find significant effects. However, model (4) shows a smaller marginal effect for *Participation to the Activities* with respect to the other models. The variable *Attention Checks =1* $\times$ *Participation to the Activities* may have captured some of the effect. In summary, our results indicate that the number of correct answers is positively and significantly correlated with higher attendance to the training modules but not with the lectures.

Figure 11 examines the effects of the variables of interest for each question separately. The marginal effects for *Participation to the activities* are generally positive and significant for most questions. This indicates an overall positive effect of attending the training activities on students' performance. However, null effects are observed for Question 2, Question 4, and Question 10, which primarily assess students' theoret-

ical knowledge. Interestingly, the marginal effects of *Participation to the lectures* for the same questions (Question 2, Question 4, and Question 10) are positive and, in the case of Question 10, significant. This suggests that attending the lectures had a positive impact on students' performance in these theoretical knowledge-based questions. For other questions, the marginal effects of *Participation to the lectures* are either not significant or negative.

Additionally, Table 19 highlights an intriguing result. While interest in computer science and cybersecurity topics did not show significant effects, the variable *Victim of a cyber attack (Dummy)* consistently displayed positive and significant effects. This suggests that past experiences with cyberattacks may be relevant in explaining students' motivation and engagement with cybersecurity topics. Past experiences might have heightened their interest and awareness in this domain, contributing to better performance.

4.5 Discussion

The goal of CyberTrials 2023 was to equip Italian high school female students with knowledge and skills in computer science and cybersecurity while also providing them with relatable female role models as a way to contrast the gender gap prevailing in the STEM field. The program complements 11 theory-driven lectures with 9 hands-on training modules, consisting of a series of CTF challenges organized following an innovative, game-based and narrative-driven framework. The combination of the CTF challenges, which also introduce a point-based scoring approach, is a novel combination of narrative and gamification elements in the field of cybersecurity education.

We employed an online survey to examine the relationship between higher engagement in such innovative educational activities and students' educational outcomes. We find a positive and significant relationship between higher engagement in the training modules of the program and the number of correct answers to questions about selected topics of computer science and cybersecurity. Also, we find a negative but not

Table 19: Poisson regressions on the number of correct answers.

	(1)	(2)	(3)	(4)
Participation to the Lectures	-0.0742	-0.0950	-0.0881	-0.0774
	(0.0537)	(0.0626)	(0.0631)	(0.0907)
Participation to the Activities	0.200**	0.195**	0.209**	0.148*
	(0.0692)	(0.0639)	(0.0665)	(0.0722)
Interest for Computer Science		-0.0306	-0.0257	-0.0511
		(0.0595)	(0.0576)	(0.0679)
Interest for Cyber Security		0.0911	0.0826	0.0841
		(0.0545)	(0.0525)	(0.0524)
Victim of a cyber attack (Dummy)		0.212**	0.206*	0.221**
		(0.0822)	(0.0812)	(0.0854)
Attention Checks (Dummy)			-0.0954	-0.602
			(0.0764)	(0.436)
Attention Checks =1 $\times$ Participation to the Lectures				-0.0641
				(0.177)
Attention Checks =1 $\times$ Participation to the Activities				0.210
				(0.172)
Constant	1.329***	1.176	1.310	1.438
	(0.241)	(0.837)	(0.835)	(0.885)
Observations	100	100	100	100
Extra controls	No	Yes	Yes	Yes
VCE	Robust	Robust	Robust	Robust

Note. Significance levels are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust Standard errors are reported in parentheses. Models (1) through (4) consider the second survey only and students in the treatment group. Models (2) through (4) consider controls in Table 18.

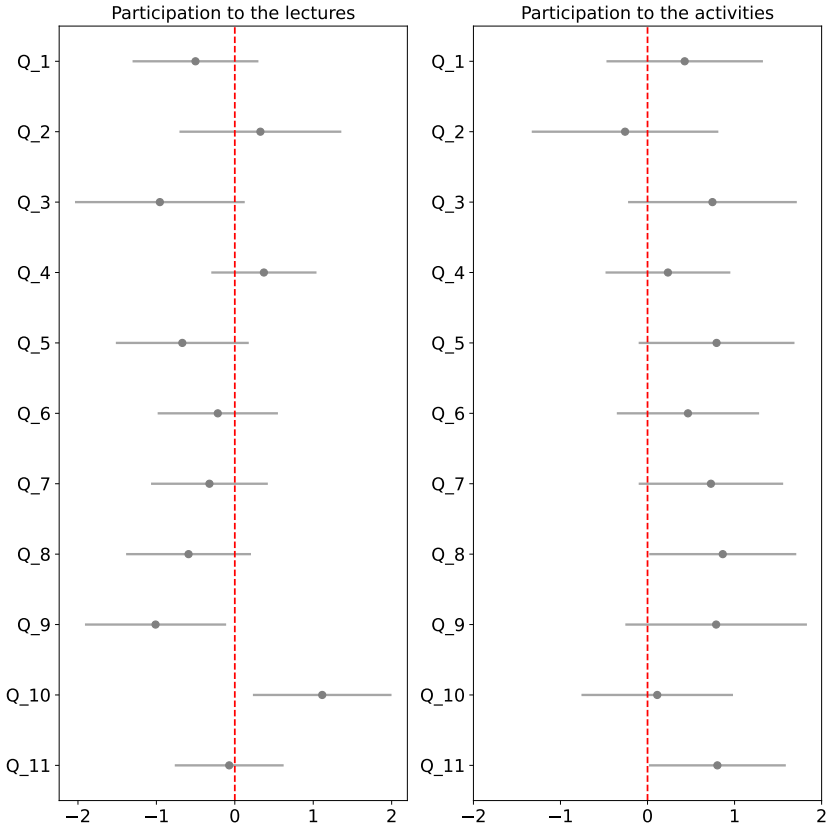


Figure 11: Marginal effects of *Participation to the lectures* and *Participation to the Activities* per question.

Note. Marginal effects and 95% confidence intervals are computed using model (1) in Table 19.

statistically significant relationship between higher engagement in the lectures of the program and the number of correct answers. Our results are robust to the inclusion of several controls, among which students' interest for the topic of the program and students' characteristics.

We repeated the analysis for each question separately. We find a positive relationship between higher engagement in training modules and the number of correct answers for most questions. Opposite results are instead found for Question 10, for which lectures proved to be more effective than training modules. Interestingly, Question 10 evaluates students' comprehension of the definition of *Scam*, which is a theoretical-based question.

Finally, we find a positive relationship between past experiences with cyberattacks and educational outcomes. Past experiences might have heightened students' interest and awareness in the cybersecurity domain, contributing to better performance.

In conclusion, the hands-on training modules appear to have successfully engaged students, as indicated by the low dropout rate and the self-reported measure of interest. Furthermore, participation in the training modules positively correlates with the number of correct responses. These facts suggest the importance of complementing theory-driven lectures with hands-on interactive, gamified activities, particularly in domains like Computer Science in general and Cybersecurity in particular, which require a balanced blend of theoretical and practical knowledge.

Future research should aim to causally evaluate the influence of CyberTrials 2023 on students' intent towards pursuing careers in STEM fields. Additionally, future works could consider individuals who identify as non-binary or in other genders, which are under-represented in the STEM literature. Finally, it is important to investigate which particular topics gain the greatest advantage from the introduction of gamified activities, as well as identify the subjects that are more effectively taught through traditional, theory-focused lectures. Such insights could significantly aid educators and practitioners in tailoring educational programs that are more effectively aligned with their students' needs.

Chapter 5

Conclusion

In this dissertation, I contributed to the growing field of experimental policy analysis with three relevant policy work in the field of consumer protection, health and education. Experimental methods are used to examine causal relationships when observational data are private and not available (Chapter 2), suffer from a censoring bias and are not fit for causal interpretation (Chapter 3) or are non-existent (Chapter 4).

The study in Chapter 2, provides an experimental analysis of virtual currencies as a form of payment for transactions of loot boxes. In particular, it examines three practices (the virtual currency effect, the money illusion effect and non-intuitive exchange rates) common to most loot box transactions, that could potentially distort the economic behaviour of the average consumer. The first result show that the currency itself can modify the spending behaviour of consumers, contributing to the debate on the neutrality of virtual currencies and experimental currency units. We suggest that future research should explore the relevance of the effect in different domains. Our second result show that, when the nominal value of a virtual currency is higher than its real value, subjects reduce their willingness to pay and become more sensitive to risk and ambiguity. This evidence suggests that the money illusion effect can be independent of computational skills. Lastly, the use of non-intuitive exchange rates do not significantly change participants' spending behaviour. One possible

explanation is that, participants employed fast but inaccurate conversion strategies. However, the similarities between conditions T93, T107 and T100, didn't allow to detect these inaccuracies. Based on our findings, we recommend policymakers to support the use of real-world currencies in loot box transactions to improve transparency and consumers' understanding.

The study in Chapter 3, examines the effect of an existing policy that makes real-time ER congestion levels available on the websites of Italian hospitals. Our study suggest that real-time information on ER congestion may increase the total number of visits compared to the common situation in which congestion are unknown. In particular, contrary to the policy maker objective, it may increase inappropriate visits and decrease appropriate visits. On the other hand, the analysis of the alternative policy suggests no significant difference in the total number of visits with respect to the condition in which congestion levels are unknown. From a policy perspective, both policies under analysis improve healthcare system transparency. Greater transparency can help patients make more informed choices, such as visiting less crowded ERs when TH services are inadequate or when uncertainty about the health status is high. On the other hand, information on ER congestion reduces the payoff uncertainty of ER visits, for example reducing the uncertainty about waiting times. Thus, to prevent overcrowding, hospitals must carefully design the way congestion information is presented on their websites. Our study also test whether there are self-others differences in ER visits. We find an overall higher number of visits to the ER in the *for-others* task. We argue that when an agent is uncertain about the health status of others, they behave more cautiously, considering the probability to be in a critical health status as higher, and thus visiting the ER more often. From a policy perspective, these results are useful to understand the demand of ER visits, and more in general of healthcare services in those situations in which the decision maker is a non-professional and possibly related with the patient. For example, this result may contribute to understand the demand of ER visits in pediatric hospitals. Finally our study investigates the presence of moral consideration towards the use of hospitals'

resources. Despite a clear and substantial preference for TH treatments, we found little to no correlation with moral considerations. This result highlights the importance of future research to understand the drivers of ER visits different from medical benefits, monetary costs and waiting costs that we included in our model.

The third study, in Chapter 2, examines the relationship between students' engagement in CyberTrials 2023 educational activities and their educational outcomes. Our results show that the program successfully engaged students, probably as a consequence of the gamification strategies introduced in the program. Moreover, our results indicate the existence of a positive relationship between higher engagement in training modules and students' educational outcomes and a null relationship between higher engagement in lectures and students' educational outcomes. Thus, we conclude by highlighting the importance of complementing theory-driven lectures with hands-on, gamified activities for computer science and cybersecurity education.

Overall, this dissertation provides three examples in which experimental methods are valid alternatives to traditional policy analysis. However, future research should develop a harmonized approach in which traditional policy analysis, field experiments, online experiments and lab experiments work as complementary methods, providing diverse evidence and insights to gather knowledge in social sciences. As a result, public authorities could overcome methodological limitations in the policy design, implementation and evaluation phases, creating more effective policy interventions.

Appendix A

Supplementary Material for Chapter 2

A.1 Supplementary Information

Ethical Approval

The current study has obtained ethical approval from the Joint Ethical Committee of Scuola Normale Superiore, Sant'Anna School of Advanced Studies, and IMT School for Advanced Studies Lucca. This approval was granted on April 27^{th}, 2023, with deliberation number 13/2023. This study has been conducted in compliance with ethical principles, IMT research policies and procedures, General Data Protection Regulation 2016/679 (GDPR) and the Italian DL. n.196/2003 procedures. No form of deception was allowed in this study. The informed consent was obtained from every participant.

Funding

This research was financially supported by IMT School for Advanced Studies Lucca through AXES funds.

Acknowledgements

The paper has benefited from comments by participants at the 7th Meeting of the Behavioral and Experimental Economics Network (Venice). Moreover, we want to thank Ennio Bilancini and Chiara Nardi for their valuable comments. This work has benefitted from the use of AI tools to improve the clarity and the readability of the final draft.

A.2 Sample Balance Tests

We examine the balance of our sample in age, gender, risk preferences, percentage of students, loot box knowledge and time taken in the experiment across treatments. We first conduct a series of Kruskal–Wallis equality-of-populations rank tests and report the corresponding p-values in Table A.1. The null hypothesis is never rejected. We then conduct 10 pairwise comparisons, one for each possible combination of conditions C, T1, T100, T93 and T107, using a Two-sample Wilcoxon rank-sum (Mann – Whitney) test for unmatched data. Table A.1 shows that the null hypothesis is never rejected. We conclude that the sample is balanced in the variables considered across treatments.

