Università
Ca' Foscari
Venezia

Dottorato di ricerca in Economia
Scuola Dottorale di Ateneo
Graduate School
Ciclo XXVII
(A.A. 2016-2017)

Essays on tax compliance, economic behavior and linguistic relativity

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/03

Tesi di dottorato di Juliana Bernhofer, Matricola 804668

Coordinatore del Dottorato
Prof. Giacomo Pasini

Tutore del dottorando
Prof. Michele Bernasconi

Essays on tax compliance, economic behavior and linguistic relativity

by

Juliana Bernhofer

Matricola 804668

A thesis submitted in partial fulfillment

for the degree of

Doctor of Philosophy in Economics

Department of Economics

Ca' Foscari University of Venice

Supervisor :

Prof. Michele Bernasconi

December, 2016

The undersigned Juliana Bernhofer, in her quality of doctoral candidate for a Ph.D. degree in Economics granted by the Università Ca' Foscari Venezia attests that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

Abstract

This dissertation presents three essays on tax compliance, economic behavior and linguistic relativity. In Chapter 1, I explore the existing literature of behavioral and experimental economics, as well as multi-agent simulation models applied to tax compliance. The analysis begins with the presentation of the classical models of expected utility theory and the discussion of the limitations arising with the assumption of full rationality. During the last decades, the inclusion of institutions, social norms and psychological factors in laboratory experiments have revealed numerous interesting insights to policymakers. However, given the complexity and the size of the research, I propose the use of agent-based simulation models as flexible tools when accounting for existing findings and to explore new research horizons. Some examples of computer-simulated agent-based models applied to the tax evasion problem are presented.

Chapter 2 tests the *Reduction of Compound Lotteries Axiom* in a tax evasion experiment. When the probability of audit and the probability of detection are considered separately, we find violations of the reduction axiom and thus of the assumption that tax payers are expected utility maximizers. Our results call for a reconsideration of existing theoretical models on individual tax compliance.

Finally, Chapter 3 is an interdisciplinary study of the role of language to economic behavior. In particular, we test if a more intense use of the *Irrealis* mood, i.e., the verb form that expresses potential situations, in an individual's spoken language impacts on their attitude towards risk by giving linguistic salience to the concept of uncertainty. We find that the use of *Irrealis* mood is positively and significantly related to the level of risk aversion. We use our findings to assess the role of (instrumented) risk aversion in investment decisions.

Acknowledgements

The last five years have been a journey of emotions and discovery, a rocky and exciting trip off the beaten paths. All this would not have been possible without the encouragement and support of my colleagues, friends and family.

I am deeply thankful to my supervisor, Michele Bernasconi. His competence as a mentor is unquestionable, but what truly made the difference was his ability to understand my difficulties without me even defining them. Numerous times he brought me politely back to earth when I was getting lost in too many ideas, while motivating and guiding me when he realized that I was on the right track. I also want to express my gratitude to Paolo Pellizzari and his useful comments both, in the first and in the final phase of my dissertation.

During my PhD I had the privilege to meet wonderful people in Venice, Barcelona and Padova. First of all I want to thank Anna from the Tratoria da Marisa for the extra-abundant portions of pasta al ragu' she handed me over and for always trusting in my eating abilities. I want to thank Lisa Negrello for her patience, moral support and for all the organizational work she has done for us during these years. Also, I am grateful to Massimo Warglien and Rosemarie Nagel for being an inspiration, both personally and professionally. Thank you Anna, Giulia, Stefano, Patrick and Mayank for your loyalty and friendship in good times and in bad times. I also want to thank Cristina for the good talks and cigarettes and Ludovico for our heated political debates and his fine sense of humor. Last but not least, thank you Matija, your support as a coauthor has been invaluable, but I want to thank you particularly for your friendship, loyalty and honesty even during hard times.

Anyone who has written a doctoral dissertation knows about the lonely hours and the uncertainties. I want to thank my old friends Claudia, Fabio, Tania, Silvio, Larry, Chiara, Henry and Teresa for being still here with me. A special thank you also goes to the Cendron family, for their generosity and for making me feel at home.

I want to thank my father Romeo for teaching me to be free even if it means to be different. I miss you deeply. Without doubt, I can say that I took my first steps into research with my mother Ursula when I started to ask the typical "why" questions. Whenever she was unsure about the answer we would go to the shelf and look it up in the encyclopedia. Nadine, you are not only my sister but my best friend. Thank you for having my back no matter what and for being the kindhearted, smart and funny person you are. Finalizing this thesis would not have been possible without my husband Johannes whom I thank for pushing me to choose research over comfort. Thank you, for your patience and for being my partner and a wonderful father to our daughter.

I conclude my acknowledgments by thanking Frida from the bottom of my heart...you show me every day what unconditional love feels like. I am looking forward to loving you for the rest of my life.

To Frida and Romeo.

Contents

List of Tables

List of Figures

Chapter 1

Experimental Economies and Tax Evasion: The Order Beyond the Market

[1]

"In this world nothing can be said to be certain, except death and taxes."

Benjamin Franklin

[1] This chapter has published as Bernhofer (2016)

Abstract: *Research on tax evasion will probably never get old. As long as there are taxes, there will also be policy-makers all over the world eager to tackle deviant conduct in the most efficient and efficacious way. To fill this purpose a number of theoretical and empirical frameworks have been developed in economics over the last decades, starting from the classical models of Allingham and Sandmo (1972) where individuals were assumed to be perfectly rational following a pure cost-benefit logic. Today, however, we look at a body of literature which has opened up to a number of new and interdisciplinary findings, also thanks to the inclusion of behavioral aspects that do not necessarily follow the paradigms of the homo economicus. To this end, the discipline of Experimental Economics has developed numerous ways to overcome the distance between economic theory and human behavior. The aim of this survey is to take the reader on a tour through some of these methodologies applied to the analysis of tax evasion, arguing that further research should focus on integrating multi-agent simulation models with outcomes from human subject experiments in order to create useful and necessary tools to administer, consolidate and represent the complex theoretical, empirical and experimental panorama of tax evasion research.*

Keywords: Tax Evasion – Agent-based Models – Experimental Economics – Artificial Markets – Behavioral Economics – Rational Choice Theory

1.1 Introduction

Tax non-compliance and fiscal fraud are giving policymakers all over the world quite a headache, leading them to invest significant efforts and resources in the attempt to tackle the issue of public revenues lost due to tax evasion. To achieve meaningful results through targeted policy interventions, it is central to understand how people perceive taxes, contributions, and sanctions, along with how they take the decision to comply. Various attempts have been made, starting from the seminal contributions of Allingham and Sandmo (1972) and Yitzhaki (1974), who follow the paradigms of expected utility theory. These two classical theoretical models are presented in Section 1.2, as they serve as a starting point for the majority of further developments in the tax-compliance literature. Despite their elegance and insightfulness, the classical models were soon questioned as it was shown that their predictions did not match existing empirical evidence (see, for example, Graetz and Wilde (1985)). In fact, people were found to evade much less than what would have been expected from a rational utility maximizer with a reasonable level of risk aversion. Sections 1.3 and 1.4 deal with issues related to the assumption of full rationality and the potential problems arising with the use of expected utility theory.

The process of cross-checking real data with model prescriptions and the acknowledgment of the limitations of analytic solutions led to modifying the basic research question from, "Why and by how much do people evade taxes?" to "Why do people pay taxes at all?" The answer should ultimately enable policy makers to find a solution to their old dilemma of, "How can compliance be increased even more?" However, determining the extent of tax evasion is and has always been an understandably challenging task. There are four main methods, according to Andreoni et al. (1998): audit data, survey data, tax amnesty data, and laboratory experiments. A frequently cited source associated with the first category is the Tax Compliance Measurement Program (TCMP) carried out by the U.S. Internal Revenue Service (IRS) from 1965 to 1988. The TCMP was a program of intensive audits designed to measure the level of noncompliance among the population. The results of the 1988 TCMP show that 40 % of the households evaded some income tax, whereas 53 % were fully compliant. The remaining 7 % instead paid more than what they actually owed. Yet available data on tax compliance is limited and often unreliable. Audit data does not give sufficient insights, as it is nearly impossible to detect all hidden income. Survey data, on the other hand, is self-reported, which casts reasonable doubt on the truthfulness of the information provided about one's own illicit behavior. Elffers et al. (1987) conducted an investigation on a Dutch sample and find no significant correlation between survey-elicited self-reports and actual behavior in their data, with 69% of documented evaders denying their evasion in a subsequent questionnaire. Tourangeau and Yan (2007) analyze the role of the social desirability bias in standardized personality measures and find that sensitive survey topics, such as illicit drug use, smoking, abortion and sexual behavior are subject to frequent understatements and a higher level of non-response. The interpretation of these results is that respondents seem to be dishonest in answering sensitive questions in order to avoid embarrassment or repercussions from third parties. However, survey design and as a consequence the level of perceived confidentiality is found to play a fundamental role, with self-administered and randomized procedures decreasing the level of underreporting for socially undesirable types of behavior.

To overcome these and other limitations, experimental economics comes into play in the attempt to provide at least a partial fix. By carefully constructing a laboratory environment as described by an economic model (e.g., tax rate, audit frequency, fine rate), the experimenter may observe whether participants behave according to the analytical predictions. Furthermore, marginal effects of single-parameter changes can be isolated *ceteris paribus*, thanks to the controlled setting of the lab. Similarly to tax amnesty data, where only a subset of participants to the amnesty program is analyzed, also experiments may present potential problems of external validity. In Section 1.5, we will present some results obtained from laboratory experiments on tax compliance and discuss the limitations of this method more in depth. The body of literature on tax evasion has grown to massive proportions during the last three decades and has become cumbersome to overlook. However, two of the main common findings are that individuals are not homogeneous in preferences and often do not act according to market-based mechanisms. Yet the plurality of stylized determinants of tax compliance and the discovery of the importance of interaction effects and dynamic approaches have given rise to new ways of modeling. Some examples of computer-simulated agent-based models applied to the tax evasion problem are illustrated in Section 1.6. The potential benefit of calibrating such models with the results stemming from real "agents" tested in controlled laboratory settings still has to be further explored. This new area of interaction between both Experimental Economics and computer-simulated realities - or an Experimental Economy - could well be one innovative tool to bridge the gap between theoretical modeling and reality on one side, while decreasing the distance between academics and policy makers on the other, thanks to the creation of manipulable interfaces that are also intuitive for non-technical users.

1.2 Classic Modeling of Tax Evasion

A classic approach to the problem of tax evasion is offered by the theoretical model contained in the seminal article of Allingham and Sandmo (1972) (henceforth AS). The model is an adaptation of Gary Becker's work on the Economics of Crime and Punishment (Becker 1974) to the case of tax compliance. The decision on how much to declare of a certain exogenously given income is presented as a gamble with two possible outcomes: being caught and not being caught. The tax authority audits the taxpayer and discovers the understatement with a certain probability p; hence the decision-maker maximizes his expected utility (EU) with respect to the declared amount x according to the following convex combination:

$$EU = (1-p)U(v + t(y-x)) + pU(v - \theta(y-x)) \quad (1.1)$$

His overall expected utility is represented as a weighted average of the utilities assigned to the two possible outcomes. The optimum amount of declared income x depends on the proportional tax rate t , the fine rate θ, and the probability of being subject to a random audit p. The utility function $U(.)$ is marginally positive with $U'(.) > 0$ and strictly decreasing with the second order derivative $U''(.) < 0$, which means that the taxpayer is a risk-averse one who prefers a certain outcome to a gamble with the same (or even a higher) payoff in expected terms.

The correct net disposable income in case of full income disclosure is described by $v = y(1-t)$. Thus one can interpret the first part of the expected utility formulation as the situation in which no audit is performed and the utility is given by the argument $(v + t(y-x))$, the correct net income v augmented by what we will call the cheater's premium, $t(y-x)$, namely, the part of tax liability he saved by not paying taxes on the undeclared part. The second part of the weighting function describes the situation in which an audit takes place: the undeclared income $(y-x)$ is detected by the authority and the fully taxed correct net income v is reduced by the cheater's penalty $\theta(y-x)$, i.e., the fine on the hidden income.

Commonly, decreasing absolute risk aversion is assumed, which describes the fact that - in absolute terms - the amount of risky investments increases with higher disposable income.

The first-order condition (FOC) for the optimal amount of income disclosure[2] of Eq. 1.1 is $\frac{\delta EU}{\delta x} = 0$ and becomes $\frac{U'(y_A))}{U'(y_N A)} = \frac{(1-p)t}{p\theta}$. We find that in the AS specification an increase in the tax rate t has ambiguous effects. On one hand, the correct net disposable

[2] The second-order condition is satisfied by the utility function's being concave.

income v decreases, which - under the assumption of decreasing absolute risk aversion[3]- should induce the individual to cheat less, leading to an income effect. On the other hand, with a higher tax rate, the attractiveness of the cheater's premium $t(y-x)$ increases, whereas the cheater's penalty $\theta(y-x)$ remains unaffected, which eventually makes tax evasion more profitable (substitution effect). The magnitude and sign of the final response depend on the shape of the utility function, in particular on how fast absolute risk aversion declines and thus its third-order derivative (Andreoni et al., 1998). With the aim of reflecting the legal framework effective in a number of countries, such as the United States and Israel, in 1974, Shlomo Yitzhaki (1974) introduced a slight modification to the specification of the AS model by making the penalty depend on the evaded taxes instead of the undeclared amount of income. The original expected utility expression becomes

$$EU = (1-p)U(v + t(y-x)) + pU(v - \theta t(y-x)) \tag{1.2}$$

and we find the following first-order conditions for optimality: $\frac{U'(y_A))}{U'(y_N A)} = \frac{(1-p)}{p\theta}$. The taxpayer will evade as long as the expected payoff per monetary unit of evasion $1-p-p\theta$ is greater than zero.