A.3 Shapiro – Wilk W test for normality and Levene's homogeneity of variance tests

Table A.2 panel a) reports the p-values for a series of Shapiro – Wilk W tests for normality across lotteries. The null hypothesis is always rejected. Hence, the normal distribution of WTP across lotteries cannot be assumed. Table A.2 panel b) reports the p-values for a series of Levene's homogeneity of variance tests across lotteries, assuming independence across conditions. The null hypothesis is rejected for lotteries R2, R3 and A4. Using the Bonferroni adjusted level of significance (Adjusted $\alpha = 0.00313$), the null hypothesis is rejected only for lottery R3.

Table A.1: Sample balance tests. We report the p-values for the Kruskal-Wallis equality-of-population rank test and the p-values for each possible pairwise Wilcoxon Rank-Sum test.

	Age	Female	Risk Preference	Student	Loot Box	Duration
Kruskal-Wallis	0.86	0.80	0.73	0.16	0.72	0.40
C/T1	0.31	0.24	0.85	0.17	0.50	0.23
C/T100	0.56	0.65	0.49	0.60	0.79	0.29
C/T93	0.91	0.83	0.35	0.26	0.41	0.08
C/T107	0.99	0.40	0.86	0.70	0.71	0.10
T1/T100	0.72	0.48	0.54	0.41	0.68	0.88
T1/T93	0.46	0.47	0.32	0.97	0.79	0.59
T1/T107	0.43	0.91	0.72	0.47	0.35	0.61
T100/T93	0.76	0.88	0.22	0.48	0.55	0.47
T100/T107	0.61	0.64	0.42	0.97	0.56	0.50
T93/T107	0.85	0.60	0.64	0.51	0.30	0.99

A.4 Framing effect and Reference effect for ambiguous lotteries

In this additional section, we examine the framing effect characterizing ambiguous lotteries. Figure 3 panel b) suggests similar levels of WTP in lotteries A1 and A4 as well as in lotteries A2 and A3. Through a series of Wilcoxon matched-pairs signed-rank tests (corrected for multiple testing), we verify the hypothesis that the framing effect has no effect. Our results suggest higher levels of WTP for lottery A4 with respect to lottery A1 (p-value=0.0008) and a null effect on lotteries A2 and A3 (p-value=0.6186). Hence, we reject the null hypothesis of no framing effect for lotteries A1 and A4. The framing effect, however, could differ between participants who played risky lotteries or ambiguous lotteries first. Participants in the first group might use risky lotteries to infer the chance of winning of ambiguous lotteries. Instead, participants in the second group do not have a reference point to compare ambiguous lotteries.

Motivated by the study of He and Hong (2018), which find that being exposed to higher levels of risk induces subject to be more risk averse,

Panel a)

Lottery	P-value	Lottery	P-value
R0	<0.001	A1	0.007
R1	<0.001	A2	<0.001
R2	<0.001	A3	<0.001
R3	<0.001	A4	<0.001
R4	<0.001	A5	<0.001
R5	<0.001		
R6	<0.001		
R7	<0.001		
R8	<0.001		
R9	<0.001		
R10	<0.001		

Panel b)

Lottery	P-value	Lottery	P-value
R0	0.11	A1	0.13
R1	0.27	A2	0.04
R2	0.03	A3	0.10
R3	<0.001	A4	0.01
R4	0.42	A5	0.98
R5	0.48		
R6	0.64		
R7	0.20		
R8	0.49		
R9	0.12		
R10	0.47		

Table A.2: Panel a) Shapiro – Wilk W test for normality. Panel b) Levene's homogeneity of variance tests.

we repeat the analysis in Section 2.4.1, comparing lotteries A1 and A4, and A2 and A3, separating participants between those who played risky lotteries first (with reference) and participants who played ambiguous lotteries first (without reference). In the presence of a reference, results suggest higher WTP in A4 wrt to A1 (p-value$= 0.0001$) and no difference between lotteries A2 and A3 (p-value$= 0.4148$). In the absence of a reference, results suggest no differences between lotteries A4 and A1 (p-value$= 0.4048$) and lotteries A2 and A3 (p-value$= 0.9711$).

A.5 Ambiguity Aversion

We compare risky and ambiguous lotteries as participants may express ambiguity aversion (Ellsberg (1961), Abbey et al. (2019)). As previously documented in Eisenberger and Weber (1995), participants with ambiguity aversion preferences report smaller values of WTP for ambiguous lotteries compared to risky lotteries. We conduct a series of comparisons using the Wilcoxon matched-pairs signed-rank test. First, we compare the WTP of lotteries A1 and A4 with the average WTP of lotteries that have a high but not sure probability of winning, namely lotteries from R6 to R9. Our results indicate similar levels of WTP between lottery A1 and the average of lotteries from R6 to R9 (p-value$= 0.6975$). On the other hand, the WTP in lottery A4 is higher with respect to the average of lotteries from R6 to R9(p-value< 0.001). Similarly, we compare the WTP of lotteries A2 and A3 with the average WTP of lotteries that have a low but positive probability of winning, namely lotteries from R1 to R4. Our results indicate smaller levels of WTP for lotteries A2 and A3 with respect to the average of lotteries from R1 to R4 (both p-values< 0.001). Finally, we compare lottery A5 with lottery R5, obtaining smaller levels of WTP in lottery A5 with respect to lottery R5 (p-value$=< 0.001$) Participants show ambiguity aversion preferences, i.e. smaller levels of WTP for ambiguous lotteries in lotteries A2, A3 and A5. Participants show ambiguity-loving preferences for lottery A4, i.e. the lottery with high probability of winning in the losing framing. Finally, no differences are found for lottery A1.

A.6 Robustness Check for Hypothesis 2

In this section, we compare the average WTP between conditions T93&T107 and T1 as a robustness check for hypothesis 2. Table A.3 shows the Cohen's d effect sizes for T93&T107-T1. The majority of the reported effect sizes are negative, with the exception of lotteries R10 and A4, for which it is exactly equal to 0. The distribution of positive and negative effect sizes is modelled using a Binomial $(14, 0.5)$. Under the null hypothesis of no difference between conditions, the probability that 0 out of 14 lotteries exhibit a positive effect size is <0.001. Therefore, we reject the null hypothesis.

A.7 Additional Results

In this section, we conduct two additional analyses. First, we compare the average WTP between conditions T100 and C, and T93&T107 and C. Then conditions T93 and T107 are compared separately with condition T100.

Table A.4 shows the Cohen's d effect sizes for T100-C and T93&T107-C. In both comparisons and for each lottery, we observe a series of small effect sizes that are not significantly different from 0. We test the null hypothesis of no difference between conditions using the approach shown in Section 2.4.1. Table A.4 reports 5 positive effect sizes between conditions T100 and C and 7 between conditions T93&T107 and C. For the first test, lottery R0 is excluded as the corresponding effect size is equal to 0. The distribution of positive and negative effect sizes is modelled using a Binomial $(15, 0.5)$. Under the null hypothesis of no difference, the probability that 5 out of 15 lotteries exhibit a positive effect size is equal to 0.0916. Therefore, the null hypothesis is not rejected. Similarly, for the second test, we use a Binomial $(16, 0.5)$, obtaining a probability of obtaining 7 positive values equal to 0.1746. Therefore, we do not reject the null hypothesis of different WTP between conditions T93&T107 and C.

A similar approach is used to compare conditions T93 and T107 with

condition T100. Assuming money illusion theories to be valid, we should expect the average WTP to be higher in T93 and smaller in T107 with respect to condition T100. Table A.5 shows that the average WTP is higher in condition T100 with respect to condition T93 in 13 out of 16 lotteries. Moreover, effect sizes are significant for lotteries R0 and R1. The probability of obtaining 13 negative values out of a Binomial(16,0.5) is 0.0085. Therefore, contrary to our expectations, we can reject the null hypothesis that the WTP is higher in T93 with respect to T100. If we consider conditions T107 and T100, we get 9 positive values and no significant effect sizes. The probability of obtaining 9 positive values out of a Binomial(16,0.5) is 0.1746. Therefore, we do not reject the null hypothesis of different WTP between conditions T107 and T100.

Table A.3: Cohen's d Effect Sizes and 95% CI

Risky Lotteries

	T93&T107-T1	
	Effect Size	CI
R0	-0.18	[-0.370, 0.040]
R1	-0.17	[-0.390, 0.070]
R2	-0.23	[-0.440, -0.020]
R3	-0.36	[-0.560, -0.150]
R4	-0.12	[-0.350, 0.100]
R5	-0.33	[-0.550, -0.090]
R6	-0.35	[-0.580, -0.130]
R7	-0.29	[-0.510, -0.050]
R8	-0.23	[-0.470, -0.001]
R9	-0.15	[-0.380, 0.070]
R10	0.00	[-0.230, 0.240]

Ambiguous Lotteries

	T93&T107-T1	
	Effect Size	CI
A1	-0.12	[-0.370, 0.100]
A2	-0.37	[-0.580, -0.160]
A3	-0.24	[-0.450, -0.020]
A4	-0.00	[-0.210, 0.230]
A5	-0.22	[-0.440, 0.020]

Table A.4: Cohen's d Effect Sizes and 95% CI

Risky Lotteries

	T100-C		T93&T107-C	
	Effect Size	CI	Effect Size	CI
R0	-0.00	[-0.200, 0.210]	0.02	[-0.210, 0.220]
R1	-0.11	[-0.330, 0.110]	-0.11	[-0.330, 0.120]
R2	-0.09	[-0.290, 0.140]	-0.14	[-0.350, 0.090]
R3	-0.17	[-0.380, 0.050]	-0.21	[-0.410, 0.010]
R4	-0.08	[-0.290, 0.150]	0.01	[-0.210, 0.260]
R5	-0.16	[-0.390, 0.070]	-0.12	[-0.340, 0.110]
R6	-0.07	[-0.300, 0.160]	-0.11	[-0.350, 0.120]
R7	-0.02	[-0.250, 0.220]	-0.11	[-0.330, 0.110]
R8	0.10	[-0.130, 0.330]	0.05	[-0.190, 0.290]
R9	0.14	[-0.080, 0.370]	0.06	[-0.180, 0.290]
R10	0.16	[-0.070, 0.380]	0.13	[-0.100, 0.360]

Ambiguous Lotteries

	T100-C		T93&T107-C	
	Effect Size	CI	Effect Size	CI
A1	0.01	[-0.210, 0.240]	0.07	[-0.160, 0.290]
A2	-0.05	[-0.280, 0.170]	-0.19	[-0.400, 0.050]
A3	-0.07	[-0.290, 0.170]	-0.07	[-0.260, 0.150]
A4	0.22	[-0.010, 0.430]	0.21	[-0.010, 0.440]
A5	-0.15	[-0.380, 0.090]	-0.19	[-0.390, 0.050]

Table A.5: Cohen's d Effect Sizes and 95% CI

Risky Lotteries

	T93-T100		T107-T100	
	Effect Size	CI	Effect Size	CI
R0	-0.20	[-0.300,-0.120]	0.14	[-0.190,0.350]
R1	-0.26	[-0.490,0.000]	0.18	[-0.080,0.430]
R2	-0.21	[-0.470,0.050]	0.07	[-0.230,0.330]
R3	-0.11	[-0.390,0.180]	0.05	[-0.230,0.310]
R4	-0.06	[-0.350,0.220]	0.25	[-0.020,0.560]
R5	-0.00	[-0.300,0.290]	0.10	[-0.170,0.380]
R6	-0.12	[-0.430,0.180]	0.05	[-0.210,0.330]
R7	-0.07	[-0.400,0.210]	-0.13	[-0.410,0.150]
R8	-0.06	[-0.350,0.250]	-0.04	[-0.310,0.240]
R9	-0.11	[-0.380,0.160]	-0.07	[-0.360,0.250]
R10	0.02	[-0.250,0.340]	-0.08	[-0.360,0.190]

Ambiguous Lotteries

	T93-T100		T107-T100	
	Effect Size	CI	Effect Size	CI
A1	0.13	[-0.160,0.440]	-0.00	[-0.280,0.270]
A2	-0.20	[-0.480,0.070]	-0.09	[-0.360,0.180]
A3	-0.11	[-0.420,0.200]	0.09	[-0.170,0.350]
A4	0.08	[-0.200,0.400]	-0.06	[-0.350,0.220]
A5	-0.11	[-0.410,0.180]	0.04	[-0.240,0.300]

A.8 Additional tables

Table A.6: Non-parametric Effect Sizes

Risky Lotteries

	T1-C		T100-T1		T93&T107-T100	
	Effect Size	CI	Effect Size	CI	Effect Size	CI
R0	0.10	[0.008, 0.200]	0.10	[0.007, 0.200]	0.02	[0.002, 0.140]
R1	0.03	[0.002, 0.150]	0.02	[0.001, 0.140]	0.07	[0.003, 0.190]
R2	0.06	[0.003, 0.170]	0.01	[0.002, 0.140]	0.02	[0.001, 0.130]
R3	0.08	[0.005, 0.190]	0.13	[0.020, 0.240]	0.00	[0.002, 0.130]
R4	0.06	[0.003, 0.160]	0.07	[0.004, 0.180]	0.03	[0.002, 0.160]
R5	0.08	[0.005, 0.200]	0.16	[0.050, 0.260]	0.01	[0.002, 0.140]
R6	0.12	[0.010, 0.230]	0.16	[0.040, 0.260]	0.01	[0.001, 0.130]
R7	0.08	[0.005, 0.190]	0.10	[0.008, 0.210]	0.05	[0.003, 0.160]
R8	0.14	[0.030, 0.250]	0.12	[0.010, 0.220]	0.02	[0.002, 0.130]
R9	0.12	[0.020, 0.230]	0.10	[0.006, 0.220]	0.02	[0.002, 0.140]
R10	0.05	[0.003, 0.170]	0.00	[0.001, 0.140]	0.00	[0.002, 0.130]

Ambiguous Lotteries

	T1-C		T100-T1		T93&T107-T100	
	Effect Size	CI	Effect Size	CI	Effect Size	CI
A1	0.09	[0.004, 0.210]	0.08	[0.006, 0.200]	0.03	[0.002, 0.150]
A2	0.13	[0.010, 0.230]	0.11	[0.020, 0.220]	0.06	[0.004, 0.170]
A3	0.09	[0.006, 0.210]	0.06	[0.004, 0.180]	0.02	[0.003, 0.140]
A4	0.11	[0.010, 0.220]	0.02	[0.002, 0.140]	0.03	[0.002, 0.140]
A5	0.03	[0.002, 0.130]	0.10	[0.007, 0.220]	0.02	[0.002, 0.130]

A.9 Additional Figure

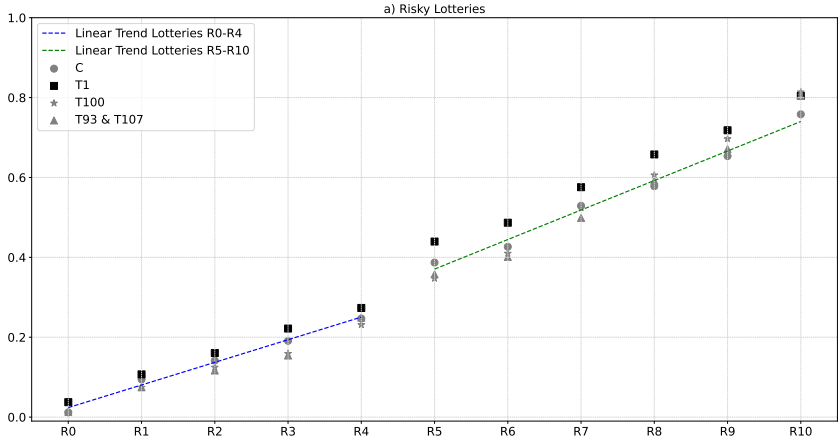


Figure A.1: Average WTP across conditions and lotteries in risky and ambiguous lotteries.

Note. The blue line depicts the trend line between lotteries R0-R4. The green line depicts the trend line between lotteries R5-R10.

Please write down the number 186 to show that you are paying attention.

The urn from which the ball in the box comes, was containing:

- **75 green** balls and
- **25 red** balls

This decision is not eligible to determine your final payment.

Please write down the number 279 to show that you are paying attention.

The number of **green** and **red** balls is high in the urn from which the ball has been extracted.

This decision is not eligible to determine your final payment.

Figure A.2: Attention Checks

Question 1

Please tell us, in general, **how willing or unwilling you are to take risks**. Please use a scale from 0 to 10, where 0 means *"completely unwilling to take risks"* and a 10 means you are *"very willing to take risks"*.

You can also use any numbers between 0 and 10 to indicate where you fall on the scale.

0 1 2 3 4 5 6 7 8 9 10

Question 2

Do you know what a *loot box* is?

- No, I don't
- Yes, I do but I do not buy them
- Yes, I do and I usually buy them

Result

Your willingness to spend was 0, while the actual cost of the box is 28.83.
You did not reveal the content of the box.

Therefore, you earn **93 TK**

What is the value in GBP?

Please answer below.

Figure A.3: Additional Questions

Appendix B

Supplementary Material for Chapter 3

B.1 Supplementary Information

Ethical Approval

The study has obtained ethical approval from the Joint Ethical Committee of Scuola Normale Superiore, Sant'Anna School of Advanced Studies, and IMT School for Advanced Studies Lucca. The approval was granted on November 11^{th}, 2024, prior to the administration of the initial online survey.

Acknowledgments

The work was partially supported by the project "THE – Tuscany Health Ecosystem" (CUP: D63C22000400001), funded by the European Union - Next Generation EU program. Also, we sincerely want to thank all the participants to the 13^{th} BEEN meeting in Bologna, the IMT Workshop on Behavioral and Experimental Economics and all the researcher from the Danish Center of Health Economics (DACHE) for the insightful feedbacks and useful comments. This work has benefitted from the use of AI tools to improve the clarity and the readability of the final draft.

B.2 Expected number of visits to the ER

In this section we use the model proposed in Section 3.3 to predict participants' behavior. In particular we compute the posterior probability μ at which an agent is indifferent between choosing ER or TH, i.e. the root of Equation B.1.

$$\sum_{\omega \in \{L,M,H,O\}} \pi(\omega)[\mu(c)u_{c\omega}(ER)+(1-\mu(c))u_{s\omega}(ER)] = \mu(c)0.2+(1-\mu(c))0.6 \tag{B.1}$$

Table B.1 and Table B.2 show the values of μ which make a risk netrual and a risk averse agent, with quadratic function of risk aversion $u(x) = \sqrt{x}$ indifferent for any condition and congestion level. To understand the meaning of this value take condition TC in Table B.1 as an example. The value corresponds to 0.39, then we expect that for $\mu > 0.39$ the agent will choose ER and for $\mu < 0.39$ the agent will choose TH. Therefore, the lower the value and the more frequent ER visits. Moreover, notice that under risk aversion the expect number of visits to the ER across conditions. Finally, one can observe that between Table B.1 and Table B.2 the expected behavior for intermediate levels of congestion is similar, and it is also similar to the behavior of a participant with risk neutrality in condition TC.

Congestion	TC	TP	TS
Low	0.39	0	0.4
Medium		0.29	
High		0.5	
Over-Congested		0.67	0.67
Total	0.39	0.36	0.35

Table B.1: Posterior probabilities μ which make a risk neutral agent indifferent between ER and TH

Congestion	TC	TP	TS
Low	0.46	0	0
Medium		0.27	0.4
High		0.5	
Over-Congested		0.75	0.75
Total	0.46	0.38	0.38

Table B.2: Posterior probabilities μ which make a risk averse agent with utility function $u(x) = \sqrt{x}$ indifferent between ER and TH.

B.3 Additional Questions - Results

In this section we examine the results of the additional questions in our study. Figure B.1 shows the results from AQ1 and AQ2. Specifically, it shows the ER uptake (black diamonds) and 95% confidence intervals across medical conditions in Table 9, with symptoms' severity categorized from 1 (most urgent) to 5 (least urgent). First, participants correctly identified flu and back pain symptoms as non-urgent. In support of this claim, the majority of participants categorized these symptoms in categories 4 or 5. Moreover, the ER uptake is low and with a low variance. Participants also correctly identified insect bite symptoms as urgent, as seen by the high ER uptake with low variance. On the other hand, for diarrhea and urinary infection symptoms ER uptake variance is high and with average values around 60%, indicating considerable indecision on the correct treatment to adopt.

Our results are consistent with the results of Botelho et al. (2019). Participants in our study overestimate the severity diarrhea, urinary infection and insect bite symptoms. In these cases, the majority of the responses (>50%) assigns symptoms to higher severity categories than what would have been correct. Therefore, we claim that participants tend to overestimate symptoms severity when the uncertainty on the health status is high. Regarding ER uptake, we get similar results for flu and back pain symptoms. On the other hand, the ER uptake of our participants is higher for diarrhea, urinary infection and insect bite symptoms.

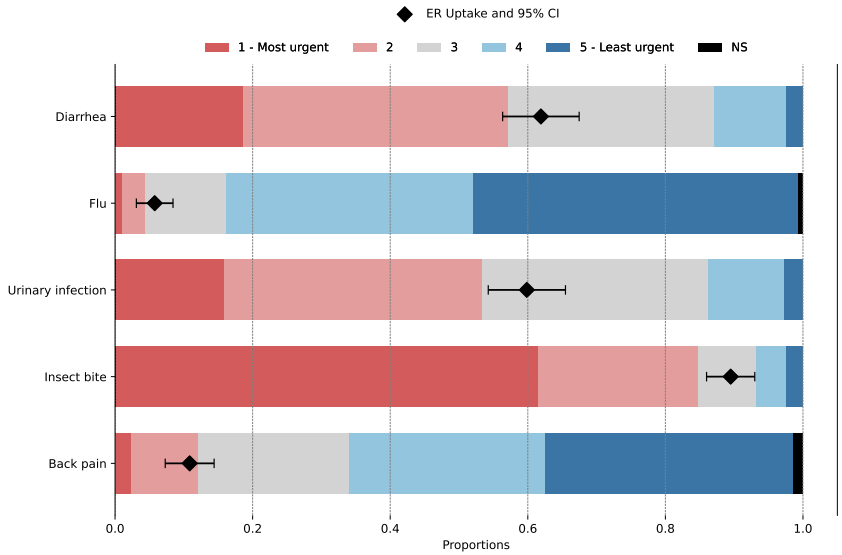


Figure B.1: Statistics on AQ1 and AQ2.

Note. We report the average ER uptake with 95% CI. Moreover, for each item, we report the perceived symptoms severity. Correct answers: Diarrhea-3; Flu-4; Urinary infection-4; Insect bite-2; Back pain-4.

We argue that eventual differences can be the consequence of different incentive schemes and recruitment characteristics, however, overall the results are comparable.

We conclude that, AQ1 and AQ2 suggest that inappropriate ER visits may be more frequent when the uncertainty on the health status is high. When the symptoms indicate a non-urgent condition, participant tend to correctly assign the severity and show a low ER uptake. Instead, for diarrhea and urinary infection symptoms, the high variability in the ER uptake and overestimation of symptoms' severity indicate a high uncertainty on the nature of the condition and the correct treatment to adopt. From a policy perspective, reducing uncertainty on the health status with behavioral intervention may help reducing inappropriate visits.

Figure B.2 presents the results from AQ3 to AQ6. In AQ3 most partic-

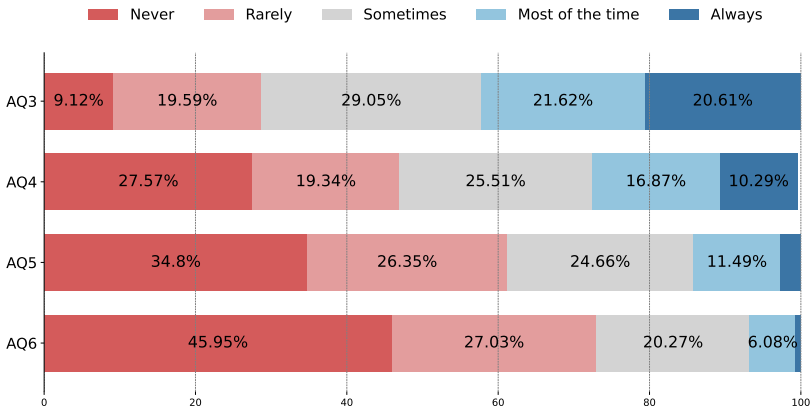


Figure B.2: Answers to questions AQ3-AQ6.

ipants reported considering considered how their visit could impact hospitals' resources. However, AQ4 revelas that these consideration didn't modified their behavior, as shown by the sum of the *Never* and *Rarely* categories (46.91%). AQ5 and AQ6 show that the behavior of participants towards the ER visits was not much affected by the risk of being infected (AQ5) or the risk of infecting (AQ6) other patients. It can be seen, however, that among these two, the risk of being infected by other patients at the ER (AQ5) was more concerning than the risk of infecting others. Figure B.3 shows the results of AQ7[1], revealing that most of the participants in our study just occasionally visited the ER in the previous 5 years. Finally, Figure B.4 shows the results of AQ8[2], revealing that the qulity of the alternatives for the participants in our study is in most of the cases either *average* or *good*.