By directly comparing the F.O.C. of Eq. 1.2 with the original AS first-order condition, we observe that the proportional multiplicative effect of taxes in the numerator, which made evasion more attractive, disappears and what remains is only the income effect (Slemrod and Yitzhaki, 2002). Overall, cheating will be reduced when the tax rate increases. Moreover, both the probability of an audit and the magnitude of the fine have a negative impact on evasion.

The Allingham-Sandmo-Yitzhaki (henceforth ASY) model stands out for its elegance and straightforwardness, and thus became the standard tool, or at least the starting point,

[3]The concepts of absolute and relative risk aversion have been developed by Arrow (1965) and Pratt (1964) independently. The first measure describes the amount of wealth placed in risky activities in absolute terms and the second expresses this amount in relative percentage terms.

in the analysis of the compliance decision. Yet an ever-growing body of empirical and experimental evidence has developed pointing out that the predictions of the ASY model do not fully hold up in reality. In the following sections we will highlight some of the criticisms to this classic utilitarian approach.

1.3 Limits of Rational-Choice Theory

Presenting the decision of tax compliance as an individual-choice problem in order to resolve what could be well defined as an aggregate social problem has proven to lack significant elements that influence the process of human decision-making. The rational choice approach has been widely challenged, as described also by the American economist James M. Buchanan:

> The economist rarely examines the presuppositions of the models with which he works. The economist simply commences with individuals as evaluating, choosing, and acting units. Regardless of the possible complexity of the processes or institutional structures from which outcomes emerge, the economist focuses on individual choices. [...] Individuals [...] are presumed able to choose in accordance with their own preferences, whatever these may be, and the economist does not feel himself obliged to inquire deeply into the content of these preferences (the arguments in individuals' utility functions).
>
> Buchanan (1987), p. 244

Standard neoclassical models of economic theory are built on the assumption of individuals' exhibiting rational behavior. Rationality in the economic context is interpreted as the individual's capability of evaluating all possible outcomes in order to take the decision that yields the greatest benefit in terms of utility. Moreover, agents are also assumed to be aware of their own preferences and to be able to maximize their utility function given certain parameters. They do so in a purely self-interested way. Accordingly, in the classic AS and ASY model of tax compliance the decision of how much to declare from one's income to the tax authority is presented as a relatively simple portfolio choice. Taxpayers must decide how much of their income they wish to allocate to the risky asset (tax evasion) and to the safe asset (tax compliance). Decision-makers are assumed to have full information about the audit probability, the fine, and the tax rate they are supposed to pay and make so-defined "rational choices under uncertainty." The latter is engineered by assigning probabilities to possible outcomes. Portfolio theory and its basic assumptions of perfect rationality have been widely challenged by the science of behavioral economics, which deals with the social and cognitive aspects in the human decision-making process. By introducing elements of psychology into economic modeling, new points of view have been presented which are not necessarily in contrast with the neoclassical models. Extensive experimental research has shown that individuals are only boundedly rational (see, for example, Conlisk (1996) for a survey on bounded rationality). The concept was already introduced in the 1950s by the work of Herbert A. Simon and has since been subsequently defined and modeled by numerous authors (Simon, 1982; Selten, 2001; Simon, 1972; Kahneman, 2003)). Examples of what could be those bounds to full rationality are information asymmetry and cognitive limitations. How does bounded rationality apply to the tax evasion problem? First, most of the time citizens do not have complete information about the true audit probabilities and form subjective probability beliefs about the frequency of verifications by the tax authority. Also, tax code complexity and bureaucracy can lead to uncertainty not only about the fine parameter θ, but also about the correct tax rate itself, as pointed out by (Andreoni et al., 1998, p. 852). Tax complexity leads to the need for tax practitioners and represents a potential source of inequity among the population, in particular with respect to education and socioeconomic status. These regressive effects stemming from inferior capabilities of interpreting the tax code and, as a consequence, finding ways to minimize the tax liability are also mentioned by Vogel (1974). Second, optimizing a utility function based on some probabilities of possible outcomes might not be a straightforward process for everyone. It is demonstrated by numerous studies, as described in Reyna and Brainerd (2008), that there is a general lack of mathematical proficiency and subsequent difficulty in judging probabilities and risks among the population. Not only is it numeracy, but also tax literacy that plays a role in determining the correct level of the expected-utility parameters to be taken into consideration. Tax literacy is tightly linked to the aforementioned tax code complexity. The first is the capability of applying the correct tax rate given a certain legal framework, whereas the latter describes the structure and accessibility of that legal framework. The more complicated the design of the set of rules, the more important the capability to interpret them correctly, in order words, *ceteris paribus*, the marginal return to tax literacy increases. Finally, it is also assumed that taxpayers take decisions individually and in a self-interested manner, but can we assume this mechanism to be compatible with the very purpose of taxes? Taxes are collected in order to finance public expenditures that again are designed to serve the community of taxpayers. This implies that tax evasion is a form of free-riding. Not paying one's taxes has the effect that the other members of the community have to pay more in order to fund the public collective project. In that sense it might be necessary to consider social interactions and norms in order to capture the role of peer effects, positive and negative reciprocity, and intrinsic motivation in the decision-making process. We will consider these elements more in depth in Sect. 1.5.

1.4 The Expected-Utility Approach Under Scrutiny

In the ASY model the decision to pay taxes is presented as a lottery with two possible outcomes: audited and not audited. Taxpayers are then expected to decide how much to declare based on the probability of being audited and possibly fined. Following this logic and given the population-specific level of risk aversion, it should hence be possible for the lawmaker to provide society with a set of rules defining audit rates, tax rates, and fines which leads to collecting the maximum tax levy in a self-regulatory manner. Individual heterogeneity is represented with regard to the attitude toward risk and captured by the functional form of the utility function, in particular its curvature. A risk-averse individual is characterized by a concave utility function with a decreasing return to wealth in marginal terms. The Arrow-Pratt measure of risk aversion in absolute terms describes the relationship between the second-order and the first-order derivative of the utility function, whereas the measure in relative terms describes the level of risk aversion with varying levels of wealth. The level of risk aversion in the context of tax compliance has been studied by numerous authors. Alm et al. (1992b) showed that estimated Arrow-Pratt levels of relative risk aversion for the United States are incompatible with the empirical evidence of tax compliance. The real levels are between 1 and 2, but only a level of 30 would support observed tax compliance rates. Frey and Feld (2002) find that the observed compliance in Switzerland of 76.52 % would require a value for the parameter of relative risk aversion of 30.75, as opposed to the observed parameter values ranging from 1 to 2. The empirical calibration values for the model were presented in Alm et al. (1992b), Andreoni et al. (1998), and Bernasconi (1998), with real-world average audit rates ranging between 1 and 3 %, and the penalty rate, which we called θ in our specification, ranging from 0.5 to 2.0. The return to tax evasion in expected terms can be obtained using $1 - p - p\theta$ and results in 91-98.5 %; hence all taxpayers should hide some of their income, which stands in contrast with the evidence showing that only 30 % of taxpayers actually evade taxes Dhami and Al-Nowaihi (2007). Again, only unreasonably high levels of risk aversion could explain the levels of tax compliance found in reality.

Criticisms to the AS and the ASY models of tax evasion link back to the very same discussions around EUT itself. Decision-makers are defined to be (rational) expected utility maximizers if they meet the four basic criteria of the Von Neumann- Morgenstern specification: (i) Completeness - preferences of individuals are well-defined; hence they are able to choose between two alternatives. (ii) Transitivity - the choices are coherent; i.e., if outcome A is preferred to outcome B and B is preferred to C, then it must be that outcome A is preferred to outcome C. (iii) Independence - if gamble A is preferred to gamble B and another gamble C is added to both of them, then preferences do not change: the new gamble (A + C) must still be preferred to the new gamble (B + C). (iv) Continuity - given the preference ranking A $\succ$ B $\succ$ C then there must exist some value of p in a convex combination of A and C which makes the decision-maker indifferent to option B, such that $pA + (1 - p)C \sim B$.

A challenge to the Von Neumann-Morgenstern utility specification is offered by Kahnemann and Tversky and their work on Prospect Theory (Kahneman and Tversky (1979)) and Cumulative Prospect Theory (Tversky and Kahneman (1992)), which shows inconsistency in preferences describing the nonlinear subjective reaction to probabilities. Starting from a reference income and moving into the gain domain, preferences are concave, whereas in the loss domain preferences are convex. The aim of determining a reference point from which to depart in defining the gain and the loss domain is to eliminate possible framing effects. Going back to our taxation framework, in Dhami and Al-Nowaihi (2007) this reference point is defined as the legal after-tax income.

Moreover, concavity of gains and convexity of losses indicates the presence of a loss aversion, where losses are perceived as worse than gains in relative terms. Cumulative Prospect Theory (CPT) uses rank-dependent expected utility theory in order to define the probability weighting function. In that way decision-makers will tend to overweight low probabilities and underweight high probabilities. In taxation terms, such a mechanism implies that a realistic audit rate of, say, 0.01 is subjectively interpreted as higher. Hence compliance for low audit rates increases with respect to the predictions of standard expected utility theory and is more in line with real world data. This hypothesis was tested in the laboratory by Alm et al. (1992b). In their experiment they set a cut-off level for the audit probability of 5 % below which a risk neutral expected utility maximizer should report zero income. Yet at a level of 2 % they still find significant compliance rates of around 50 %. Such a result could fit expected utility theory only by assuming extreme values of risk aversion. Still, the results are consistent with the predictions of Cumulative Prospect Theory which allows for subjects to subjectively perceive a higher audit rate than the given one.

They also find that the reactions to increases in audit probabilities are non-linear, with compliance rising less than the audit rate in relative terms. An additional variation of their experiment consisted in a treatment with no possibility of being detected. Nonetheless, the average compliance rate was 20 % which makes the comparison with the neutral lottery set-up somewhat questionable and files Cumulative Prospect Theory as only a piece of the compliance puzzle.

1.5 Institutions, Social Norms and Psychological Factors - New Evidence from the Lab

Indeed, it has become quite clear that there are several further "ingredients" to be considered in order to get a better picture of the various aspects involved in the tax compliance process, aside from the fine rate, the available income, the audit frequency and the tax rate. The *homo economicus*, the economic man, who is assumed to act rationally and in a self-interested manner, would be better off not paying taxes at all if we consider real-world audit rates, even under the assumption of extreme levels of risk aversion. In this section, we will present some results of tax compliance experiments testing for the impact of both the classical parameters and behaviorally driven elements which have been gradually introduced in the attempt to accommodate observed tax compliance data. A typical tax compliance experiment is computer-based and consists of one or more rounds during which subjects are asked to take a decision on how much of their previously assigned income they want to declare to the tax authority, given the audit frequency, the fine in cases of detection and the tax rate. Thereafter their report may be randomly drawn for an audit and if they declared less than their gross income, the penalty is applied. It is also to be mentioned, however, that experiments in economics are often subject to criticism with regards to their external validity. Guala and Mittone (2005) dedicate a section to the issues related to the tax compliance environment in the laboratory, naming as examples problems of scale, the game-like behavior of subjects, the absence of social incentives (the "real" social environment is not part of the experiment) and the absence of social actors. The authors admit to the difficulty of generalizing laboratory results to the real world due to their inherent context-specificity. Nevertheless, tax experiments may offer valuable cause-effect explanations and might often even be the only chance to get additional data on the behavioral dynamics behind the tax compliance decision (Alm et al., 1992a), given the difficulty of gathering truthful and reliable data on tax compliance. Another criticism common to laboratory experiments in general is the use of student subjects. This is addressed by Alm et al. (2010) and tested in a tax compliance experiment conducted with both, students and staff. Some variation was introduced between groups with regards to the level of certainty about the tax liability and the existence of social programs aimed at positively inducing taxpayers (namely Income Tax Credit and Unemployment Benefits). They find that average compliance rates within subjects (staff and students) indeed differ, but that changes within group treatments are alike. Similarly to Guala and Mittone (2005), their findings suggest that laboratory experiments are able to offer insights regarding marginal effects of parameter changes.

1.5.1 Testing the Classic Microeconomic Predictions

Experiments that test traditional microeconomic models such as those we have seen (AS and ASY), focus on manipulations of the enforcement regime, the tax rate and the fine rate. In an experiment conducted on law students, Friedland (1982) found for example that responsiveness to information about threat probability (audit) is higher than to information about threat magnitudes (fines). Moreover, a number of experiments have been conducted to assess reactions to variations in the tax burdens. The ASY model predicts the compliance rate to be increasing in the tax rate, however, this cannot be confirmed by a number of findings coming from the lab, which is commonly called the Yitzhaki puzzle. Alm et al. (1992a) find that tax evasion increases with the tax burden which is in-line with the empirical findings of Clotfelter (1983), even with the tax rate elasticity being similar[4]. Also the experimental subjects of Bernasconi et al. (2014) tend to increase their compliance when the tax burden decreases and vice versa. In addition they find that the reaction to tax cuts and tax rises is asymmetric with faster reaction to the first than to the latter. Cultural factors and social norms might also play a role as shown by Alm et al. (1995), whose results will be presented more in detail in the next section. They find that Spanish test subjects behave according to the ASY model, increasing their compliance with higher tax burdens at a positive rate of 0.94 whereas the U.S. subject pool confirms once again previous findings as in Clotfelter (1983) and Alm et al. (1992a) with a negative elasticity of around -0.5. An attempt to modify the ASY model in order to match the experimental and empirical evidence has been made by Dhami and Al-Nowaihi (2007) who show that under prospect theory, hence depending on the reference point, tax evasion is increasing in the tax rate[5]. Overall, similarly to empirical results, also in the laboratory higher-than-rational levels of compliance are usually observed, which prompted researchers to investigate the role of determinants other than the classical triad of parameters.