B.4 Robustness check

In the proposed experimental design, participants observe a probability μ for each class of posterior. These probabilities are randomly drawn

[1]How often have you visited the Emergency room in the last 5 years
[2]What is the quality of the alternatives to the emergency room in your area

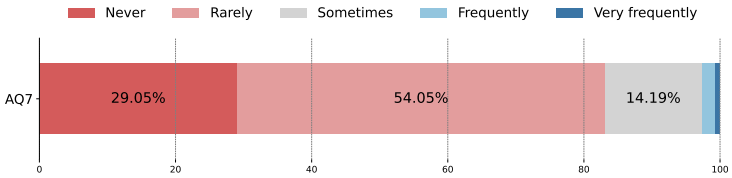


Figure B.3: Answers to AQ7.

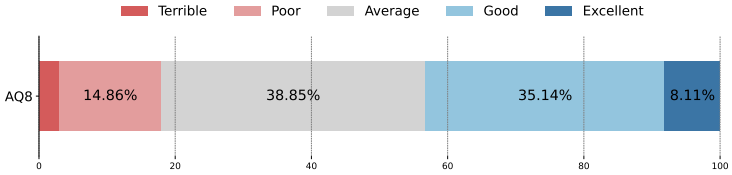


Figure B.4: Answers to AQ8.

from a uniform distribution with extreme values specified in Table 7. An incorrect randomized assignment could systematically show higher or lower probabilities in one condition, introducing relevant bias in our results. However, we show that this is not the case. Table B.3 presents the median values of posterior probabilities shown to participant in each posterior class and each experimental condition. Table B.4 provides a series of non-parametric analysis showing that posterior probabilities are balanced across conditions and posterior classes.

B.5 Addtional Tables

Table B.3: Median probability per class of posterior and relative standard errors

	TC	TP	TS
R0	9.00	9.00	10.00
	0.28	0.29	0.31
R20	29.50	30.00	28.00
	0.29	0.28	0.29
R40	49.00	50.00	50.00
	0.29	0.29	0.29
R60	70.00	70.00	69.00
	0.28	0.30	0.28
R80	89.50	90.00	90.00
	0.30	0.31	0.31

Note. This table shows the median probability for each class of posterior and each condition. It also shows the standard errors.

Table B.4: Posterior Balance test

Covariate	TC/TP	TC/TS	TP/TS	Kwallis
R0	0.707	0.555	0.847	0.837
R20	0.461	0.616	0.204	0.448
R40	0.738	0.658	0.924	0.899
R60	0.800	0.702	0.964	0.932
R80	0.736	0.910	0.827	0.943

Note. This table report a series of balance tests on the posterior probabilities. In columns TC/TP, TC/TS and TP/TS, we report a series of Wilcoxon rank-sum tests for unmatched data. In column K-Wallis we report the Kruskal–Wallis equality-of-populations rank tests across conditions TC, TP and TS. Significance levels are indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Heterogeneity Analysis. Linear regressions on $ER_self_{i,t,\mu,\omega}$ without covariates.

	Low		Medium		High		Over-Congested	
	TP	TS	TP	TS	TP	TS	TP	TS
R0	0.435*** (0.058)	0.354*** (0.057)	0.085* (0.047)	-0.029 (0.035)	-0.018 (0.037)	-0.029 (0.035)	-0.039 (0.034)	-0.06* (0.031)
R20	0.644*** (0.055)	0.517*** (0.059)	0.303*** (0.062)	0.022 (0.051)	-0.016 (0.048)	0.022 (0.051)	-0.027 (0.048)	-0.1** (0.04)
R40	0.227*** (0.055)	0.168*** (0.059)	0.186*** (0.058)	-0.013 (0.067)	-0.144** (0.069)	-0.013 (0.067)	-0.443*** (0.064)	-0.357*** (0.067)
R60	-0.001 (0.025)	-0.031 (0.03)	-0.011 (0.027)	-0.031 (0.03)	-0.073** (0.035)	-0.031 (0.03)	-0.321*** (0.052)	-0.263*** (0.049)
R80	0.04 (0.024)	0.04* (0.024)	0.019 (0.028)	0.02 (0.028)	-0.002 (0.031)	0.02 (0.028)	-0.063 (0.039)	-0.011 (0.033)
Unlikely	0.465*** (0.056)	0.415*** (0.055)	0.063 (0.039)	0.021 (0.034)	0.002 (0.031)	0.021 (0.034)	0.022 (0.034)	0.001 (0.031)
Unknwon	0.335*** (0.062)	0.258*** (0.066)	0.253*** (0.066)	0.147** (0.069)	-0.211*** (0.069)	0.147** (0.069)	-0.355*** (0.062)	-0.318*** (0.064)
Likely	-0.012 (0.033)	-0.021 (0.034)	-0.012 (0.033)	-0.011 (0.033)	-0.032 (0.036)	-0.011 (0.033)	-0.259*** (0.052)	-0.172*** (0.047)
All Posteriors	0.267*** (0.023)	0.213*** (0.023)	0.111*** (0.025)	0.016 (0.025)	-0.062** (0.025)	0.016 (0.025)	-0.186*** (0.025)	-0.16*** (0.025)

Note. A separate linear regression has been estimated for any combination of posteriors and congestion levels. Standard errors are robust to heteroskedasticity, having only 1 observation per id. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table B.6: Heterogeneity Analysis. Linear regressions on $ER_self_{i,t,\mu,\omega}$ with covariates.

	Low		Medium		High		Over-Congested	
	TP	TS	TP	TS	TP	TS	TP	TS
R0	0.432*** (0.1)	0.35*** (0.058)	0.07 (0.075)	-0.037 (0.035)	-0.075 (0.075)	-0.031 (0.034)	-0.028 (0.06)	-0.055* (0.03)
R20	0.579*** (0.097)	0.499*** (0.06)	0.351*** (0.118)	0.031 (0.054)	-0.052 (0.099)	0.034 (0.052)	-0.037 (0.093)	-0.081** (0.041)
R40	0.315*** (0.096)	0.175*** (0.063)	0.348*** (0.103)	0.015 (0.069)	0.033 (0.117)	0.044 (0.069)	-0.289** (0.118)	-0.319*** (0.069)
R60	0.002 (0.054)	-0.042 (0.033)	0.061 (0.058)	-0.021 (0.031)	0.017 (0.064)	-0.014 (0.031)	-0.214** (0.098)	-0.256*** (0.052)
R80	0.141* (0.072)	0.041* (0.024)	0.187*** (0.072)	0.021 (0.029)	0.19*** (0.072)	0.024 (0.029)	0.111 (0.078)	-0.004 (0.032)
Unlikely	0.276*** (0.101)	0.387*** (0.057)	-0.031 (0.086)	0.028 (0.038)	-0.034 (0.087)	0.039 (0.038)	-0.074 (0.085)	0.016 (0.035)
Unknwon	0.369*** (0.104)	0.268*** (0.067)	0.352*** (0.113)	0.171** (0.071)	-0.173 (0.114)	0.176** (0.071)	-0.264** (0.105)	-0.28*** (0.066)
Likely	0.063 (0.081)	-0.027 (0.036)	0.147* (0.077)	-0.01 (0.031)	0.129* (0.078)	0.0 (0.031)	-0.084 (0.097)	-0.162*** (0.047)
All Posteriors	0.272*** (0.037)	0.206*** (0.024)	0.186*** (0.04)	0.025 (0.026)	0.004 (0.041)	0.034 (0.026)	-0.11*** (0.041)	-0.143*** (0.025)

Note. A separate linear regression has been estimated for any combination of posteriors and congestion levels. Standard errors are robust to heteroskedasticity, having only 1 observation per id. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table B.7: Linear regressions on $ER_self_{i,t,\mu,\omega}$ for intermediate levels of congestion.

	Risky		Ambiguous	
	(1)	(2)	(3)	(4)
TP	0.033 (0.021)	0.101*** (0.035)	0.010 (0.025)	0.055 (0.046)
TS	-0.006 (0.021)	0.004 (0.021)	0.052* (0.028)	0.065** (0.028)
Age		0.002 (0.001)		0.000 (0.001)
Female?		0.009 (0.019)		-0.010 (0.023)
Student?		0.018 (0.017)		0.007 (0.020)
Children?		0.016 (0.023)		0.045 (0.032)
Caring Responsabilities?		0.013 (0.016)		-0.010 (0.018)
Chronic Illness?		0.050 (0.031)		0.060** (0.028)
Employed?		-0.047 (0.035)		0.032 (0.037)
Risk Preferences		0.000 (0.004)		-0.006 (0.005)
Attention Checks		0.053** (0.025)		0.032 (0.036)
Previous experiences		0.018* (0.010)		0.005 (0.011)
Quality of alternatives		0.011 (0.008)		-0.001 (0.012)
Constant	0.564*** (0.014)	0.297*** (0.096)	0.503*** (0.019)	0.375*** (0.133)
N	2960	2940	1776	1764

Note. The baseline treatment is condition TC. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the individual level and reported in parentheses. Estimates are computed in the long format; hence, there are 16 observations per participant, one for each combination of posterior and congestion level. Models (1) and (2) consider risky posteriors. Models (3) and (4) consider ambiguous posteriors. Models (1) and (3) are estimated without further controls. Models (2) and (4) are estimated with the reported controls.

Table B.8: Heterogeneity Analysis. Linear regression on $ER_self_{i,t,\mu,\omega}$ with TP as the baseline and TS as the regressor without covariates

	TS			
	(1) Low	(2) Medium	(3) High	(4) Over-Congested
R0	-0.081 (0.071)	-0.114*** (0.044)	-0.011 (0.033)	-0.021 (0.025)
R20	-0.127** (0.064)	-0.282*** (0.063)	0.038 (0.05)	-0.073* (0.038)
R40	-0.059 (0.047)	-0.199*** (0.059)	0.131* (0.07)	0.086 (0.064)
R60	-0.03 (0.03)	-0.019 (0.032)	0.042 (0.039)	0.058 (0.067)
R80	0.0 (0.014)	0.001 (0.025)	0.021 (0.028)	0.053 (0.04)
Unlikely	-0.051 (0.072)	-0.043 (0.041)	0.019 (0.034)	-0.022 (0.034)
Unknwon	-0.078 (0.056)	-0.106* (0.065)	0.358*** (0.067)	0.037 (0.054)
Likely	-0.009 (0.036)	0.001 (0.034)	0.022 (0.037)	0.087 (0.063)
All Posteriors	-0.054*** (0.021)	-0.095*** (0.025)	0.077*** (0.025)	0.026 (0.024)

Note. A separate linear regression has been estimated for any combination of posteriors and congestion levels. . Model (1) compares $ER_self_{i,TP,\mu,L}$ with $ER_self_{i,TS,\mu,L}$. Model (2) compares $ER_self_{i,TP,\mu,M}$ with $ER_self_{i,TS,\mu,M\&H}$. Model (2) compares $ER_self_{i,TP,\mu,H}$ with $ER_self_{i,TS,\mu,M\&H}$. Model (4) compares $ER_self_{i,TP,\mu,C}$ with $ER_self_{i,TS,\mu,C}$. Standard errors are robust to heteroskedasticity, having only 1 observation per id. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table B.9: Heterogeneity Analysis. Linear regression on $ER_self_{i,t,\mu,\omega}$ with TP as the baseline and TS as the regressor with covariates

	TS			
	(1) Low	(2) Medium	(3) High	(4) Over-Congested
R0	-0.121 (0.135)	-0.129 (0.086)	0.049 (0.09)	-0.064 (0.052)
R20	-0.146 (0.111)	-0.427*** (0.113)	0.043 (0.109)	-0.114 (0.089)
R40	-0.179* (0.1)	-0.403*** (0.112)	-0.052 (0.14)	-0.071 (0.106)
R60	-0.051 (0.067)	-0.108 (0.078)	-0.062 (0.084)	-0.098 (0.135)
R80	-0.024 (0.049)	-0.134* (0.079)	-0.148* (0.081)	-0.085 (0.092)
Unlikely	0.076 (0.127)	0.005 (0.093)	-0.004 (0.059)	0.037 (0.086)
Unknwon	-0.106 (0.113)	-0.214* (0.124)	0.371*** (0.134)	-0.025 (0.086)
Likely	-0.051 (0.07)	-0.179* (0.097)	-0.146 (0.101)	-0.105 (0.131)
All Posteriors	-0.075* (0.039)	-0.199*** (0.045)	0.006 (0.047)	-0.066 (0.045)

Note. A separate linear regression has been estimated for any combination of posteriors and congestion levels. . Model (1) compares $ER_self_{i,TP,\mu,L}$ with $ER_self_{i,TS,\mu,L}$. Model (2) compares $ER_self_{i,TP,\mu,M}$ with $ER_self_{i,TS,\mu,M\&H}$. Model (2) compares $ER_self_{i,TP,\mu,H}$ with $ER_self_{i,TS,\mu,M\&H}$. Model (4) compares $ER_self_{i,TP,\mu,C}$ with $ER_self_{i,TS,\mu,C}$. Standard errors are robust to heteroskedasticity, having only 1 observation per id. Significance levels are indicated as follows: * p < 0.10, ** p < 0.05, *** p < 0.01.

B.6 Additional figures

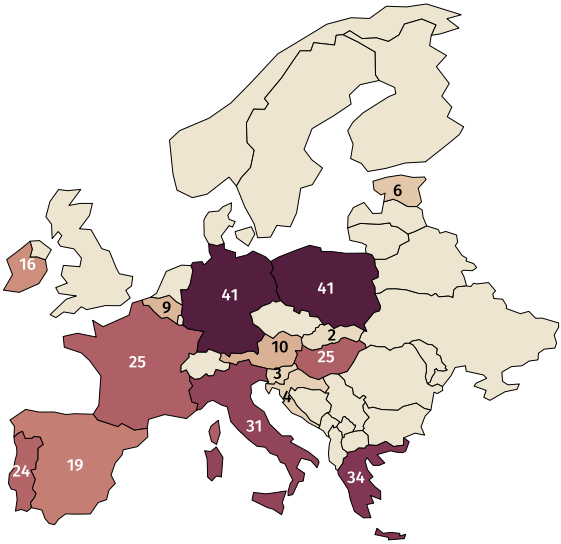


Figure B.5: Country of residence of the participants.

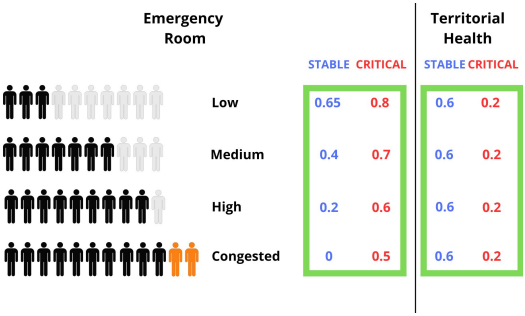


Figure B.6: Condition TC

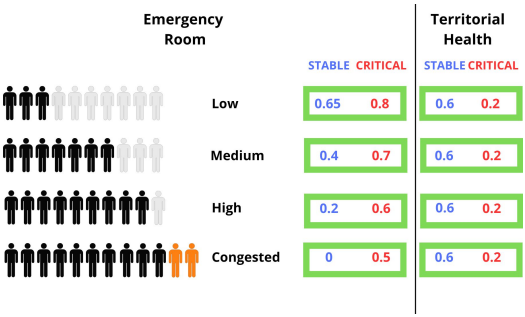


Figure B.7: Condition TP

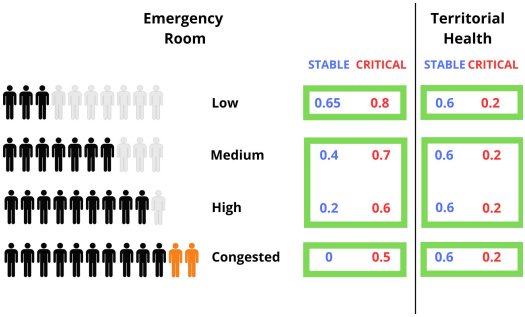


Figure B.8: Condition TS

Appendix C

Supplementary Material for Chapter 4

C.1 Supplementary Information

Ethical Approval

The study has obtained ethical approval from the Joint Ethical Committee of Scuola Normale Superiore, Sant'Anna School of Advanced Studies, and IMT School for Advanced Studies Lucca. The approval was granted on January 16^{th}, 2023, prior to the administration of the initial online survey. Significant emphasis was placed on maintaining the anonymity of the students throughout the research process.

Acknowledgments

We sincerely want to thank all the participants to the 5th International Conference on Higher Education Learning Methodologies and Technologies Online (HElMeTO) September 13-15, 2023 Foggia, for the insightful feedbacks and useful comments. This work has benefitted from the use of AI tools to improve the clarity and the readability of the final draft.

Bibliography

Aaskoven, Maiken Skovrider, Trine Kjær, and Dorte Gyrd-Hansen (Jan. 2022). "Effects of parental health shocks on children's school achievements: A register-based population study". In: *Journal of Health Economics* 81. ISSN: 18791646. DOI: 10.1016/j.jhealeco.2021.102573.

Abbey, James D. et al. (Jan. 2019). "Remanufacturing and consumers' risky choices: Behavioral modeling and the role of ambiguity aversion". In: *Journal of Operations Management* 65.1, pp. 4–21. ISSN: 0272-6963. DOI: 10.1002/joom.1001. URL: https://onlinelibrary.wiley.com/doi/10.1002/joom.1001.

Acker, Daniella and Nigel W. Duck (Apr. 2013). "Do Investors Suffer from Money Illusion? A Direct Test of the Modigliani–Cohn Hypothesis". In: *Review of Finance* 17.2, pp. 565–596. ISSN: 1573-692X. DOI: 10.1093/rof/rfs005. URL: https://academic.oup.com/rof/article/17/2/565/1598296.

Afilalo, Jonathan et al. (Dec. 2004). "Nonurgent Emergency Department Patient Characteristics and Barriers to Primary Care". In: *Academic Emergency Medicine* 11.12, pp. 1302–1310. ISSN: 1069-6563. DOI: 10.1197/j.aem.2004.08.032.

Ahuja, Vishal, Carlos A. Alvarez, and Bradley R. Staats (Sept. 2020). "Maintaining Continuity in Service: An Empirical Examination of Primary Care Physicians". In: *Manufacturing & Service Operations Management* 22.5, pp. 1088–1106. ISSN: 1523-4614. DOI: 10.1287/msom.2019.0808.

Alghamdi, Saleh and Natalia Beloff (July 2015). "Virtual currency concept: Its implementation, impacts and legislation". In: *2015 Science and Information Conference (SAI)*. IEEE, pp. 175–183. ISBN: 978-1-4799-8547-0. DOI: 10.1109/SAI.2015.7237142.

Anantharaman, V. (Apr. 2008). "Impact of health care system interventions on emergency department utilization and overcrowding in Singapore". In: *International Journal of Emergency Medicine* 1.1, pp. 11–20. ISSN: 1865-1372. DOI: 10.1007/s12245-008-0004-8.

Anvari, Farid, Dorina-Diana Verdeș, and Davide Marchiori (Jan. 2022). "Percent framing attenuates the magnitude effect in a preference-matching task of intertemporal choice". In: *PLOS ONE* 17.1. Ed. by David V. Smith, e0262620. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0262620. URL: https://dx.plos.org/10.1371/journal.pone.0262620.

Athey, Susan and Guido W. Imbens (May 2017). "The State of Applied Econometrics: Causality and Policy Evaluation". In: *Journal of Economic Perspectives* 31.2, pp. 3–32. ISSN: 0895-3309. DOI: 10.1257/jep.31.2.3. URL: https://pubs.aeaweb.org/doi/10.1257/jep.31.2.3.

Basiglio, Stefania ; Daniela ; Del Boca, and Chiara D Pronzato (2023). "The Impact of the "Coding Girls" Program on High School Students' Educational Choices". URL: www.iza.org.

Batteux, Eleonore, Eamonn Ferguson, and Richard J. Tunney (May 2019). "Do our risk preferences change when we make decisions for others? A meta-analysis of self-other differences in decisions involving risk". In: *PLOS ONE* 14.5. Ed. by Pablo Brañas-Garza, e0216566. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0216566. URL: https://dx.plos.org/10.1371/journal.pone.0216566.

Becker, David J. et al. (Oct. 2013). "Co-payments and the use of emergency department services in the Children'S Health Insurance Program". In: *Medical Care Research and Review* 70.5, pp. 514–530. ISSN: 10775587. DOI: 10.1177/1077558713491501.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964). "Measuring utility by a single-response sequential method." In: *Behavioral science* 9.3, pp. 226–232. ISSN: 00057940. DOI: 10.1002/bs.3830090304.

Beede, David et al. (2011). "Women in STEM: A Gender Gap to Innovation". In: *Economics and Statistics Administration Issue Brief*.

Bertoni, Andrea and Andrea Maffia (Apr. 2024). "Promoting creative insubordination using Escape Games in mathematics". In: *International Journal of Mathematical Education in Science and Technology* 55.4, pp. 1054–1064. ISSN: 0020-739X. DOI: 10.1080/0020739X.2022.2113468. URL: https://www.tandfonline.com/doi/full/10.1080/0020739X.2022.2113468.

Bilancini, Ennio, Leonardo Boncinelli, and Roberto Di Paolo (June 2023). "Game-based education promotes practices supporting sustainable water use". In: *Ecological Economics* 208, p. 107801. ISSN: 09218009. DOI: 10.1016/j.ecolecon.2023.107801. URL: https://linkinghub.elsevier.com/retrieve/pii/S0921800923000642.

Bird, Stephen R et al. (2007). "Integrated care facilitation for older patients with complex health care needs reduces hospital demand". In: *Australian Health Review* 31.3, p. 451. ISSN: 0156-5788. DOI: 10.1071/AH070451. URL: http://www.publish.csiro.au/?paper=AH070451.

Bisanti, Matteo et al. (Sept. 2022). "Digital Escape Games in Educational Programs for Financial Literacy". In: *European Conference on Games Based Learning* 16.1, pp. 666–674. ISSN: 2049-100X. DOI: 10.34190/ecgbl.16.1.857.

Böhm, Robert, Cornelia Betsch, and Lars Korn (Nov. 2016). "Selfish-rational non-vaccination: Experimental evidence from an interactive vaccination game". In: *Journal of Economic Behavior and Organization* 131, pp. 183–195. ISSN: 01672681. DOI: 10.1016/j.jebo.2015.11.008.

Bonde, Mads T et al. (July 2014). "Improving biotech education through gamified laboratory simulations". In: *Nature Biotechnology* 32.7, pp. 694–697. ISSN: 1087-0156. DOI: 10.1038/nbt.2955. URL: https://www.nature.com/articles/nbt.2955.

Botelho, Anabela et al. (July 2019). "Overestimation of health urgency as a cause for emergency services inappropriate use: Insights from an exploratory economics experiment in Portugal". In: *Health & Social Care in the Community* 27.4, pp. 1031–1041. ISSN: 0966-0410. DOI: 10.1111/hsc.12720. URL: https://onlinelibrary.wiley.com/doi/10.1111/hsc.12720.

Brand, Jennie E., Xiang Zhou, and Yu Xie (July 2023). "Recent Developments in Causal Inference and Machine Learning". In: *Annual Review of Sociology* 49.1, pp. 81–110. ISSN: 0360-0572. DOI: 10.1146/annurev-soc-030420-015345. URL: https://www.annualreviews.org/doi/10.1146/annurev-soc-030420-015345.

Breda, Thomas et al. (2020). "Do Female Role Models Reduce the Gender Gap in Science? Evidence from French High Schools". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3584926. URL: https://www.ssrn.com/abstract=3584926.

Brunnermeier, Markus K. and Christian Julliard (Jan. 2008). "Money Illusion and Housing Frenzies". In: *Review of Financial Studies* 21.1, pp. 135–180. ISSN: 0893-9454. DOI: 10.1093/rfs/hhm043.

Bystydzienski, Jill M., Margaret Eisenhart, and Monica Bruning (Mar. 2015). "High School Is Not Too Late: Developing Girls' Interest and Engagement in Engineering Careers". In: *The Career Development Quarterly* 63.1, pp. 88–95. ISSN: 0889-4019. DOI: 10.1002/j.2161-0045.2015.00097.x. URL: https://onlinelibrary.wiley.com/doi/10.1002/j.2161-0045.2015.00097.x.

Callaway, Brantly and Pedro H.C. Sant'Anna (Dec. 2021). "Difference-in-Differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230. ISSN: 03044076. DOI: 10.1016/j.jeconom.2020.12.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407620303948.

Camerer, Colin F. et al. (Mar. 2016). "Evaluating replicability of laboratory experiments in economics". In: *Science* 351.6280, pp. 1433–1436. ISSN: 0036-8075. DOI: 10.1126/science.aaf0918. URL: https://www.science.org/doi/10.1126/science.aaf0918.