Tax Compliance as a Social Norm

The juxtaposition of the profit-maximizing *homo economicus* in the Smithian sense with Dahrendorf's more other-regarding *homo sociologicus* became a necessary adjustment in

[4] around -0.5

[5] Strictly increasing for interior solutions and non-decreasing in case of boundary solutions when $D^* = 0$ or $D^* = W$.

the attempt to disentangle the fundamental drivers of the tax compliance decision. The latter characterization describes an economic agent who acts according to social norms and exhibits feelings such as guilt and anxiety. Roughly said, social norms are behavioral rules shared by other people who tend to judge them in a similar way. Elster (1989) cites some examples of social norms, such as norms of reciprocity, work norms and norms of cooperation. Special forms of norms of cooperation are given by norms of fairness. The relevance to the tax compliance decision covers various aspects. First, perceived peer-to-peer fairness which follows a logic similar to: "if others (don't) pay their taxes then I am (not) going to pay them as well". Second, the tax system itself might be evaluated in terms of fairness before taking a compliance decision. The items under scrutiny could be the magnitude of the fine, the frequency and modality of performed audits, the tax progressivity and thus the level of tax equity with respect to one's income. Finally, taxes are levied in order to finance public projects which may lead to an evaluation of the personal gain from paying taxes and receiving public good consumption in return, or also of the efficiency of public spending. One way to test the relevance of social norms is to conduct cross-country surveys. By comparing responses from different cultures with similar fiscal systems, different tax attitudes and compliance rates emerge. Alm et al. (1995) provide a rough summary of the main findings of such studies. Drivers of tax compliance can be classified into moral (compliers view tax evasion as immoral and "moral appeal" tends to have positive effects on compliance), reputational (low social standing of tax evaders), peer effects (friends of tax evaders tend to evade more), perception of fairness, trust, and social cohesion. Alm et al. (1995) conducted a tax compliance experiment in the laboratory with Spanish subjects replicating an earlier study that was run in the United States. From the comparison of the two studies, it emerges that the Spanish subjects tend to comply less than their American counterparts in absolute terms. In the absence of a public good, with a fine rate of 2, a 30 % tax rate, and a 5 % probability of being audited, compliance of the American subjects is 27 % on average, whereas in Spain it amounts to only 7 %. However, the Spaniards turn out to be much more sensitive toward fiscal policies, such as changes in the tax rate, the audit rate, or the magnitude of the fine. While reminding the reader that the only difference between the two experiments lies in the cultural origin of the subject pools, the authors conclude that the social norm of compliance, which can also be defined as "tax morality", might be the reason for the difference in responses. It also emerges that there are different types of taxpayers: those who always comply and those who never comply, utility maximizers, subjects that behave according to prospect theory overweighting low probabilities, highly policy-sensitive subjects, and some who are at times cooperative and at times free-riders.

The power of social norms is also determined by the interaction and enforcement among individuals. The relevance of peer effects is bolstered by Vogel's analysis of a survey with Swedish respondents (Vogel, 1974), which shows that contacts with tax evaders decreases tax compliance, weakening the social norm of compliance and, as a consequence, also the stigma of evading. On the other hand, conforming to tax-paying peers might also yield a positive return, based on the individual's level of intrinsic tax morale. A formal model, which is still embedded in the EUT framework, was developed by Myles and Naylor (1996). Conformity to social groups yields an additional payoff to the taxpayer, which depends on the size of the group itself. Moreover, the non-evasion equilibrium could potentially be turned over by small changes in the tax rate, leading to an evasion epidemic with tipping point behavior. Tightly linked to the concept of social norms and group conformity are the models considering the psychic costs of evasion, as described by Gordon (1989). The evaded amount becomes a function of an additional parameter which is determined by the personal level of morale and peer effects. The positive relationship between tax morale and tax compliance has been tested and confirmed numerous times in the laboratory (for an extensive survey and discussion of experimental results see Torgler (2002). In a separate article, Torgler (2003) shows empirically that the level of tax morale itself is influenced by formal and informal institutions, such as direct participation rights and trust in the government. He defines tax morale as "the intrinsic motivation to pay taxes" or "the willingness to pay taxes by the individuals." We have seen that the tax-compliance decision is not only determined by the absolute levels of the classic parameters, which are the fine rate, the tax rate, and the probability of being caught in a random audit. Psychological and cultural factors also play a role, as well as peer interactions. The latter exhibit imitative patterns based on lagged events and give rise to the need for a dynamic modeling approach. In the following section we will present the tool of agent-based modeling, which represents a way to bridle the rise in complexity of stylized facts that potentially influence the tax-compliance decision.

1.6 From Top Down to Bottom Up - From Experimental Economics to an Experimental Economy

The ultimate purpose of research on tax evasion is undoubtedly to find policies able to increase tax compliance and hence the overall tax levy. A number of interplays and complexities characterizing the system (e.g., country) under analysis have to be considered in order to fit the outcome-predicting model as closely as possible to the underlying reality without too much loss of generality. It is, however, a challenging task and not always possible to disentangle the behavioral and economic elements affecting the tax-compliance decision by analyzing the available empirical data. Parameter values needed for a correct calibration could be obtained, for example, by performing field studies and conducting laboratory experiments. The idea of using experimental data (but not solely) in order to feed a computer-simulated replica of society was suggested by Duffy (2006), and even though calibration with experimental data has not yet become a widespread habit, we will see one agent-based model that put these suggestions into practice. Agent-based models offer an alternative approach to deal with complexity, and the available tools make it possible to take interactions and heterogeneities into account without necessarily abandoning the simplicity of representation. It is possible to artificially recreate an experimental economy that reacts according to the model we choose. The parameter values can be calibrated according to experimental or empirical findings; and, most importantly, this technique allows simulation with heterogeneous agents and social interactions in a dynamic environment. The results of agent-based simulations are then compared with real data, allowing for a more detailed understanding of the underlying social and behavioral dynamics.

1.6.1 Group Conformity and Social Norms

Bloomquist (2006) reviewed three agent-based simulation models applied to an environment of tax compliance, namely, Mittone and Patelli (2000), Davis et al. (2003), and Bloomquist (2004). Mittone and Patelli (2000) use the model of Myles and Naylor (1996) as a basis, which considers group conformity and the social norm of tax compliance. Psychological costs are also included in the model, as originally proposed by Gordon (1989), but without making them depend on the evaded amount. The underlying idea is that, no matter how much income is hidden, once the decision to evade is put into practice, the "honest citizen" status is lost. Agent heterogeneity is captured by introducing three types of subjects: the honest taxpayer, the imitative taxpayer, and the perfect free-rider, each with his or her own specific utility function. Honest agents achieve positive marginal utility effects from conforming to social rules; free-riders will contribute as little as possible; and imitative taxpayers will use the population mean of compliance as a benchmark, which is also in a way in line with the findings on peer effects described by Vogel (1974) and the findings of Porcano (1988), that the perception of existing evasion has a positive and significant effect on the own level of evasion. Additional utility gains are obtained from the introduction of a public good that depends on the amount of tax levy, as considered by the theoretical model of Cowell and Gordon (1988). Finally, the behavioral characteristics of single agents are not static, but subject to a stochastic updating process, a genetic procedure where probabilities of typesurvival are calculated based on individual utility gains over total population utility gains. Decision rules on how much to declare follow a learning mechanism with choices being updated based on the success or failure of past compliance decisions. The random element in the behavioral switching algorithm, in combination with the feature of imitative behavior, triggers a cycle that allows for the model to react to audit rates that are close to zero, pushing compliance to a near-zero level, even if all agents were initially honest. On the other hand, when audits are introduced, only honest taxpayers remain after a certain number of rounds. It is interesting to notice that the type of audit procedure, uniform versus those aimed at the lower tail of contributions, does not change this result. Another model considering peer-oriented behavior is the Multi-Agent Based Simulation (MABS) developed by Davis et al. (2003) using the software Mathematica. The behavioral classification of taxpayers is similar to the previous model: there are honest, susceptible (to others' behavior), and evading agents. Initially, there are two randomly assigned types: honest and evading. An honest taxpayer might, depending on the "infection rate," become a susceptible one in case a (randomly chosen) acquaintance happens to be an evader. Susceptible agents form beliefs about the severity of the tax enforcement regime by observing the mean audit frequency of their peers in the previous period. They become evaders if this perceived severity is below a certain threshold. Finally, evaders become honest taxpayers if the belief about the severity of the audit regime is above a certain threshold or if they observe a social norm of tax compliance. The existence of a social norm of tax compliance is confirmed by having a certain number of honest taxpayers among the own acquaintances. Also, evading agents become honest after a tax audit. The authors follow the literature on availability (Tversky and Kahneman 1973) and vividness, assuming that the subjectively determined probability of being audited is judged to be higher after a recent audit experience. This assumption, however, stands in contrast with the experimental findings of Mittone (2006) and Maciejovsky et al. (2007), who observed negative post-audit responses, which are likely to be due to the socalled Bomb-Crater Effect, according to which most recent events are judged to be unlikely to occur again immediately. A second cause is a loss-repair mechanism as described by Maciejovsky et al. (2007) and suggested also by Andreoni et al. (1998) where the fined taxpayer tries to recover the sum by evading more in the subsequent round. By manipulating the starting proportion of evading and honest agents and letting the systems evolve over 2,000 rounds, Davis et al. (2003) find that changes in the audit rates from 0.002 to 0.030 in steps of 0.002 lead to "tipping point" behavior with abrupt changes in compliance equilibria. In all set-ups societies converge to total honesty at an audit rate of only 0.03. Although the latter result is not supported by empirical findings, it is still notable that the audit rate may be used as a device to prevent a non-compliance epidemic from happening. The authors suggest establishing a similar experimental environment with social norms and group conformity in order to confirm the robustness of their finding.

1.6.2 The Tax Compliance Simulator (TCS) - Playing with Complexity

Bloomquist (2004) uses the software NetLogo in order to simulate a more complex agent-based environment. His Tax Compliance Simulator (TCS) is capable of testing the effects of variations in audit rates, fine rates, income visibility (wages and salaries versus other sources of income), auditor efficacy, and audit celerity after the event of an evasion. What is measured are direct effects of audits given by the additional levy, indirect effects of audits experienced by peers, and post-audit responses. The interface of the TCS is also quite intuitive for a non-technical user and was illustrated in (Bloomquist, 2006, p. 423). It is composed by various sliders (such as tax rate, audit rate, penalty, etc.) through which it is possible to manipulate the desired parameter values before starting the simulation. Diagrams at the top show the evolution of the variables of interest, as, for example, the amount of reported tax over time, whereas the distribution dynamic of full evaders, partial compliers, and full compliers appears in the window containing "turtles," the NetLogo labeling of what we called agents. The TCS features both overweighting of low audit probabilities and overestimation. Taxpayers tend to overweight low probabilities of audit, as predicted by prospect theory (Kahneman and Tversky (1979)) and experimentally assessed by Alm et al. (1992a). All probabilities were given, but there was still evidence of people typically overweighting these probabilities. Overestimation, on the other hand, is a bias which is a less numerical and more psychological phenomenon: it may depend on the subjects' perception of the auditing mechanism and also past audit experiences, as hypothesized in Kastlunger et al. (2009). In order to help account for the opportunities to evade, it is possible to determine the percentage of visible income. Visible income being subject to third-party information reporting, such as salaries, is assumed to be entirely declared, which is in line with empirical data. However, once the agent does not declare her full income it is not always true that when an audit is performed, the full evasion is detected. To account for such partial detection, a detection rate is introduced and the cost component of the decision to evade is modified accordingly. Finally, also borrowing constraints which could incentivize lower compliance and discount rates for delayed detection with respect to the evasion event itself are taken into account. Unlike the first two models, the TCS uses actual empirical evidence from IRS audit data when calibrating the parameters of the MABS model in order to achieve an outcome which is as close as possible to observed levels of compliance.

1.6.3 Experimental Economics - Calibration with Experimental Data

In an agent-based exercise simulating tax compliance behavior of small-business owners Bloomquist (2011) employs a relatively simple evolutionary game-theoretic 5 Experimental Economies and Tax Evasion: The Order Beyond the Market 95 approach and uses experimental data in order to calibrate the model programmed in NetLogo. Taxpayers this time are of four different types: Honest, Strategic, Defiant, and Random. Behavior of small-business owners is hypothesized to be similar to that of laboratory subjects and in order to prove this claim the author relies on thirdparty experimental results (Alm and McKee (2006); Alm et al. (2008)). In these taxcompliance experiments subjects were tested for various behavioral mechanisms, but for comparability reasons only data of the "no treatment" subject pools was extrapolated for the calibration purpose at issue. Compliance rate histograms clearly exhibit the typical bimodal distribution around zero and one, hence confirming the existence of honest and defiant taxpayers. Initially the combination of behavioral taxpayer types is selected such that laboratory results in terms of mean and mode are matched as closely as possible. In this first run no neighborhood effects were included because laboratory subjects were not able to see what others were doing. From the comparison of the simulation results for five different audit rates from 0 to 0.40 with and without risk aversion,[6] it can be inferred that the inclusion of risk aversion with probability weighting gives more precise results for audit rates ranging from 0 to 0.10, confirming the findings of Bernasconi (1998). As a second step four different scenarios[7] of agent-based modeling were matched to real data gathered from the IRS National Research Program (NRP) study. The subsample of individuals with income stemming only from Sole Proprietorship (Schedule C - Profit or Loss From Business) was used and cases of

[6] Risk aversion in this case describes the mechanism of overweighting small probabilities of being audited as described (also in Bernasconi (1998)).