Campbell, Jamie I.D. (Aug. 2005). *The Handbook of Mathematical Cognition*. Psychology Press, pp. 197–212. ISBN: 9781135423667. DOI: 10.4324/9780203998045. URL: https://www.taylorfrancis.com/books/9781135423667.

Cannon, Edmund and Giam Pietro Cipriani (2006). "Euro-Illusion: A Natural Experiment". In: *Journal of Money, Credit and Banking* 38.5, pp. 1391–1403. URL: https://www.jstor.org/stable/3839012.

Caplan, Gideon A. et al. (Sept. 2004). "A Randomized, Controlled Trial of Comprehensive Geriatric Assessment and Multidisciplinary Intervention After Discharge of Elderly from the Emergency Department—The DEED II Study". In: *Journal of the American Geriatrics Society* 52.9, pp. 1417–1423. ISSN: 0002-8614. DOI: 10.1111/j.1532-5415.2004.52401.x. URL: https://agsjournals.onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.2004.52401.x.

Carret, Maria L.V., Anaclaudia G. Fassa, and Ichiro Kawachi (2007). "Demand for emergency health service: Factors associated with inappropriate use". In: *BMC Health Services Research* 7. ISSN: 14726963. DOI: 10.1186/1472-6963-7-131.

Carret, Maria Laura Vidal, Ana Claudia Gastal Fassa, and Marlos Rodrigues Domingues (Jan. 2009). "Inappropriate use of emergency services: a systematic review of prevalence and associated factors". In: *Cadernos de Saúde Pública* 25.1, pp. 7–28. ISSN: 0102-311X. DOI: 10.1590/S0102-311X2009000100002. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X200900010000
lng=en&tlng=en.

Catola, Marco et al. (Aug. 2023). "Multilevel public goods game: Levelling up, substitution and crowding-in effects". In: *Journal of Economic Psychology* 97, p. 102626. ISSN: 01674870. DOI: 10.1016/j.joep.2023.102626.

Cerulli-harms, Annette et al. (2020). *Loot boxes in online games and their effect on consumers, in particular young consumers*. Tech. rep.

Chakravarty, Sujoy et al. (Apr. 2011). "Are You Risk Averse over Other People's Money?" In: *Southern Economic Journal* 77.4, pp. 901–913. ISSN: 0038-4038. DOI: 10.4284/0038-4038-77.4.901. URL: https://onlinelibrary.wiley.com/doi/10.4284/0038-4038-77.4.901.

Chang, Andrew C. and Phillip Li (Oct. 2015). "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"". In: *Finance and Economics Discussion Series* 2015.0.83, pp. 1–26. ISSN: 1936-2854. DOI: 10.17016/feds.2015.083. URL: https://www.federalreserve.gov/econres/feds/is-economics-research-replicable-sixty-published-papers-from-thirteen-journals-say-quotusually-notquot.htm.

Chiarello, Fabio et al. (n.d.). *The Language of Life: A Game-Based Workshop for Introducing Protein Biosynthesis*. Tech. rep.

Coenen, Olaf and Hendrik Lesser (2021). *Key Facts 2020 The Year We Played Togheter*. Tech. rep. ISFE (Europe's Video Game Industry), EGDF (European Games Developer Federation).

Cohen, R. B., C. Polk, and T. Vuolteenaho (May 2005). "Money Illusion in the Stock Market: The Modigliani-Cohn Hypothesis". In: *The Quarterly Journal of Economics* 120.2, pp. 639–668. ISSN: 0033-5533. DOI: 10.1093/qje/120.2.639. URL: https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/120.2.639.

Coleman, P (Nov. 2001). "Will alternative immediate care services reduce demands for non-urgent treatment at accident and emergency?" In: *Emergency Medicine Journal* 18.6, pp. 482–487. ISSN: 14720205. DOI: 10.1136/emj.18.6.482. URL: https://emj.bmj.com/lookup/doi/10.1136/emj.18.6.482.

Cornel, Cariana J., Dale C. Rowe, and Caralea M. Cornel (Sept. 2017). "Starships and Cybersecurity: Teaching Security Concepts through Immersive Gaming Experiences". In: *Proceedings of the 18th Annual Conference on Information Technology Education*. New York, NY, USA: ACM, pp. 27–32. ISBN: 9781450351003. DOI: 10.1145/3125659.3125696.

Costa, Gabriele et al. (Aug. 2023). "Why Mary Can Hack: Effectively Introducing High School Girls to Cybersecurity". In: *Proceedings of the 18th International Conference on Availability, Reliability and Security*. New York, NY, USA: ACM, pp. 1–8. ISBN: 9798400707728. DOI: 10.1145/3600160.3605009. URL: https://dl.acm.org/doi/10.1145/3600160.3605009.

Darnell, Jeffrey C et al. (1985). "After-Hours Telephone Access to Physicians with Access to Computerized Medical Records: Experience in an Inner-City General Medicine Clinic". In: *Medical Care* 23.1, pp. 20–26. URL: https://www.jstor.org/stable/3765022.

Daysal, N. Meltem et al. (Feb. 2024). *Do Medical Treatments Work for Work? Evidence from Breast Cancer Patients*. Tech. rep. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w32167. URL: http://www.nber.org/papers/w32167.pdf.

Decker, Simon and Hendrik Schmitz (Dec. 2016). "Health shocks and risk aversion". In: *Journal of Health Economics* 50, pp. 156–170. ISSN: 01676296. DOI: 10.1016/j.jhealeco.2016.09.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167629616302260.

Delfabbro, Paul and Daniel L. King (Sept. 2020). "Gaming-gambling convergence: evaluating evidence for the 'gateway' hypothesis". In: *International Gambling Studies* 20.3, pp. 380–392. ISSN: 1445-9795. DOI: 10.1080/14459795.2020.1768430. URL: https://www.tandfonline.com/doi/full/10.1080/14459795.2020.1768430.

DellaVigna, Stefano and Elizabeth Linos (2022). "RCTs to Scale: Comprehensive Evidence From Two Nudge Units". In: *Econometrica* 90.1, pp. 81–116. ISSN: 0012-9682. DOI: 10.3982/ecta18709.

DeSalvo, Alyson et al. (Mar. 2000). "Patient education and emergency room visits". In: *British Journal of Clinical Governance* 5.1, pp. 35–38. ISSN: 1466-4100. DOI: 10.1108/14664100010333017. URL: https://www.emerald.com/insight/content/doi/10.1108/14664100010333017/full/html.

Di Paolo, Roberto and Veronica Pizziol (June 2024). "Gamification and Sustainable Water Use: The Case of the BLUTUBE Educational Program". In: *Simulation & Gaming* 55.3, pp. 391–417. ISSN: 1046-8781. DOI: 10.1177/10468781231181652. URL: https://journals.sagepub.com/doi/10.1177/10468781231181652.

Drichoutis, Andreas C., Jayson L. Lusk, and Rodolfo M. Nayga (Dec. 2015). "The veil of experimental currency units in second price auc-

tions". In: *Journal of the Economic Science Association* 1.2, pp. 182–196. ISSN: 2199-6776. DOI: 10.1007/s40881-015-0014-2.

Drummond, Aaron and James D. Sauer (June 2018). "Video game loot boxes are psychologically akin to gambling". In: *Nature Human Behaviour* 2.8, pp. 530–532. ISSN: 2397-3374. DOI: 10.1038/s41562-018-0360-1. URL: https://www.nature.com/articles/s41562-018-0360-1.

Dupoux M. et al. (2025). *Unlocking the full potential of behavioural insights for policy: From influ-encing the individual to shaping the system*. Tech. rep. Luxemburg: Publications Office of the European Union: European Commission: Joint Research Centre. DOI: 10.2760/7367599. URL: https://data.eu-ropa.eu/doi/10.2760/7367599.

Duvendack, Maren, Richard Palmer-Jones, and W. Robert Reed (May 2017). "What Is Meant by "Replication" and Why Does It Encounter Resistance in Economics?" In: *American Economic Review* 107.5, pp. 46–51. ISSN: 0002-8282. DOI: 10.1257/aer.p20171031.

Eisenberger, Roselies and Martin Weber (1995). "Willingness-to-Pay and Willingness-to-Accept for Risky and Ambiguous Lotteries". In: *Journal of Risk and Uncertainty* 10, pp. 223–233. URL: https://www.jstor.org/stable/41760767.

Ellsberg, Daniel (Nov. 1961). "Risk, Ambiguity, and the Savage Axioms". In: *The Quarterly Journal of Economics* 75.4, p. 643. ISSN: 00335533. DOI: 10.2307/1884324. URL: https://www.jstor.org/stable/1884324.

Eriksen, Kristoffer W., Ola Kvaløy, and Miguel Luzuriaga (June 2020). "Risk-taking on behalf of others". In: *Journal of Behavioral and Experimental Finance* 26, p. 100283. ISSN: 22146350. DOI: 10.1016/j.jbef.2020.100283. URL: https://linkinghub.elsevier.com/retrieve/pii/S221463502030023X.

Estey, Angela et al. (2003). "Understanding the causes of overcrowding in emergency departments in the Capital Health Region in Alberta: A focus group study". In: *Canadian Journal of Emergency Medicine* 5.2, pp. 87–94. ISSN: 14818043. DOI: 10.1017/S1481803500008216.

European Central Bank (2015). *Virtual currency schemes-a further analysis*. Tech. rep. Frankfurt am Main: European Central Bank. DOI: 10.2866/662172. URL: www.ecb.europa.eu.

European Parliament and The Council of the European Union (2015). *Unfair Commercial Practices Directive*. Tech. rep. Official Journal of the European Union.

Falk, Armin, Anke Becker, et al. (Nov. 2018). "Global evidence on economic preferences". In: *Quarterly Journal of Economics* 133.4, pp. 1645–1692. ISSN: 15314650. DOI: 10.1093/qje/qjy013.

Falk, Armin and James J. Heckman (Oct. 2009). "Lab Experiments Are a Major Source of Knowledge in the Social Sciences". In: *Science* 326.5952, pp. 535–538. ISSN: 0036-8075. DOI: 10.1126/science.1168244. URL: https://www.science.org/doi/10.1126/science.1168244.

Fehr, Ernst and Simon Gächter (Sept. 2000). "Cooperation and Punishment in Public Goods Experiments". In: *American Economic Review* 90.4, pp. 980–994. ISSN: 0002-8282. DOI: 10.1257/aer.90.4.980.

Fehr, Ernst and Jean-Robert Tyran (Dec. 2001). "Does Money Illusion Matter?" In: *American Economic Review* 91.5, pp. 1239–1262. ISSN: 0002-8282. DOI: 10.1257/aer.91.5.1239. URL: https://pubs.aeaweb.org/doi/10.1257/aer.91.5.1239.

— (Feb. 2007). "Money illusion and coordination failure". In: *Games and Economic Behavior* 58.2, pp. 246–268. ISSN: 08998256. DOI: 10.1016/j.geb.2006.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0899825606000649.

Flores-Mateo, Gemma et al. (May 2012). "Effectiveness of Organizational Interventions to Reduce Emergency Department Utilization: A Systematic Review". In: *PLoS ONE* 7.5. Ed. by Joseph S. Ross, e35903. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0035903. URL: https://dx.plos.org/10.1371/journal.pone.0035903.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann (Apr. 2022). "Individual-level loss aversion in riskless and risky choices". In: *Theory and Decision* 92.3-4, pp. 599–624. ISSN: 15737187. DOI: 10.1007/s11238-021-09839-8.

Gamble, Amelie (Aug. 2006). "Euro illusion or the reverse? Effects of currency and income on evaluations of prices of consumer products". In: *Journal of Economic Psychology* 27.4, pp. 531–542. ISSN: 01674870. DOI: 10.1016/j.joep.2006.01.006.

Gamble, Amelie et al. (2002). "Euro Illusion: Psychological Insightsinto Price Evaluations with a UnitaryCurrency". In: *European Psychologist* 7.4, pp. 302–311. DOI: 10.1027/1016-9040.7.4.302.

Garcia-Retamero, Rocio and Mirta Galesic (Mar. 2012). "Doc, what would you do if you were me? On self–other discrepancies in medical decision making." In: *Journal of Experimental Psychology: Applied* 18.1, pp. 38–51. ISSN: 1939-2192. DOI: 10.1037/a0026018. URL: https://doi.apa.org/doi/10.1037/a0026018.

Garrafa, Emirena et al. (Dec. 2020). "When fear backfires: Emergency department accesses during the Covid-19 pandemic". In: *Health Policy* 124.12, pp. 1333–1339. ISSN: 01688510. DOI: 10.1016/j.healthpol.2020.10.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S016885102030261X.

He, Tai-Sen and Fuhai Hong (Dec. 2018). "Risk breeds risk aversion". In: *Experimental Economics* 21.4, pp. 815–835. ISSN: 1386-4157. DOI: 10.1007/s10683-017-9553-0.

Holin Lin and Chuen Tsai Sun (May 2011). "Cash trade in free-to-play online games". In: *Games and Culture* 6.3, pp. 270–287. ISSN: 15554120. DOI: 10.1177/1555412010364981.

Holt, Charles A and Susan K Laury (Nov. 2002). "Risk Aversion and Incentive Effects". In: *American Economic Review* 92.5, pp. 1644–1655. ISSN: 0002-8282. DOI: 10.1257/000282802762024700. URL: https://pubs.aeaweb.org/doi/10.1257/000282802762024700.

Huang, Yunhui, Kai H. Lim, and Zhijie Lin (Mar. 2021). "Leveraging the numerosity effect to influence perceived expensiveness of virtual items". In: *Information Systems Research* 32.1, pp. 93–114. ISSN: 15265536. DOI: 10.1287/isre.2020.0955.

Imbo, Ineke, Sandrine Duverne, and Patrick Lemaire (Sept. 2007). "Working memory, strategy execution, and strategy selection in mental arithmetic". In: *Quarterly Journal of Experimental Psychology* 60.9, pp. 1246–1264. ISSN: 17470218. DOI: 10.1080/17470210600943419.

Inchingolo, Giannandrea et al. (2023). "PIXEL: Challenges of Designing a Professional Board Game for Astronomy Education". In: *Proceedings of the 17th European Conference on Game-Based Learning: ECGBL 2023*. Academic Conferences and publishing limited.

Jin, Ge et al. (Feb. 2018). "Game based Cybersecurity Training for High School Students". In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. New York, NY, USA: ACM, pp. 68–73. ISBN: 9781450351034. DOI: 10.1145/3159450.3159591.

Jung, Hyemin et al. (Nov. 2014). "The impact of an emergency fee increase on the composition of patients visiting emergency departments". In: *Journal of Preventive Medicine and Public Health* 47.6, pp. 309–316. ISSN: 19758375. DOI: 10.3961/jpmph.14.044.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (Feb. 1991). "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias". In: *Journal of Economic Perspectives* 5.1, pp. 193–206. ISSN: 0895-3309. DOI: 10.1257/jep.5.1.193. URL: https://pubs.aeaweb.org/doi/10.1257/jep.5.1.193.

Kahneman, Daniel and Amos Tversky (Mar. 1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2, p. 263. ISSN: 00129682. DOI: 10.2307/1914185. URL: https://www.jstor.org/stable/1914185?origin=crossref.

Kalogiannakis, Michail, Stamatios Papadakis, and Alkinoos-Ioannis Zourmpakis (Jan. 2021). "Gamification in Science Education. A Systematic Review of the Literature". In: *Education Sciences* 11.1, p. 22. ISSN: 2227-7102. DOI: 10.3390/educsci11010022. URL: https://www.mdpi.com/2227-7102/11/1/22.

Kerr, Norbert L. (Aug. 1998). "HARKing: Hypothesizing After the Results are Known". In: *Personality and Social Psychology Review* 2.3, pp. 196–217. ISSN: 1088-8683. DOI: 10.1207/s15327957pspr0203{_}4.

King, Daniel L. and Paul H. Delfabbro (Nov. 2018). "Predatory monetization schemes in video games (e.g. 'loot boxes') and internet gaming disorder". In: *Addiction* 113.11, pp. 1967–1969. ISSN: 0965-2140. DOI: 10.1111/add.14286. URL: https://onlinelibrary.wiley.com/doi/10.1111/add.14286.

King, Daniel L., Paul H. Delfabbro, et al. (Dec. 2019). "Unfair play? Video games as exploitative monetized services: An examination of game patents from a consumer protection perspective". In: *Computers in Human Behavior* 101, pp. 131–143. ISSN: 07475632. DOI: 10.1016/j.chb.2019.07.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0747563219302602.

Lattimer, V. (Apr. 2000). "Cost analysis of nurse telephone consultation in out of hours primary care: evidence from a randomised controlled trial". In: *BMJ* 320.7241, pp. 1053–1057. ISSN: 09598138. DOI: 10.1136/bmj.320.7241.1053. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.320.7241.1053.

Lattimer, V. et al. (Oct. 1998). "Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial". In: *BMJ* 317.7165, pp. 1054–1059. ISSN: 0959-8138. DOI: 10.1136/bmj.317.7165.1054. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.317.7165.1054.

Leahy, D. (Sept. 2022). "Rocking the Boat: Loot Boxes in Online Digital Games, the Regulatory Challenge, and the EU's Unfair Commercial Practices Directive". In: *Journal of Consumer Policy* 45.3, pp. 561–592. ISSN: 15730700. DOI: 10.1007/s10603-022-09522-7.

Lega, Federico and Alessandro Mengoni (Dec. 2008). "Why non-urgent patients choose emergency over primary care services? Empirical ev-

idence and managerial implications". In: *Health Policy* 88.2-3, pp. 326–338. ISSN: 01688510. DOI: 10.1016/j.healthpol.2008.04.005.
Lemaire, P., M. Lecacheur, and M. Ferréol-Barbey (Apr. 2001). "Strategy Use in Currency Conversion Tasks: The Examples of Euro into French Franc and French Franc into Euro Conversions". In: *Current psychology letters* 2001/1, 4. ISSN: 1376-2095. DOI: 10.4000/cpl.176.
Lemaire, Patrick (Dec. 2007). "The Impact of the Euro Changeover on Between-Currency Conversions". In: *Journal of Consumer Policy* 30.4, pp. 383–391. ISSN: 0168-7034. DOI: 10.1007/s10603-007-9045-2. URL: http://link.springer.com/10.1007/s10603-007-9045-2.
Lemaire, Patrick and Mireille Lecacheur (2001). "Older and younger adults' strategy use and execution in currency conversion tasks: Insights from French franc to euro and euro to French franc conversions". In: *Journal of Experimental Psychology: Applied* 7.3, pp. 195–206. ISSN: 1076898X. DOI: 10.1037/1076-898X.7.3.195.
Leune, Kees and Salvatore J. Petrilli (Sept. 2017). "Using Capture-the-Flag to Enhance the Effectiveness of Cybersecurity Education". In: *Proceedings of the 18th Annual Conference on Information Technology Education*. New York, NY, USA: ACM, pp. 47–52. ISBN: 9781450351003. DOI: 10.1145/3125659.3125686.
Liu, Fang and Panagiotakos Demosthenes (Dec. 2022). *Real-world data: a brief review of the methods, applications, challenges and opportunities*. DOI: 10.1186/s12874-022-01768-6.
Lowe, Robert A, Rongwei Fu, and Charles A Gallia (2010). *Impact of Policy Changes on Emergency Department Use by Medicaid Enrollees in Oregon*. Tech. rep. 7, pp. 619–627.
Macey, Joseph and Juho Hamari (Jan. 2019). "eSports, skins and loot boxes: Participants, practices and problematic behaviour associated with emergent forms of gambling". In: *New Media and Society* 21.1, pp. 20–41. ISSN: 14617315. DOI: 10.1177/1461444818786216.
Maffia, Andrea and Liliana Silva (May 2022). "Teachers' Struggling in Identifying the Semiotic Potential of Mathematical Board Games". In: *Handbook of Research on International Approaches and Practices for Gamifying Mathematics*. IGI Global, pp. 354–373. DOI: 10.4018/978-1-7998-9660-9.ch017. URL: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-9660-9.ch017.
Maier, Maximilian et al. (Aug. 2022). "No evidence for nudging after adjusting for publication bias". In: *Proceedings of the National Academy of*

Sciences 119.31. ISSN: 0027-8424. DOI: 10.1073/pnas.2200300119. URL: https://pnas.org/doi/full/10.1073/pnas.2200300119.

Martorana, Antonio (2022). "Investigating the Experiences of Female CTF Players". PhD thesis.

Mattiassi, Alan, Andrea Ghirarduzzi, and Giovanni Bacaro (2023). "L'Universal Game Design for Learning: una metodologia innovativa da integrare nei curricola scolastici e universitari". In: *QuaderniCIRD. Rivista del Centro Interdipartimentale per la Ricerca Didattica dell'Università di Trieste Journal of the Interdepartmental Center for Educational Research of the University of Trieste-ISSN: 2039-8646*. URL: https://www.openstarts.units.it/entities/publication/91999384-93da-4efb-a9b7-bed3feb10eed/details.

Mazar, Nina, On Amir, and Dan Ariely (Dec. 2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance". In: *Journal of Marketing Research* 45.6, pp. 633–644. ISSN: 0022-2437. DOI: 10.1509/jmkr.45.6.633. URL: http://journals.sagepub.com/doi/10.1509/jmkr.45.6.633.

Mazar, Nina, Hilke Plassmann, and Nicole Robitaille (2016). "Pain of Paying? A Metaphor Gone Literal: Evidence from Neural and Behavioral Science". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2901808. URL: http://www.ssrn.com/abstract=2901808.

McDaniel, Lucas, Erik Talvi, and Brian Hay (Jan. 2016). "Capture the Flag as Cyber Security Introduction". In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, pp. 5479–5486. ISBN: 978-0-7695-5670-3. DOI: 10.1109/HICSS.2016.677.

McKinstry, Brian et al. (2002). *Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices*. Tech. rep.

Mertens, Stephanie et al. (Jan. 2022). "The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains". In: *Proceedings of the National Academy of Sciences* 119.1. ISSN: 0027-8424. DOI: 10.1073/pnas.2107346118. URL: https://pnas.org/doi/full/10.1073/pnas.2107346118.

Miller, Klaus M. et al. (Feb. 2011). "How Should Consumers' Willingness to Pay be Measured? An Empirical Comparison of State-of-the-Art Approaches". In: *Journal of Marketing Research* 48.1, pp. 172–184. ISSN: 0022-2437. DOI: 10.1509/jmkr.48.1.172.

Montinari, Natalia ; and Michaela Rancan (2013). "Social preferences under risk: The role of social distance". URL: https://hdl.handle.net/10419/98455.

Morales, Antonio J. and Enrique Fatas (Dec. 2021). "Price competition and nominal illusion: experimental evidence and a behavioural model". In: *SERIEs* 12.4, pp. 607–632. ISSN: 18694195. DOI: 10.1007/s13209-021-00235-7.

Morgan, Sofie Rahman et al. (Oct. 2013). "Non–Emergency Department Interventions to Reduce <scp>ED</scp> Utilization: A Systematic Review". In: *Academic Emergency Medicine* 20.10. Ed. by Shahriar Zehtabchi, pp. 969–985. ISSN: 1069-6563. DOI: 10.1111/acem.12219. URL: https://onlinelibrary.wiley.com/doi/10.1111/acem.12219.

Morley, Claire et al. (Aug. 2018). "Emergency department crowding: A systematic review of causes, consequences and solutions". In: *PLOS ONE* 13.8. Ed. by Fernanda Bellolio, e0203316. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0203316. URL: https://dx.plos.org/10.1371/journal.pone.0203316.

Nesti, Romina (2017). "Game-Based Learning". In: *Gioco e progettazione ludica in educazione. Pisa: Edizioni ETS*.

Oliveira, Wilk et al. (Jan. 2023). "Tailored gamification in education: A literature review and future agenda". In: *Education and Information Technologies* 28.1, pp. 373–406. ISSN: 1360-2357. DOI: 10.1007/s10639-022-11122-4. URL: https://link.springer.com/10.1007/s10639-022-11122-4.

Osborne, Charlie and Steve Morgan (2022). *WOMEN IN CYBERSECURITY 2022 REPORT*. Tech. rep.

Palan, Stefan and Christian Schitter (Mar. 2018). "Prolific.ac—A subject pool for online experiments". In: *Journal of Behavioral and Experimental Finance* 17, pp. 22–27. ISSN: 22146369. DOI: 10.1016/j.jbef.2017.12.004.

Park, Joowon, Clarence Lee, and Manoj Thomas (Jan. 2021). "Why do cashless payments increase unhealthy consumption? The decision-risk inattention hypothesis". In: *Journal of the Association for Consumer Research* 6.1, pp. 21–32. ISSN: 23781823. DOI: 10.1086/710251.

Pesare, Enrica et al. (Dec. 2016). "Game-based learning and Gamification to promote engagement and motivation in medical learning contexts". In: *Smart Learning Environments* 3.1, p. 5. ISSN: 2196-7091. DOI: 10.1186/s40561-016-0028-0. URL: http://slejournal.

springeropen.com/articles/10.1186/s40561-016-0028-0.