[7] Risk aversion versus no risk aversion and neighborhood effects versus no neighborhood effects.

overcompliance were normalized to full compliance. None of these four simulations was able to match the average NRP compliance rate at which point the agent group defined as "random taxpayers" was excluded from the runs, assuming that small-business owners exhibit less random behavior than the students who participated in the lab. The best match with the new setup was found in absence of neighborhood effects, which is also clear from the comparison of the histogram of compliance rates of the simulation with real data excluding neighborhood effects from playing a dominant role.

1.6.4 A Model of Citizenship

Pellizzari and Rizzi (2014) developed an agent-based model of tax compliance that contemplates two types of agents: taxpayers and the government. Taxpayers maximize their utility based on net income and also considering the perceived level of public expenditure. Again, the role of the public good is considered as being relevant to the tax-compliance decision, but in a more sophisticated and dynamic way than we already saw in Mittone and Patelli (2000). In addition, an array of individual characteristics are included - namely, risk aversion; relative preference for public expenditure; an innate tendency to pay taxes, which we could define as intrinsic tax morale; and, as in Mittone and Patelli (2000), group conformity modeled as the expectation about other agents' compliance behavior. Audit probabilities are not exactly known and inferred subjectively by observing audit dynamics among peers. Moreover, the authors propose three different institutional frameworks based on the power a government is able to exert (high, average, and low). The taxpayers are divided into three types based on their level of "citizenship," which is defined as a combination of the preference for public expenditure, group conformity, and tax morale. Under a weak enforcement regime and with a low level of citizenship the authors observed high levels of tax evasion. Even with high levels of enforcement, full compliance cannot be achieved. Moreover, ceteris paribus, the higher the tax rate, the lower the compliance level of their agents. This finding is in contrast with the predictions of the theoretical EUT model developed by Yitzhaki (1974) but confirms a large body of empirical evidence. Government power still plays a significant role in societies with an average level of citizenship, even though elasticities decrease with respect to low levels of citizenship. Finally, the role of government almost disappears for high levels of citizenship, where tax evasion approaches near-zero levels regardless of the enforcement regime. Overall, citizenship is found to have a larger marginal effect on compliance than enforcement power by the government, but both concepts are necessary to enhance the level of tax compliance in a society. To take the findings further, it could also be interesting to read these results in light of the somewhat contrasting prescriptions coming from the literature on motivation crowding-out, surveyed for example by Frey and Jegen (2001): the intrinsic motivation to pay taxes could therefore be decreased by the mere presence of an extrinsic mechanism of punishment.

1.7 Conclusions

The aim of this survey was to take the reader on a tour through the very rich body of existing literature on tax evasion. Not only the plurality of determinants of tax compliance, but also the multiplicity of methodologies and analytical approaches make it a challenging task to provide policy makers with useful indications. We started from the classical models of tax evasion which represent an ideal starting point for more realistic considerations and modifications. Thereafter, institutional, social and psychological factors, among others, have been found to be highly relevant when a taxpayer decides how much to declare to the tax authority. Also, survey data and laboratory experiments have shown that interaction among taxpayers in the form of peer effects and psychic costs of evasion cannot be disregarded. As the famous physicist Stephen Hawking so wisely predicted, "the next century will be the century of complexity." We now face the moment in which we have to tame and coordinate all these different elements in order to maintain the informative value of new findings. To do so, the ceteris paribus approach and the search for equilibria should be relaxed, giving space to a new notion of *ceteris mutabilibus*[8] and asymptotic dynamics. Innovative tools are needed, and the impressive computing power of modern devices serves this purpose well. We presented some examples of agent-based models to provide the reader with food for thought by highlighting their flexibility in accounting for both heterogeneous agents as well as different kinds of parameters, in addition to the possibility for interaction and the dynamic nature of such simulations. Additionally, individual characteristics can be matched with evidence stemming from human subjects tested in the lab, as we have seen in Bloomquist (2011). The future focus of research on tax compliance should hence be on continuing the multidisciplinary approach in determining the drivers of tax evasion, on one hand, while properly administering, elaborating and integrating old and new findings, on the other, with the final aim being to provide policymakers with ever-improving policy advice on how to increase the overall level of tax compliance.

[8] "All else changeable."

Chapter 2

Catch Me If You Can: Testing the reduction of compound lotteries axiom in a tax compliance experiment

The experiment was designed and conducted with the support of Prof. Michele Bernasconi, whom I also thank for all his additional helpful comments and suggestions.

Abstract: *Our paper tests the Reduction of Compound Lotteries Axiom in a tax evasion experiment. Differently to existing theoretical and empirical models, we disentangle the probability of audit, testing it - ceteris paribus - against a more realistic situation with a probability of audit and a subsequent probability of detection of the undeclared income. Various authors have shown that abstract framing in human subject experiments often leads to a violation of the ROCL axiom and also in our framed set-up we do observe statistically significant differences between the two treatments, in contrast with the predictions of Expected Utility Theory. Furthermore we find evidence that gender plays a significant role, both with regards to the reduction principle, and in post-audit responses. In the light of our results, we conclude that (1) audit and detection should be considered separately in theoretical and empirical models of tax compliance, (2) dynamic rather than static approaches are preferable in order to account for post-audit responses and (3) gender-specific behavioral patterns must not be disregarded in policy design.*

Keywords: Reduction of compound lotteries - Tax evasion - Bomb crater effect - Laboratory experiments

2.1 Introduction

As outlined by numerous authors, such as Bar-Hillel (1973), Kahneman and Tversky (1979) and Harrison et al. (2015), the *reduction of compound lotteries axiom (ROCL)*, one of the fundamental underpinnings of expected utility theory, does not seem to hold when experimentally tested. As a consequence a number of alternative approaches have been developed and discussed up to the present (see for example Segal (1990), Bernasconi and Loomes (1992), and Tversky and Kahneman (1992)).

Generally, these and similar experiments embed the reduction axiom in a neutral setting with subjects facing a lottery-type choice problem which is presented to them as a gamble. While a sterile context-free approach might on one side be alluring when testing theoretical models, as it allows for a careful control of a number of variables, it is also to be considered that the very lack of a realistic framing might produce results that fail to provide reliable insights when the aim is to apply economic theory to policy impact evaluation. Existing research that specifically tests context-free against realistic framing has come to contrasting conclusions (see for example Alm et al. (1992b) and Abbink and Hennig-Schmidt (2006) versus Laury and Taylor (2008) and Voors et al. (2012)). With our experimental assessment we take a step forward from "Speaking to Theorists" which happens in standard laboratory environments testing economic theory, towards "Whispering in the Ears of Princes"(Roth, 1987), where the latter approach is meant to bridge the gap between real policy concerns and theoretical economic models.

In our study we frame the decision as a choice of tax compliance and divide the overall probability of being audited and detected into one probability of being audited and one separate probability of being detected. To the best of our knowledge, none of the numerous experiments on tax evasion feature a formal distinction between the audit and the detection rate. Our null hypothesis is that participants behave accordingly to existing economic models of rational choice that view decision makers as being able and willing to reduce compound events to simple ones by applying straightforward algebra. If this hypothesis was accepted, we should not find any treatment effects in the single-stage set-up compared to the two-stage equivalent.

However, our results show that subjects do not reduce lotteries according to the reduction axiom and that this effect is increasing with the overall audit-detection rate. Also, they exhibit a different sensitivity toward audit and detection rate in post-audit and post-fine responses. Finally, we find significant gender effects that seem to indicate that both, the application of the ROCL axiom and post-audit responses depend on whether the decision maker is male or female.

In section 2.2 of this chapter, we present the classical models of tax evasion as a starting point and testable null hypothesis. We also introduce the reader to the existing literature on the reduction axiom, its non-expected utility alternatives and its application to our tax compliance case. In section 2.3 the experiment is described in detail and results are presented in section 2.4. Besides a direct test of the reduction of compound lotteries axiom, our experimental design also allows for a precise assessment of post-audit responses which are discussed in the regression analyses. Section 3.4 concludes.

2.2 Theoretical Background

2.2.1 Tax Evasion

The Classical Model of Allingham and Sandmo

The problem of tax evasion, or in more positive terms, individuals' attitudes towards tax compliance has been extensively analyzed from numerous perspectives. Within the economics literature, the first to provide a theoretical framework were Allingham and Sandmo (1972) who developed an adapted version of Becker's (1968) formulation of crime and punishment. In the original Beckerian model the decision whether to commit a crime is driven by economic considerations, with potential outlaws weighing expected benefits in terms of monetary and psychic income against expected costs given by a pecuniary sanction. They are assumed to be *homines oeconomici* acting in a self-interested and perfectly rational manner. The individuals' decision model proposed by Becker (1968) is nested into a broader social welfare analysis aimed at providing optimal public and private policy responses to criminal behavior.

In a similar fashion, the decision-maker in Allingham and Sandmo (1972) (henceforth A-S) acts in the guise of a taxpayer who is again assumed to be a rational expected-utility maximizer faced with a dilemma of fiscal compliance.

The expected-utility maximization problem described by equation (2.1) is represented using the same notation as in the original A-S specification in the attempt to facilitate the readers' references to existing research. It weighs the classical two prospects of tax compliance with the taxpayer having to decide on how much to declare (X):

$$E[U] = pU(W - \tau X - \pi(W - X)) + (1 - p)U(W - \tau X)) \tag{2.1}$$

The first prospect is the situation in which the taxpayer is audited at a probability p. In this case the authority discovers their full actual amount of income W and will apply a fine π if partial compliance or full evasion is detected. Hence, the taxpayer's actual gross income W is reduced by the taxes they paid on the income they decided to declare, given by τX, and by the penalty rate π on the undeclared part of income $(W - X)$. As it is intuitive, the penalty rate π must be higher than the tax rate τ.

In the second situation the taxpayer is not audited by the authority, hence getting away with the eventual misdemeanor. In this case they keep their full actual income W, reduced only by the taxes on the income they decided to declare τX.

The probability of being caught p, the constant tax rate τ and the penalty rate π[1], as well as the actual income W, are given exogenously in this basic model. Taxpayers are assumed to be risk averse with their marginal utility being everywhere positive and strictly decreasing with $U'(W) > 0$ and $U''(W) < 0$.

Thus, if there was no uncertainty about the actions of the tax authority, the maximization problem yields two corner solutions, resulting in the following optimal choices for the only argument X: full compliance ($X = W$) in case of a certain audit with $p = 1$, resulting in the outcome $W - \tau W$, or $(1 - \tau)W$, which is the true net income, and total evasion $X = 0$ if detection is known to be impossible ($p = 0$) with the taxpayer getting to keep his full actual income W.

In the intermediate cases where $0 < p < 1$, given the other exogenous parameters, they need to find an interior solution for the optimal amount of declared income X maximizing their expected utility. To keep notation clear, we keep on following the lines of A-S, but substitute the two prospects with C (caught) and NC (not caught):

$$\begin{aligned} C &= W - \tau X - \pi(W - X) \\ NC &= W - \tau X \end{aligned} \tag{2.2}$$

The first-order condition for the maximization problem (2.1) can be written as

$$\frac{dE[U]}{dX} = p(\pi - \tau)U'(C) - (1 - p)\tau U'(NC) = 0 \tag{2.3}$$

By rearranging the terms, we obtain the marginal rate of substitution (MRS) and may assess the parameter values for an interior solution to exist:

[1] Allingham and Sandmo (1972) point out that the penalty rate is assumed to be of pecuniary and not penal nature and that the same τ might be "uncertain from the point of view of the taxpayers", but decide to ignore these points at this level in order to simplify the theoretical analysis.

$$\frac{U'(C)}{U'(NC)} = \frac{(1-p)\tau}{p(\pi-\tau)} \tag{2.4}$$

If the marginal utilities in these two states are equal, such that $U'(C) = U'(NC)$, then the decision maker is indifferent between being caught and not being caught for every declared amount X, which is exactly when the penalty weighted by the probability of being fined is actuarially equivalent to the tax rate τ:

$$\begin{aligned} 1 &= \frac{(1-p)\tau}{p(\pi-\tau)} \\ p(\pi-\tau) &= (1-p)\tau \\ p\pi &= \tau \end{aligned} \tag{2.5}$$

In this case, the risk-averse taxpayer would choose to declare the full amount of income. The same would hold true if the probability-weighted fine rate exceeded the tax rate, such that $p\pi > \tau$. On the other hand, parameter values for which $p\pi < \tau$, with $1 > p > 0$ and $\pi > \tau$, increase the marginal return to undercompliance and the optimal solution will be an interior maximum with $0 < X < W$[2].