Peter Vedsted, Morten Bondo Christe (Jan. 2001). "The effect of an out-of-hours reform on attendance at casualty wards. The Danish example". In: *Scandinavian Journal of Primary Health Care* 19.2, pp. 95–98. ISSN: 0281-3432. DOI: 10.1080/028134301750235303. URL: http://www.tandfonline.com/doi/full/10.1080/028134301750235

Petrovskaya, Elena and David Zendle (2021). "Predatory Monetisation? A Categorisation of Unfair, Misleading and Aggressive Monetisation Techniques in Digital Games from the Player Perspective". In: *Journal of Business Ethics*. ISSN: 15730697. DOI: 10.1007/s10551-021-04970-6.

Pines, Jesse M. and Richard T. Griffey (Aug. 2015). "What We Have Learned From a Decade of ED Crowding Research". In: *Academic Emergency Medicine* 22.8, pp. 985–987. ISSN: 10696563. DOI: 10.1111/acem.12716.

Polman, Evan (Nov. 2012). "Self–other decision making and loss aversion". In: *Organizational Behavior and Human Decision Processes* 119.2, pp. 141–150. ISSN: 07495978. DOI: 10.1016/j.obhdp.2012.06.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0749597812000775.

Prelec, Drazen and George Loewenstein (Feb. 1998). "The Red and the Black: Mental Accounting of Savings and Debt". In: *Marketing Science* 17.1, pp. 4–28. ISSN: 0732-2399. DOI: 10.1287/mksc.17.1.4. URL: https://pubsonline.informs.org/doi/10.1287/mksc.17.1.4.

Qualtrics (Sept. 2005). *Qualtrics*. Provo, Utah, USA.

Raghubir, Priya and Joydeep Srivastava (Dec. 2002). "Effect of Face Value on Product Valuation in Foreign Currencies". In: *Journal of Consumer Research* 29.3, pp. 335–347. ISSN: 0093-5301. DOI: 10.1086/344430. URL: https://academic.oup.com/jcr/article-lookup/doi/10.1086/344430.

Randazzo, Maria Pia (Apr. 2024). *Lo stato dell'emergenza urgenza alla luce del DM77*. Tech. rep. Roma: AGENAS.

Rector, T S, P J Venus, and A J Laine (Dec. 1999). "Impact of mailing information about nonurgent care on emergency department visits by Medicaid beneficiaries enrolled in managed care." In: *The American journal of managed care* 5.12, pp. 1505–12. ISSN: 1088-0224. URL: http://www.ncbi.nlm.nih.gov/pubmed/11066617.

Reinstein, David and Gerhard Riener (Mar. 2012). "Decomposing desert and tangibility effects in a charitable giving experiment". In: *Experimental Economics* 15.1, pp. 229–240. ISSN: 13864157. DOI: 10.1007/s10683-011-9298-0.

Ricciardi, Walter and Rosanna Tarricone (Dec. 2021). "The evolution of the Italian National Health Service". In: *The Lancet* 398.10317, pp. 2193–2206. ISSN: 01406736. DOI: 10.1016/S0140-6736(21)01733-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673621017335.

Richards, D. A (Nov. 2002). "Nurse telephone triage for same day appointments in general practice: multiple interrupted time series trial of effect on workload and costs". In: *BMJ* 325.7374, pp. 1214–1214. ISSN: 09598138. DOI: 10.1136/bmj.325.7374.1214. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.325.7374.1214.

Rubin, Mark (May 2020). "Does preregistration improve the credibility of research findings?" In: *The Quantitative Methods for Psychology* 16.4, pp. 376–390. ISSN: 2292-1354. DOI: 10.20982/tqmp.16.4.p376.

Saayman, Andrea et al. (2022). "Money illusion under tourists: deceived by larger numbers?" In: *Current Issues in Tourism* 25.5, pp. 792–807. ISSN: 13683500. DOI: 10.1080/13683500.2021.1887825.

Salminen, Mikko, Simo Järvelä, and Niklas Ravaja (2018). "Economic decision-making in free-to-play games: A laboratory experiment to study the effects of currency conversion". In: *GamiFIN Conference 2018*.

Sánchez-Martín, Jesús, Florentina Cañada-Cañada, and María Antonia Dávila-Acedo (Dec. 2017). "Just a game? Gamifying a general science class at university". In: *Thinking Skills and Creativity* 26, pp. 51–59. ISSN: 18711871. DOI: 10.1016/j.tsc.2017.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1871187117300585.

Scott, John C. et al. (Sept. 2004). "Effectiveness of a Group Outpatient Visit Model for Chronically Ill Older Health Maintenance Organization Members: A 2-Year Randomized Trial of the Cooperative Health Care Clinic". In: *Journal of the American Geriatrics Society* 52.9, pp. 1463–1470. ISSN: 0002-8614. DOI: 10.1111/j.1532-5415.2004.52408.x. URL: https://agsjournals.onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.2004.52408.x.

Selby, Joe V., Bruce H. Fireman, and Bix E. Swain (Mar. 1996). "Effect of a Copayment on Use of the Emergency Department in a Health Maintenance Organization". In: *New England Journal of Medicine* 334.10, pp. 635–642. ISSN: 0028-4793. DOI: 10.1056/NEJM199603073341006. URL: http://www.nejm.org/doi/abs/10.1056/NEJM199603073341006.

Shafir, E., P. Diamond, and A. Tversky (May 1997). "Money Illusion". In: *The Quarterly Journal of Economics* 112.2, pp. 341–374. ISSN: 0033-5533. DOI: 10.1162/003355397555208. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/003355397555208.

Shalowitz, David I., Elizabeth Garrett-Mayer, and David Wendler (Mar. 2006). "The Accuracy of Surrogate Decision Makers". In: *Archives of Internal Medicine* 166.5, p. 493. ISSN: 0003-9926. DOI: 10.1001/archinte.166.5.493. URL: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.166.5.493.

Shen, Junyi and Hiromasa Takahashi (Dec. 2013). "A cash effect in ultimatum game experiments". In: *Journal of Socio-Economics* 47, pp. 94–102. ISSN: 10535357. DOI: 10.1016/j.socec.2013.09.005.

Stern, Jeff, Emily Reid, and Kari Bancroft (Feb. 2015). "Teaching Introductory Computer Science for a Diverse Student Body". In: *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. New York, NY, USA: ACM, pp. 705–705. ISBN: 9781450329668. DOI: 10.1145/2676723.2678291.

Stone, Eric R., Andrew J. Yates, and Allison S. Caruthers (Sept. 2002). "Risk Taking in Decision Making for Others Versus the Self ¹". In: *Journal of Applied Social Psychology* 32.9, pp. 1797–1824. ISSN: 0021-9029. DOI: 10.1111/j.1559-1816.2002.tb00260.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2002.tb00260.x.

Stone, Eric R., Choi YoonSun, et al. (May 2013). "I can take the risk, but you should be safe: Self-other differences in situations involving physical safety". In: *Judgment and Decision Making* 8.3, pp. 250–267. ISSN: 19302975. DOI: 10.1017/s1930297500005969.

Su, Yuliu et al. (2021). "Nonurgent Patients' Preferences for Emergency Department Versus General Practitioner and Effects of Incentives: A Discrete Choice Experiment". In: *MDM Policy and Practice* 6.2. ISSN: 23814683. DOI: 10.1177/23814683211027552.

Sun, Shujing, Susan F. Lu, and Huaxia Rui (Sept. 2020). "Does Telemedicine Reduce Emergency Room Congestion? Evidence from New York State". In: *Information Systems Research* 31.3, pp. 972–986. ISSN: 1047-7047. DOI: 10.1287/isre.2020.0926. URL: https://pubsonline.informs.org/doi/10.1287/isre.2020.0926.

Švábenský, Valdemar et al. (Mar. 2021). "Cybersecurity knowledge and skills taught in capture the flag challenges". In: *Computers & Security* 102, p. 102154. ISSN: 01674048. DOI: 10.1016/j.cose.2020.

102154. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404820304272.

Swacha, Jakub (Feb. 2021). "State of Research on Gamification in Education: A Bibliometric Survey". In: *Education Sciences* 11.2, p. 69. ISSN: 2227-7102. DOI: 10.3390/educsci11020069. URL: https://www.mdpi.com/2227-7102/11/2/69.

Teger, Allan I. and Nathan Kogan (Sept. 1975). "Decision-making for Others under Reciprocal and Non-reciprocal Conditions". In: *British Journal of Social and Clinical Psychology* 14.3, pp. 215–222. ISSN: 0007-1293. DOI: 10.1111/j.2044-8260.1975.tb00174.x. URL: https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/j.2044-8260.1975.tb00174.x.

Toccafondi, Niccolò, Roberto Di Paolo, and Sibilla Di Guida (2025). *Virtual Currencies in Online Gaming Increase the Willingness to Pay for Loot Boxes: An Experimental Analysis*. DOI: 10.2139/ssrn.5146007.

Tversky, Amos and Daniel Kahneman (Sept. 1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. ISSN: 0036-8075. DOI: 10.1126/science.185.4157.1124.

Uscher-Pines, Lori et al. (Jan. 2013). "Emergency department visits for nonurgent conditions: systematic literature review." In: *The American journal of managed care* 19.1, pp. 47–59. ISSN: 1936-2692. URL: http://www.ncbi.nlm.nih.gov/pubmed/23379744%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4156292.

Van den Heede, Koen and Carine Van de Voorde (Dec. 2016). "Interventions to reduce emergency department utilisation: A review of reviews". In: *Health Policy* 120.12, pp. 1337–1349. ISSN: 01688510. DOI: 10.1016/j.healthpol.2016.10.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168851016302688.

Van Der Linden, M. Christien et al. (Nov. 2017). "Two emergency departments, 6000 km apart: Differences in patient flow and staff perceptions about crowding". In: *International Emergency Nursing* 35, pp. 30–36. ISSN: 1755599X. DOI: 10.1016/j.ienj.2017.06.002.

Vlasceanu, Madalina et al. (Feb. 2024). "Addressing climate change with behavioral science: A global intervention tournament in 63 countries". In: *Science Advances* 10.6, p. 217. ISSN: 2375-2548. DOI: 10.1126/sciadv.adj5778. URL: https://www.science.org/doi/10.1126/sciadv.adj5778.

Wang, Ming-Te and Jessica L. Degol (Mar. 2017). "Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowl-

edge, Implications for Practice, Policy, and Future Directions". In: *Educational Psychology Review* 29.1, pp. 119–140. ISSN: 1040-726X. DOI: 10.1007/s10648-015-9355-x. URL: http://link.springer.com/10.1007/s10648-015-9355-x.

Zahedi, Leila et al. (Aug. 2021). "Gamification in education: a mixed-methods study of gender on computer science students' academic performance and identity development". In: *Journal of Computing in Higher Education* 33.2, pp. 441–474. ISSN: 1042-1726. DOI: 10.1007/s12528-021-09271-5. URL: https://link.springer.com/10.1007/s12528-021-09271-5.

Zendle, David and Paul Cairns (Nov. 2018). "Video game loot boxes are linked to problem gambling: Results of a large-scale survey". In: *PLoS ONE* 13.11. ISSN: 19326203. DOI: 10.1371/journal.pone.0206767.

Zendle, David, Paul Cairns, et al. (Jan. 2020). "Paying for loot boxes is linked to problem gambling, regardless of specific features like cash-out and pay-to-win". In: *Computers in Human Behavior* 102, pp. 181–191. ISSN: 07475632. DOI: 10.1016/j.chb.2019.07.003.

Zendle, David, Rachel Meyer, and Harriet Over (June 2019). "Adolescents and loot boxes: Links with problem gambling and motivations for purchase". In: *Royal Society Open Science* 6.6. ISSN: 20545703. DOI: 10.1098/rsos.190049.

Zendle, David, Lukasz Walasek, et al. (Apr. 2021). "Links between problem gambling and spending on booster packs in collectible card games: A conceptual replication of research on loot boxes". In: *PLoS ONE* 16.4 April. ISSN: 19326203. DOI: 10.1371/journal.pone.0247855.

Zikmund-Fisher, Brian J. et al. (June 2006). "A matter of perspective". In: *Journal of General Internal Medicine* 21.6, pp. 618–622. ISSN: 0884-8734. DOI: 10.1111/j.1525-1497.2006.00410.x. URL: http://link.springer.com/10.1111/j.1525-1497.2006.00410.x.

Unless otherwise expressly stated, all original material of whatever nature created by Niccolò Toccafondi and included in this thesis, is licensed under a Creative Commons Attribution Noncommercial Share Alike 3.0 Italy License.

Check on Creative Commons site:

https://creativecommons.org/licenses/by-nc-sa/3.0/it/legalcode/

https://creativecommons.org/licenses/by-nc-sa/3.0/it/deed.en

Ask the author about other uses.