The Yitzhaki Objection

Soon thereafter Yitzhaki (1974), in the attempt to bring the model closer to reality and eliminate contradictory results, proposed an approach in which fines are proportional to the amount of evaded taxes instead of the undisclosed proportion of income. In that way, as the tax rate increases cheating will decrease. In order to decide on the level of tax

[2] As already mentioned, the assumption of strict concavity of a function f(.) prescribes that it must be strictly decreasing with $u''(.) < 0$. Hence the second-order condition for the solution to be indeed a maximum is satisfied by the strict concavity of the utility function

$$\frac{d^2E[U]}{dX^2} = p(\pi-\tau)^2U''(C) + (1-p)\tau^2U''(NC) < 0 \tag{2.6}$$

compliance, taxpayers must therefore maximize the following expected utility function with respect to the income they are going to declare[3] expressed by x:

$$EU = (1-p)U(v + \tau(y-x)) + pU(v - \phi\tau(y-x)) \tag{2.7}$$

The baseline value v refers to the net income under the assumption of full compliance $v = y(1-\tau)$, which is gross income minus taxes. The parameter p describes again the probability of getting caught in a random audit, therefore the prospect related to $(1-p)$ represents the case where no audit is performed. Hence, the hypothetically correct net income is augmented by what we coin the *cheater's premium*, given by the taxes they have not paid on the undeclared part of income, *i.e.* the total gross income y minus the declared income x. In the second situation, the audit is performed with probability p, and again we start from the baseline value v, but now the evader's correct net income is reduced by the *cheater's penalty* as they has to pay not only the evaded taxes, but also a fine. The multiplier (fine rate) is given by $\phi > 1$ and is this time proportional to the evaded tax τ.

The first-order condition for optimal evasion is given by

$$\frac{dE[U]}{dX} = -(1-p)\tau U'(NC) + p(\phi\tau - \tau)U'(C) = 0 \tag{2.8}$$

Again, we compare marginal utilities of the two prospects and find the parameter values for which the taxpayer is indifferent between full compliance and evasion:

$$\frac{U'(C)}{U'(NC)} = \frac{(1-p)\tau}{p\tau(\phi-1)} \tag{2.9}$$

[3]see Andreoni et al. (1998) for a similar mathematical representation.

$$
\begin{aligned}
1 &= \frac{(1-p)}{p(\phi-1)} \\
p(\phi-1) &= (1-p) \\
p\phi &= 1
\end{aligned}
\tag{2.10}
$$

As a result, the risk-averse taxpayer will be fully compliant for parameter values $p\phi \geq 1$, whereas they will find an interior solution optimal as long as $p\phi < 1$. Consequently, given the intuitive condition that the fine has to be as least as high as the evaded amount of tax, *i.e.* $\phi > 1$, it must be that the audit rate p is strictly smaller than one. For parameter values $\phi = \pi/\tau$ the model yields the same solutions as in the A-S set-up[4].

In the A-S case it is not clear how an increase in the tax rate will impact on the evaded amount, as there are both, an income effect which is due to the decrease of wealth exerting negative pressure on evasion, coupled with a substitution effect. The latter goes in the opposite direction making evasion more attractive as the relative price of compliance increases with rising tax rates. The sign of the overall effect ultimately depends on the shape of the utility function. In the Yitzhaki model on the other hand, an increase in the tax rate only leads to a negative income effect, hence under the assumption of decreasing absolute risk aversion, the fully compliant taxpayer will have less disposable income and prefer to evade less.

Both models treat all parameters as orthogonal and exogenously assigned. Given the decision makers' utility function, the tax authority can thus increase compliance by acting on the fine (rate), the tax rate and the audit probability. In the next section we will spend some words on how sanctions and audit probabilities influence tax compliance and why we finally concentrated our analysis on the latter.

[4]The second-order condition is satisfied by the strict concavity of the utility function

$$
\frac{d^2E[U]}{dx^2} = (1-p)\tau^2U''(NC) + p(\phi-1)\phi\tau^2U''(NC) < 0 \tag{2.11}
$$

The relevance of p

A wide body of literature exists on the impact of sanctions and the probability of them to be imposed. Friedland (1982) for example conducted a tax evasion experiment on law students and found that responsiveness to information about threat probability (*audit*) is higher than to information about threat magnitudes (*fines*). Furthermore Slemrod et al. (2001) find significant positive reactions in tax compliance as a consequence to information feedback on forthcoming audit probabilities in a controlled experiment on taxpayers from Minnesota. Maciejovsky et al. (2007), Guala and Mittone (2005) and Choo et al. (2013) study the lagged effect of previous audits on future compliance behavior, finding similar results of negative post-audit responses.

The importance of being sanctioned to the tax compliance decision underpins our research idea, which however takes the argument a little more into detail. If we consider a simple, realistic situation of tax audit, it is clearly not always the case that an investigation automatically leads to the detection of the full amount of undeclared income. To accommodate this possibility Feinstein (1991) for example develops a fractional detection model that allows for partial detection and estimates a multiplier of approximately 2 in the U.S. for the TCMP[5] years 1982 and 1985. That is, if a certain amount of evaded income X is detected, the latent true level of tax evasion is twice as high. In their comprehensive work about tax compliance, Andreoni et al. (1998) address the issue of partial detection in section 7.2. highlighting the resulting bias when classifying taxpayers into honest and cheaters. An explicit distinction between audit rate and detection probability is made by Kleven et al. (2011) who describes evasion behavior in an experiment in Denmark. For third-party reported income, they observed a low level of evasion, whereas evasion was consistent for self-reported income. This difference was described as being due to a lower level of the perceived detection probability for the latter. Also, they conjecture

[5] Tax Compliance Measurement Programs

that past audits may have a negative impact on the perceived probability of detection. They assume, however, that the overall detection probability is a product of the probability of audit and the probability of detection conditional on audit and that the overall (perceived) detection rate depends on the amount of evaded income.

The actual application of a fine may, in fact, depend on a number of different factors, such as transparency of the tax code, efficiency of the tax authority, the observability of the taxable income and ultimately also on the cognitive and socio-demographic characteristics of the taxpayer. The experimental laboratory represents one approach for the isolation of such effects, but to the best of our knowledge, experiments distinguishing formally between the probability of being audited and the probability of the evaded income being detected by the tax authority have yet to be conducted. As of today, audit and detection are widely used as synonyms within the existing literature on tax compliance decisions.

2.2.2 The Reduction of Compound Lotteries Axiom

Auditing - from one-stage to two-stage

The underlying assumptions of a theoretical model on tax evasion which considers only one audit/detection probability p must be either that detection of the undeclared income is guaranteed in the case of an audit, or that taxpayers respect the Reduction of Compound Lotteries axiom. The first of these hypothetical assumptions could suffer from lack of realism as it relies on the strong conjecture that there is no information asymmetry in the audit procedure, whereas the second refers to a specific cognitive characteristic of the taxpayer which we aim to test with our study.

The reduction of compound lotteries (ROCL henceforth) axiom classifies the decision maker as rational, able and willing to mentally reduce multi-stage compound lotteries into single-stage lotteries by multiplying probabilities, as shown in the simple example in figure 2.1.

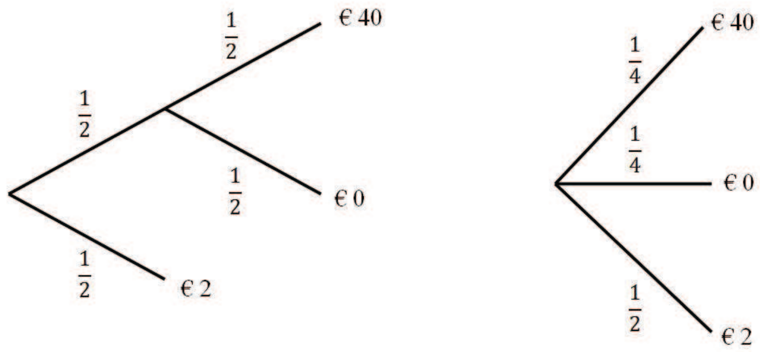


Figure 2.1: Reduction of Compound Lotteries

In the first lottery the decision-maker has a 50% chance of winning 2 euro and a 50% of winning a ticket to another lottery. In that case they again have the chance to get either 40 euro or nothing with equal chance. If the ROCL was not violated, the decision-maker should be betting the same amount of money on both lotteries, as the single-stage lottery on the right side merely compounds the probabilities of the two-stage equivalent.

Similarly, in our setting expected utility maximizing taxpayers should not exhibit any difference in behavior between the following two scenarios:

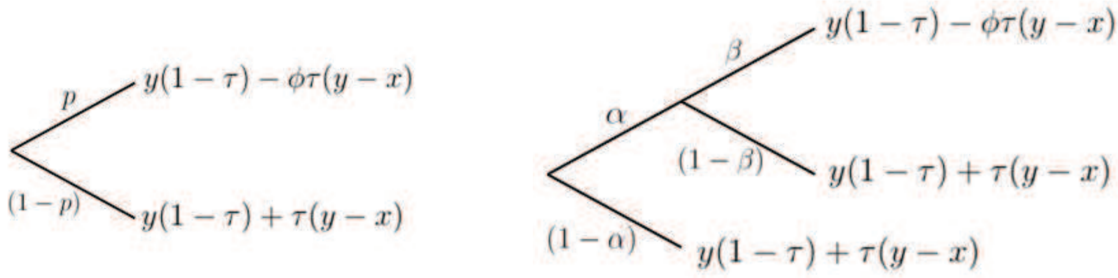


Figure 2.2: One-Stage versus Two-stage evasion decisions

In the one-stage scenario on the left there are two possible outcomes based on the audit rate p, whereas the two-stage scenario on the right side features an audit rate α and a probability of detection β, where $p = \alpha\beta$ and $(1 - p) = \alpha(1 - \beta) + (1 - \alpha) = 1 - \alpha\beta$

Our testable null hypothesis is that decision makers behave in line with the axioms of expected utility theory by being indifferent in terms of declared income between the following single-stage and the two-stage environment:

$$EU = (1-p)U(v+\tau(y-x)) + pU(v-\phi\tau(y-x)) \tag{2.12}$$

$$EU = (1-\alpha\beta)U(v+\tau(y-x)) + \alpha\beta U(v-\phi\tau(y-x)) \tag{2.13}$$

The focus of our study is therefore to assess whether $p = \alpha\beta$. Our alternative hypothesis is that the ROCL axiom is violated in which case we would observe compliance rates that differ in the one-stage with respect to the two-stage treatment.

The independence axiom and the reduction principle

The reduction principle goes back to the early axiomatization of expected utility of von Neumann and Morgenstern (1944, p. 26) and was described by the authors as follows: *"This*[6] *is the statement that it is irrelevant whether a combination of two constituents is obtained in two successive steps, - first the probabilities* $\alpha, (1-\alpha)$*, then the probabilities* $\beta, (1-\beta)$*; or in one operation, - the probabilities* $\gamma, (1-\gamma)$ *where* $\gamma = \alpha\beta$*."*. Even though the authors themselves admit to the possibility of there existing a (dis-)utility of gambling which the reduction axiom - among all EUT axioms - comes closest to exclude, they defend the legitimacy of their axiomatization with the lack of a better theory capable of accounting for psychological factors. The ROCL axiom was again taken up by Samuelson (1952, p. 671), who - when describing the equivalence of *compound income-probability situations* - reminds the reader that *"only algebra, not human behavior, is involved in this definition"*. Also the latter author opens up to relaxing the independence axiom - which comprises ROCL - under certain circumstances, for example if timing indifference is not

[6] axiom (3:C:b) under (3:C) *Algebra of Combining*

assumed. He writes as follows: *"Having blown cold concerning the theory, I should like to conclude by again blowing hot. It is desirable, though not mandatory, that opponents of the Bernoulli theory should make a direct attack upon the independence axiom. Thus, in dealing with compound lotteries with prizes of the form* (A_1)*, or* (A_2) *and* (B_1) *or* (B_2)*, you might explicitly introduce some time interval between drawings, so that suspense elements might enter to contaminate the choices; etc."*.

To provide one of the numerous examples of how the reduction axiom could fail, we extract the comparison between problems 3, 4 and 10 from Kahneman and Tversky (1979). The decision tasks were presented to the experimental subjects in the following manner:

> **Problem 3:** Choose between a sure outcome of 3000 (X_1) and the lottery $Y_1 = (4000, 0.8; 0, 0.2)$
>
> **Problem 4:** Choose between the lottery $X_2 = (3000, 0.25; 0, 0.75)$ and the lottery $Y_2 = (4000, 0.2; 0, 0.8)$
>
> **Problem 10:** Consider the two-stage game which has a 75% chance of not winning anything and a 25% probability of getting to a second stage. In the latter case you have a choice between getting 3000 (X_1) for sure and playing lottery $Y_1 = (4000, 0.8; 0, 0.2)$. The choice must be made before the game starts.

The problems can be represented as decision trees in figure 2.3. Problems 3 and 4 were answered by 95 subjects, out of which 80% preferred the sure option X_1 of problem 3, whereas 65% chose Y_2 over X_2 in problem 4[7]. By considering that the trees in problem 4 are the reduced forms of the lotteries displayed in problem 10, we can immediately spot visually that the (mixture) independence axiom of expected utility was not satisfied[8].

[7] The authors observe a *reflection effect* where results are mirrored when the experiment is conducted with losses instead of gains, which is consistent with the predictions of Prospect Theory, the core of Kahneman and Tversky (1979)

[8] In order to be able to distinguish between the two different types of independence axioms for single-stage and two-stage lotteries, Segal (1990) introduces the compound independence axiom for the lat-

According to the mixture independence axiom, if $X \succ Y$ then this preference relation must also hold if both prospects are convexly combined with a common Z, such that $pX + (1-p)Z \succ pY + (1-p)Z$. In the example of Kahneman and Tversky (1979), given that $X_1 \succ Y_1$, their convex combination $(0.25, 0.75)$ with Z, which in this case is given by the degenerate lottery $(0, 1)$ at the bottom 75% branch, should produce the preference relation $X_2 \succ Y_2$. This violation of EUT is called the *common ratio effect* (Cubitt et al., 1998; Nebout and Dubois, 2014) and could be explained by a preference for certainty in the reported example[9]. The reduction of compound lotteries axiom, on the other hand, requires individuals to be indifferent between the lottery A (B) of problem 10 and the single-stage counterpart X_2 (Y_2) of problem 4. Yet, of 141 participants, 78% chose option A over option B, which is clearly inconsistent with the preferences elicited in problem 4. The ROCL axiom was therefore violated, with the two-stage lottery B not being evaluated in the same manner as its actuarially equivalent single-stage counterpart Y_2.

However, the choices of problem 10 are consistent with those of problem 3, which means that in the case of a salient (common) first stage, individuals seem to behave consistently with the predictions of the compound independence axiom. Kahneman and Tversky (1979) conjecture an *isolation effect* when their subjects attach the same preferences to Problem 10 as to problem 3. In other words, they mentally eliminate the common part given by $\{(0.75, (.); 0.25, Z)\}$ and concentrate only on the two respective sub-lotteries, that are equal to X_1 and Y_1.

As explained in depth by Segal (1990), the mixture independence axiom can only be considered valid if both, the compound independence axiom and the reduction of compound lotteries axiom are jointly respected. The same author suggests in a dedicated chapter (Segal, 1992) that the mechanism with which people reduce two-stage lotteries is

ter, whereas mixture independence refers to the independence axiom in the classical von Neumann-Morgenstern sense.

[9] Kahneman and Tversky (1979) interpret the preference reversal as *certainty effect*

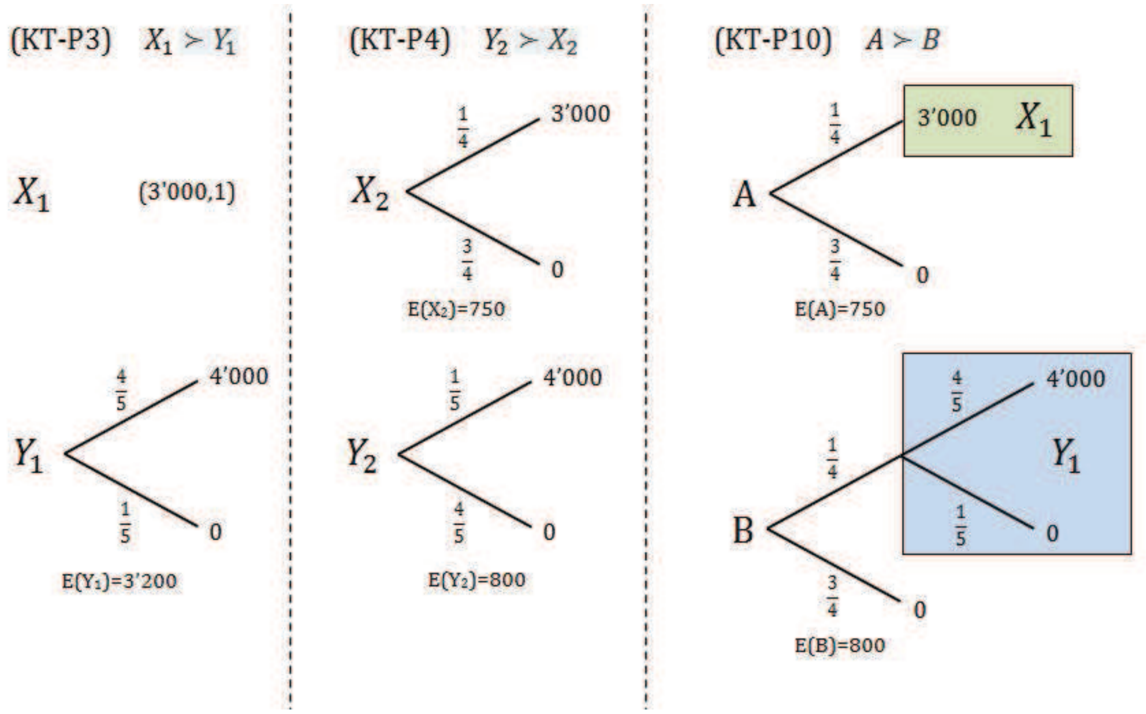


Figure 2.3: Decision trees and preferences of problems 3, 4 and 10 of Kahneman and Tversky (1979)

that of the compound independence axiom, in particular by the technique of *substitution of certainty equivalents*. Moreover, and for that very reason, he concludes that CIA and ROCL are not compatible in the same analysis of compound events as they refer to two distinct ways of reducing subsequent stochastic stages.

Machina (1989) provides an extensive survey of the most common examples of violations of the independence axiom applied to two-stage lotteries. However, he assumes ROCL to hold and interprets the failure of the mixture independence axiom as a direct consequence of the compound independence axiom.

Numerous other authors, such as Bar-Hillel (1973), Segal (1987) Bernasconi (1992), Starmer and Sugden (1991), Nebout and Dubois (2014) and Harrison et al. (2015) found that ROCL does not hold and a number of alternative approaches have been developed in order to accommodate such findings.

Alternative two-stage evaluation functionals

We have already mentioned one model which was introduced by Segal (1987) and consists in a folding-back technique which starts at the final tree nodes. The decision maker substitutes the lotteries at the final branches with their certainty equivalents, ending up with simple lotteries to be evaluated. If the decision maker is indifferent between a simple lottery and its certainty equivalent, the compound independence axiom is satisfied, but the same cannot be said about ROCL.

By transforming probabilities according to the anticipated (rank-dependent) utility theory of Quiggin (1982), he shows how ROCL becomes just a special case when the probability transformation function $g(p)$ is linear. The advantage of this model is that it can account for pessimism and optimism, *i.e.* it can assign higher (lower) weights to extreme probabilistic values. The procedure consists in ranking the outcomes from lowest to highest, such that $X = (x_1, p_1, ..., x_n, x_n)$, where $x_1 \leq x_2 \leq ... \leq x_n$. The lottery X is evaluated by adapting outcomes with a utility function $u(x)$ and by weighing probabilities with $g(p)$, where $g(0) = 0$ and $g(1) = 1$. The overall value function for lottery X would take the following form:

$$V(X) = u(x_n)g(p_n) + \sum_{i=1}^{n-1} u(x_i)[g(\sum_{j=i}^{n} p_j - g(\sum_{j=i+1}^{n} p_j)]$$

Taking our example represented as probability trees in figure 2.2, we call the prospect of getting away with evasion NC (not caught), whereas C (caught) refers to the prospect where evaded taxes and a fine have to be paid. Moreover, assuming for simplicity that $u(x) = x$, the overall value function for our two-stage set-up takes the following form:

$$V(2S) = g(\alpha)[g(\beta)NC + f(1-\beta)C] + g(1-\alpha)C$$

As the weighting function $g(p)$ is applied to the cumulative probability distribution, another property of this model is that first-order stochastic dominance is not violated. Segal (1990) also discusses a weaker alternative to the reduction axiom when comparing two-stage lotteries, which is given by the compound dominance axioms. Strong compound (stochastic) dominance is implied by ROCL whereas the converse is not true.

The advantages of rank-dependent utility were taken up also by Tversky and Kahneman (1992) who further developed Prospect Theory into a more elaborate Cumulative Prospect Theory. By allowing for utility functions to behave differently based on whether the decision maker feels to be in the loss or in the gain domain with respect to their reference point, the latter theory offers an even more flexible tool able to account for non-expected utility deviations due to different behavioral and psychological drivers. On similar lines, also Regret Theory by Loomes and Sugden (1982) allows for the possibility of subjective probability distortions and conjectures that the decision maker elaborates prospects based on the ex-ante juxtaposition of expected regret and rejoice once the lottery is over.

In conclusion of this section, assuming that the tax auditor is not perfect in detecting the evaded amount of income, the theoretical and empirical representation of the audit/detection mechanism as one single probability within an expected utility framework requires the strong assumption of ROCL. As we have seen, when applied to lottery experiments, this is most of the times not the case. Our experiments specifically tests for the ROCL axiom in a tax compliance setting and - in case the null hypothesis was to be rejected - calls for a reconsideration on how tax evasion is modeled.

2.3 The Experiment

2.3.1 Experimental Design

The control groups in our experiment face a standard tax compliance decision, where the audit rate p, the tax rate τ and the fine rate ϕ are exogenously given[10]. The participants in the treatment groups instead have to decide how much of their income to disclose based on a situation where both, audit and detection rates α and β are given separately, featuring the same percentages in compound terms, as in the control treatments. There is no uncertainty attached to any of the probabilities and they are communicated to the participants upfront. Linking our design back to the environment of experimental lottery games, the latter treatment corresponds to the two-stage lottery, whereas the degenerate first treatment is a framed equivalent to a single-stage lottery.

In our tax compliance experiment participants are asked to decide, in each of fourteen subsequent periods, how much of their assigned income they would like to declare given a certain tax rate, a fine rate and a probability of being audited and caught. In order to add some variation to the flow, their exogenous income per period was drawn randomly from a discrete uniform distribution ranging from 80 to 160 ECU (*Experimental Currency Units*) in steps of 10. At the end of the experiment total profits in ECU were converted at a rate of 180 ECU= 1€. We conducted six different sessions, suitable for between-subject analyses. After period 14, participants were provided with a new set of instructions. Those who just concluded the one-stage (two-stage) treatment then had to decide in a two-stage (one-stage) environment respectively. In that way we also gathered data for within-subject assessments.

Even though we may face the risk of income effects due to wealth accumulation over time, we carefully assessed our particular case and ultimately decided to opt for a pay-

[10]The control group represents the degenerate case where detection is certain, so referring to our two-stage set-up, where $\beta = 1$ and $\alpha = p$

off mechanism that rewards each single period as opposed to a random-lottery incentive procedure. Problems relating to potential biases that could arise with the use of random-lottery incentive mechanisms (RLIM henceforth) have been analyzed for example by Starmer and Sugden (1991), Harrison and Swarthout (2014) and Harrison et al. (2015). The "1-in-K" payment method, where $K > 1$, might potentially create distortions when basic axioms of decision theory are experimentally tested. This is due to the fact that a procedure that randomly draws one period to be paid out to the subjects at the end of the experiment adds an additional stage to the lottery choice task and might influence the decisions of the participants. The laboratory experiment conducted in Harrison et al. (2015) consisted of two different treatments. The first group was presented with a battery of 40 different choices between a single-stage lottery and its two-stage counterpart. Here a "1-in-40" RLIM was used which randomly drew one of the 40 lottery outcomes for final payment. The second group had to solve a one-shot choice problem which was drawn randomly from the battery of 40 pairs, the lottery was played and they were paid with the "1-in-1" payment protocol. Harrison et al. (2015) found no violation of the ROCL axiom in the "1-in-1" subject group, whereas ROCL was violated in the "1-in-40" treatment. As the authors state themselves, this could also be due to other drivers than the incentive mechanism, such as the lack of time inducing the first subject group to choose faster hence think less about the reduction of compound lotteries. In his work about preference reversals and the independence axiom, Holt (1986) describes how the use of a RLIM could lead to a preference reversal if the independence axiom is not satisfied. This reversal is due to a "dilution" of choices as a consequence of the additional probabilistic stage. He also concludes that the RLIM does not elicit true preferences if the reduction principle holds. The ROCL principle does not hold in the experiment conducted by Starmer and Sugden (1991) which leads the authors to conclude that "experimental researchers need not be too concerned about this particular [RLIM] problem".

In conclusion, as we were conducting an exploratory experiment in which the very test of the ROCL axiom was the focus of our study, we could therefore not state ex ante that the independence axiom was valid, hence implementing RLIM beforehand would have been somewhat "bipolar" in our case (see Harrison and Swarthout (2014) for an interpretation of the term in this context).

2.3.2 Procedures and Experimental Flow

Our computerized laboratory experiment was programmed and conducted using the software z-Tree (Fischbacher, 2007)[11] on three different occasions between February and March 2014. In order to enhance realism and consistency of our experimental design and to minimize potential latent cultural confounders (see for example Alm and Torgler (2006) for a cross-country study on differing levels of tax morale), particular attention was paid to the use of the Italian language throughout the entire process of recruitment, instruction and experimental programming. We selected only those subjects with fluency in Italian and contacted them on-line through an internal platform inviting them to register for a paid experiment on individual decision making at the Ca' Foscari Laboratory for Experimental Economics in Venice, Italy. Our final subject pool consisted of 95 participants, most of them from Ca' Foscari undergraduate and master programs, who were then divided into six different sessions.

At the beginning of each session subjects were asked to take a seat in front of one of the laboratory computers, each of them separated by cardboard dividers in order to prevent participants from being observed or from seeing what others were doing. Once seated in their cubicles, they were invited to read the paper instructions carefully[12] that had

[11] Z-Tree is a user-friendly and widely used toolbox for programming and conducting computerized economic experiments. Each subject is assigned one individual client, a z-Leaf that transmits the responses to the server of the experimenter, the z-Tree.

[12] The full version of the instructions in Italian is included in the appendix 2.A, English translations are available upon request.

been handed out to each of them, but which were also read out loud by the experimenter. Figure 2.4 depicts the experimental flow of the 2-stage treatment[13].

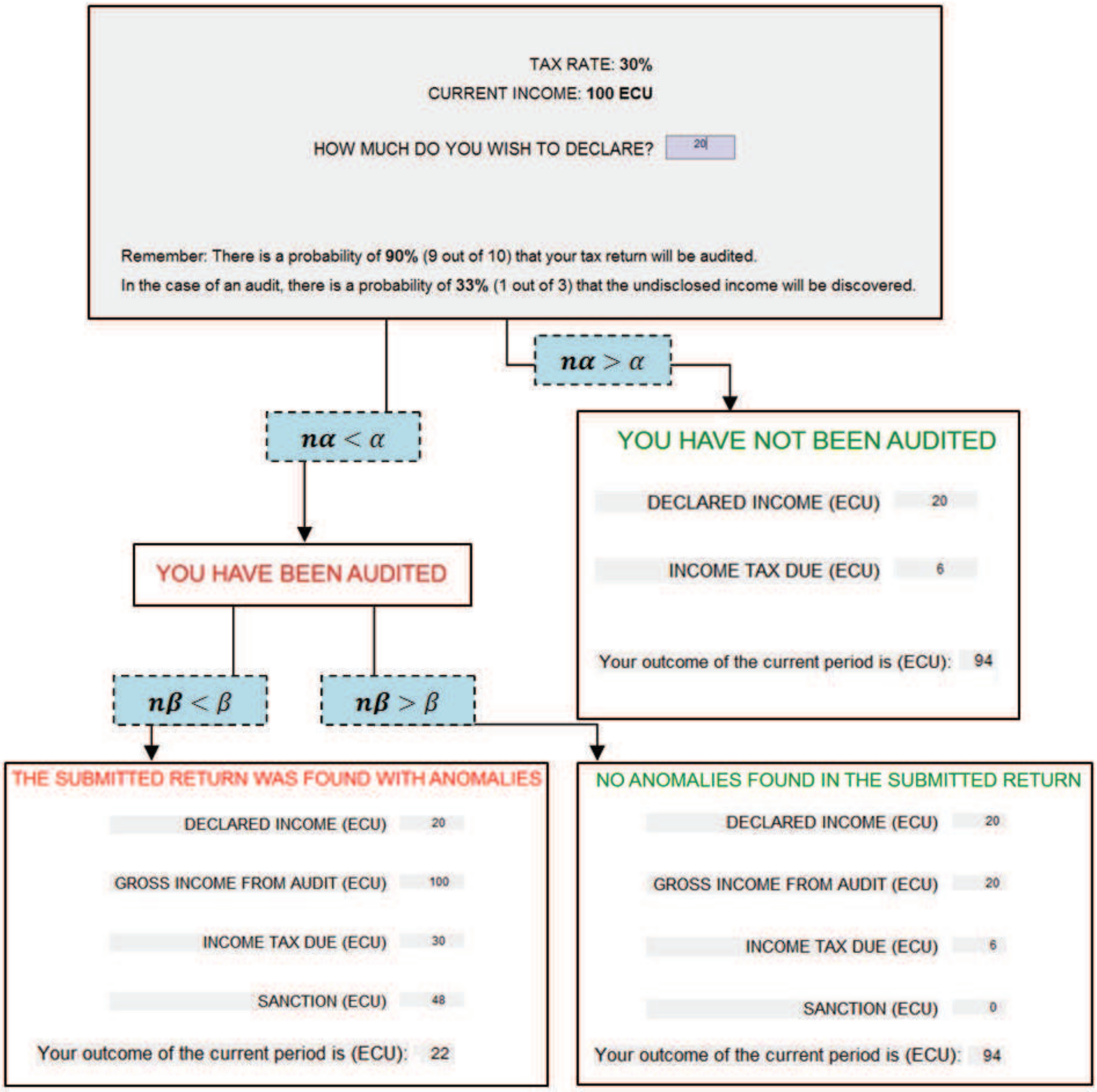


Figure 2.4: Experimental flow - 2-stage treatment - the 90/33 treatment

On the first screen of the experiment subjects are informed about their gross income of that period and asked to take a decision about the part of income they wish to declare.

[13]The original experiment was conducted in Italian, the English translation is provided merely for an illustrative purpose. The original treatment in Italian is available upon request.

To take their decision they are also provided with a calculator which can be accessed by clicking on the corresponding icon on the bottom left of the screen. The next picture lasts for 3 seconds and confirms the reception of the report, reminding the participant that the report might be audited. In the background, a random number $n\alpha$ (np in the one-stage treatment) from 0 to 1 is now drawn to determine whether an audit will be performed. If $n\alpha > \alpha$ ($np > p$), no audit is performed and subjects in both treatments are shown their final result which is given by their gross income minus the taxes calculated on the declared amount. An audit is performed whenever $n\alpha < \alpha$ ($np < p$) in which case a blinking text lasting 8 seconds appears on the screen informing the participant of an audit in progress. In the one-stage treatment if no evasion is detected, the final outcome of the period - gross income minus taxes - is displayed. If, on the other hand, the declared amount is found to be lower than the actual gross income, they are shown a summary containing the sanction resulting from the audit. In the two-stage treatment, another background process starts which draws a new random number $n\beta$. If $n\beta$ is lower than the detection rate β and the declared amount X was lower than the gross income Y, the undeclared income is detected and the participant is shown the screen containing the sanction. In the case of $n\beta > \beta$ or of full compliance $X = Y$ they end up visualizing the "green" summary on the right side of the left branch in Figure 2.4. To conclude, a short questionnaire on demographics, risk aversion, fiscal understanding and beliefs was administered[14] and the participants were invited to collect their total payoff.

2.3.3 Experimental parameters

Table 2.1 provides an overview of the six sessions and the corresponding parameters.

Both, tax rates and fine rates were held constant at 30% and twice the evaded taxes respectively, whereas the overall probability of being detected was either 10%, 20% or

[14] Questionnaire available upon request

Session	τ	fine	Income Y	subjects	Rounds 1-14		Compound	Rounds 15-28		Compound
					α	β	p	α	β	p
1139	0.3	2	u[80,160]	18	0.2	1	0.2	0.9	$^1/_9$	0.1
1330	0.3	2	u[80,160]	15	0.9	$^1/_3$	0.3	0.1	1	0.1
1005	0.3	2	u[80,160]	13	0.5	0.4	0.2	0.3	1	0.3
1137	0.3	2	u[80,160]	16	0.9	$^1/_9$	0.1	0.2	1	0.2
1318	0.3	2	u[80,160]	18	0.1	1	0.1	0.9	$^1/_3$	0.3
1001	0.3	2	u[80,160]	15	0.3	1	0.3	0.5	0.4	0.2
			TOTAL	95						

Table 2.1: Overview of experimental treatments

30% in compound terms. In each session participants played both, the one-stage and the two-stage task for fourteen rounds each, but with different compound probabilities resulting in a total of 28 rounds. Note that our subjects were not told how many rounds and treatments they were about to play. By keeping some degree of uncertainty about the time horizon we wanted to reduce endgame effects to a minimum. This phenomenon of extreme choices at the very end of experiments has been observed by Selten and Stoecker (1986) and analyzed more in-depth by Reuben and Suetens (2012). It mainly occurs in experiments characterized by past and/or future interaction between participants. Even though our subjects do not interact in any way, we wanted to be on the safe side. One of the possibilities of an endgame effect in this context could be a temptation to perform some sort of extreme actions at the end of the game.

2.4 Results

2.4.1 Mean compliance rates

We have collected a total of 2660 income reports handed in by 95 subjects over 28 periods. The compliance rate was calculated as a ratio between declared income and total gross income. Under the assumption of risk neutrality, a utility-maximizing agent should declare zero income. However, nearly a third of all income reports, 29% (767 reports) were honest, fully declaring the assigned gross income, whereas around the same percentage, 27% (729 reports) declared zero income.

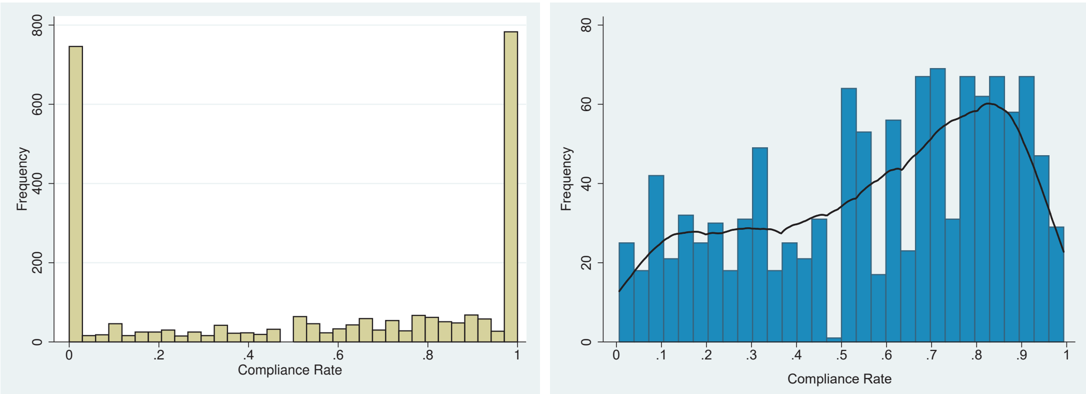


Figure 2.5: Histograms overall compliance and partial compliance with kernel density function

The histogram of all compliance rates is reported on the left side of figure 2.5. The two peaks at 0 and 1 are the corner solutions described and indicate censored-type behavior.

The remaining 1164 reports of partial compliance are distributed between 0 and 1 (excluded) as shown in figure 2.5 on the right. The overlaid kernel density function highlights an average tendency towards declaring more than 50% of the gross income.

Table 2.2 summarizes the overall average of the compliance rates for the six treatment variations, divided into whether they stem from the first or the second phase.

		1st phase		2nd phase	
		1-Stage	2-Stage	1-stage	2-stage
Overall (compound) probability of being detected	10%	.5342	.5348	.3870	.3315
	20%	.5664	.6406	.4337	.4663
	30%	.5426	.8470	.6645	.6140

Table 2.2: Average compliance rates by treatment

First, we consider only the treatments conducted in the first phase. A non-parametric Mann–Whitney U test[15] shows that there is no significant difference in average compliance

[15] A t-test is not applicable as compliance rates are bounded between 0 and 1 and can thus not be assumed to be normally distributed. The Shapiro-Wilk test of normality clearly rejects the a normal distribution.

rates in the 10% set-up between the 1-stage and the 2-stage (audit rate 90% - detection rate $^1/_9$) treatment: the null hypothesis of equality in means cannot be rejected at a level of p=.73. However, for the 20% compound detection rate we find statistically significant differences (z = -2.166 and, p<.05) in the 1-stage vs. the 2-stage (audit rate 50% - detection rate 40%) treatment, which become very strong when the compound probability of detection increases to 30% (z = -8.602, p<.001), where the 2-stage treatment was divided into a 90% audit probability and a $^1/_3$ detection rate.

In other words, as the compound percentage level increases, the average compliance rates become significantly higher when subjects are asked to reason with two distinct probabilities, one for the audit rate and one for the detection rate, as opposed to when they are confronted with only one single rate of audit with automatic detection, though both of them yield the same in compound terms, hence indicating a violation of the reduction of compound lotteries axiom.

In the second phase, however, non-parametric differences in means between the one-stage and the two-stage treatment are not significant anymore, which could indicate a learning effect induced by the administration of the second treatment. By being given a similar, yet slightly different set-up, they could have understood our final purpose and decided to take a major computational effort. What is interesting to observe, is that the mean compliance rates of the first phase (see table 2.2, 1st phase versus 2nd phase) are significantly different, with a compliance path that seems to converge towards an emphasized attention towards compound detection rates. Compliance rates in the second phase are significantly different in a vertical sense, *i.e.* with respect to the overall probability of being detected. In appendix 2.B we report a graphical representation of the level of compliance for each individual in the 28 periods[16]. Not surprisingly, a Wilcoxon

[16]The first four digits of the individual ID in the graph refer to the session identifiers that are reported in the first columns of table 2.1.

matched-pairs signed-ranks test shows that within-subject means of the two phases are not the same.

2.4.2 Period-wise analysis

Phase 1

As a next step we look at our experimental data from a period-wise perspective.

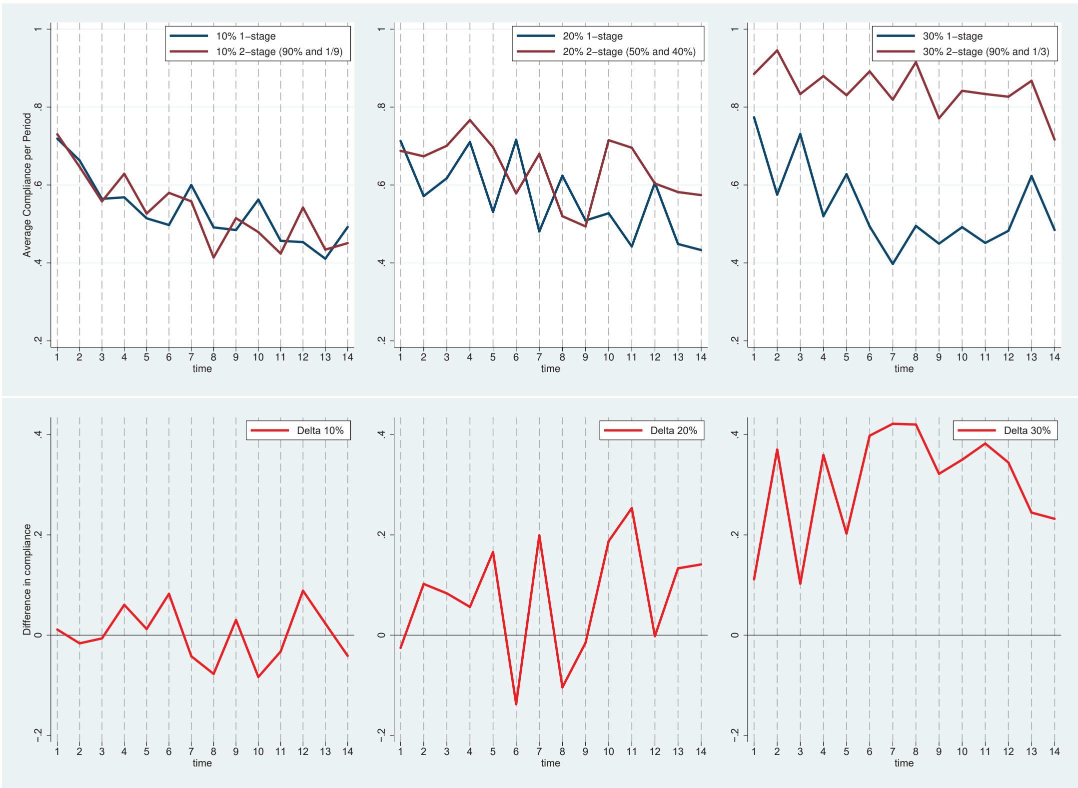


Figure 2.6: Average compliance rates per period and differences - Phase 1

The upper set of graphs in figure 2.6 compares the average compliance rates for each decision period of the one-stage to the two-stage treatment for all three compound parameter values, 10%, 20% and 30% considering only phase 1 (1330 reports).

Differences between the period averages of the two-stage treatment and the one-stage control are displayed in the lower set of figures[17]. Shifts in absolute terms between compliance rates over time are not statistically significant for the 10% and the 20% treatment levels though the latter exhibits noisier behavior. However, a sharp and highly significant shift is observed for the 30% treatment where compliance in the two-stage set-up is around 25% higher (p=0.001).

Phase 2

The graphical representation of the average compliance rates per period in phase 2 of figure 2.7 confirm that there is shift in compliance for the three different audit rates.

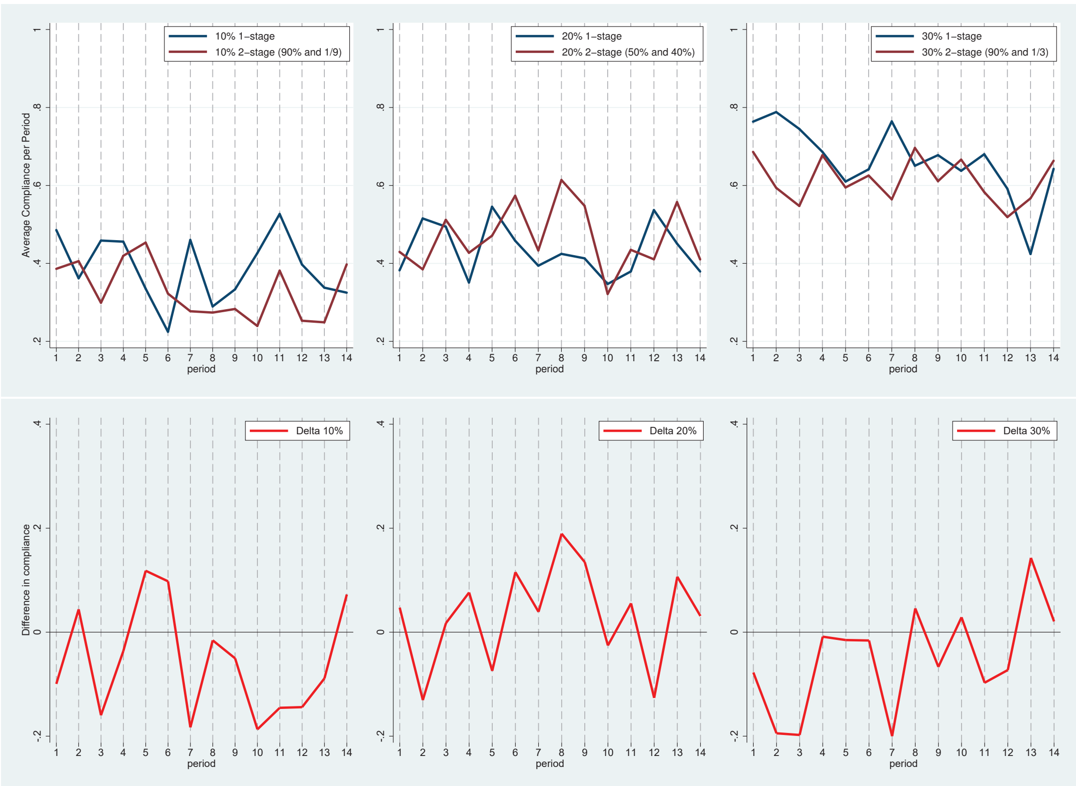


Figure 2.7: Average compliance rates per period and differences - Phase 2

[17] see Bernasconi et al. (2014) for a similar graphical representation

However, compliance does not differ significantly based on the stage-treatment, indicating that no ROCL violations were detectable in the second phase.

2.4.3 Regression analysis

For our regression analysis we chose a random-effects Tobit Model with the dependent variable, reported income, being censored from the left at zero and from the right at the gross income endowment of the period. In addition, a random-effects model can account for individual effects in the data. We discovered a total of three policy-inert participants who always declare their full income in each of the 28 periods they played, thus we excluded them from our analysis assuming they have a strong intrinsic motivation for their standardized behavior and are thus not affected by policy. Including them would not change the significance and signs of our results. However, it does have an impact on the level of the regression coefficients.

Income is defined as the amount of ECU assigned to the subjects each period and it was drawn from a uniform distribution between 80 and 160 ECU in steps of 10 ECU. With the regressor *Caught evading in (t-1)* we identify those partial compliers that fell into our clutches and had to pay a fine in the previous period (t-1). For a better insight into underlying gender effects, we display post-audit results separately for male and female subjects in the appendix. The *2-stage* dummy variable is our treatment variable and takes value 1 when the subject was asked to decide in the two-stage environment and 0 for the one-stage set-up.

Econ Discipline identifies those who have a study background in economics and thus supposedly a higher ex-ante level of ease with quantitative tasks. It is worth mentioning however, that (Rubinstein, 2006b,a) found mixed results with respect to profit-maximizing attitudes and quantitative skills of economics students. What emerges clearly from his experiments is that students with training in economics exhibit a significantly higher overall zeal towards profit maximization regardless of the moral implications of their choices. On the other hand, in a comparative analysis of a choice that requires some mathematical skills, their answers did not differ with respect to other study disciplines. Given these findings, we too might face difficulties in interpreting the compliance level of economics students as this could be due to their level of tax morale, their quantitative skills or both.

Compound detection rate refers to the overall final detection probability, which corresponds to p in the one-stage treatment and $\alpha * \beta$ in the two-stage treatment and takes values 10%, 20% and 30%. The time trend over all periods is identified by the variable *Time*. The level of *risk propensity* is elicited in a questionnaire which was given to the participants at the end of the experiment. We asked a general risk question as in Dohmen et al. (2011), *i.e.* "How do you deal with potentially risky situations?". Our variable ranges from 1 (I always avoid risky situations and choose the safer option) to 7 (I am willing to take risks and always choose the riskier option). *Ceteris paribus* individuals with a lower level of risk aversion are expected to declare less, as also found in similar empirical studies by Bazart and Bonein (2014) and Bernasconi et al. (2014). We also control for the amount of time spent to take the compliance decision, which is given by *Seconds to Compliance Decision*. In the alternative models (2)-(4) we added a set of additional determinants of tax compliance: tax morale, perceived tax fairness and gender. The deterrent variable of *Tax Morale* was determined by asking participants whether or not they agree that paying taxes depends on the individual level of morale. A higher level of tax morale should affect compliance positively (see for example Torgler (2002) and Torgler and Schneider (2007)). Conversely, the inverse of perceived tax fairness is measured by the variable *Taxes too high* which is given by the difference between what an individual believes to be the tax rate his or her family unit has to pay on their income and the tax rate he or she considers to be fair. If individuals perceive the tax burden as being too high, we should observe a lower level of overall compliance, as shown also by Andreoni et al. (1998). Finally, we also control for gender with the variable *Female*. However, empirical findings in existing literature appear to be mixed (see Kastlunger et al. (2010) and references therein). Summary statistics for the variables are reported in Appendix 2.C.1.

Regression analysis of the periods 1-14 (phase 1)

As a first step we considered only the data from phase 1, *i.e.*, the first 14 time periods with a total of 1260 income reports with the results of our analysis being displayed in table 2.3.

Looking at the coefficients starting from the top of the table, we start by controlling for income, which varied within a range of 80 to 160 ECU in each period in order to avoid boredom during the experiment. The effect is small and positive, yet not highly significant. The impact of a detection in the previous period is negative and significant, reflecting similar findings of Mittone (2006) and Maciejovsky et al. (2007). This post-audit response might be due to the "Bomb Crater Effect". This effect is a form of "backwards gambler's fallacy". The reason why the two phenomena are not the same is described in Mittone (2006), p 824, who defines the gambler's fallacy (GF) as follows:

"The GF is related to a form of distortion produced by the subjects' perception of the probability distribution of an uncertain event, due to specific sequences of observations. For example, imagine that a gambler is betting on red or black in roulette. After a long sequence of black, her/his expectancy of the occurrence of red increases. As well known, the probability of red and black is always the same, independently from any possible sequence of occurrences, but the gambler feels that the red becomes increasingly more likely."

The Bomb Crater Effect, on the other hand, describes soldiers in wartimes hiding in bomb craters because they believe that where a bomb hits once it does not hit twice. Ac-

Table 2.3: Random-effects Tobit Model (First phase only)

	(1)	(2)	(3)	(4)
Income	0.220*	0.220*	0.214*	0.216*
	(2.31)	(2.31)	(2.25)	(2.28)
Caught evading taxes in (t-1)	-21.30***	-21.26***	-20.75***	-21.04***
	(-3.38)	(-3.38)	(-3.30)	(-3.34)
2-stage	36.02**	35.58**	40.22**	33.46**
	(3.02)	(2.99)	(3.24)	(2.85)
Econ Discipline	-33.38*	-32.40*	-47.52**	-35.36*
	(-2.26)	(-2.19)	(-3.26)	(-2.41)
Compound detection rate	20.83**	20.56**	20.25**	20.53**
	(2.84)	(2.81)	(2.71)	(2.85)
Time	-2.100***	-2.094***	-2.095***	-2.091***
	(-3.62)	(-3.61)	(-3.62)	(-3.61)
Risk propensity	-15.33**	-16.32**		-16.77**
	(-2.83)	(-2.95)		(-3.08)
Seconds to Compliance Decision	0.409***	0.415***	0.409***	0.416***
	(3.31)	(3.35)	(3.30)	(3.37)
Tax Morale		-2.580	-2.480	-3.813
		(-0.79)	(-0.73)	(-1.15)
Taxes too high			-11.19	-10.61
			(-1.73)	(-1.72)
Female			20.68	
			(1.56)	
Constant	96.80***	109.6***	80.65**	138.1***
	(4.32)	(3.95)	(2.71)	(4.30)
ρ	0.404	0.403	0.415	0.393
Observations	1170	1170	1170	1170
Log-likelihood	-3745.016	-3744.706	-3746.589	-3743.248
Wald χ^2	92.399	93.101	88.313	97.031

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

cording to this mechanism, the subjective distortion of the likelihood of events exhibiting in reality invariant objective probabilities, would therefore induce our subjects to think that two subsequent audits are near to impossible. Our data would be in-line with such a hypothesis as we observe a sharp decrease in compliance right after an audit.

A complementary behavioral mechanism explaining our results could be the urge for "loss repair" which is also studied by Maciejovsky et al. (2007) and mentioned in Andreoni et al. (1998). Again Maciejovsky et al. (2007) make an attempt to disentangle misperception of chance from loss repair, but we will return to this topic in the dedicated section 2.4.3 of the paper. For now, it is interesting to note how the significant effect after the experience of having been caught and the payment of a fine seems to be significant only for our male participants, as shown in the interaction coefficients with gender in the appendix 2.C.3. Also, we do not observed lagged effects that exceed the first period immediately after a fine was paid. The core of our research, however, was to test whether subjects are indifferent between being presented with a one-stage tax compliance problem as opposed to its two-stage counterpart. Our regression analysis ultimately confirms that the Reduction of Compound Lotteries Axiom is indeed violated, with the variable *2-stage* exhibiting a strongly positive effect on tax compliance. Even though expected utility theory would predict rational decision makers to be indifferent, our participants declare a significantly higher amount in the 2-stage setting, as opposed to the 1-stage equivalent. Again, when interacting the treatment variable with gender (see appendix 2.C.3 for regression results) we find the somewhat puzzling result of the ROCL axiom being violated only by female subjects.

Not surprisingly, our data also confirms some significant stylized results: the overall probability of getting caught evading taxes, *i.e.*, the compound detection rate, affects compliance positively, withal we observe a negative time trend. Also, dedicating more time to the compliance decision seems to increase overall tax compliance, though only by a very small amount. We do not find any evidence of women declaring overall more of their income, which is in contrast to the findings of Torgler and Valev (2010) that women have a higher aversion towards tax evasion and corruption. This result is robust even when we do not control for risk attitude (results reported in model (3)). As expected, having an economics background reduces tax compliance, which is likely to be due to a higher level of training and not to a different level of tax morale. Overall, our sample does not seem to be driven by tax morale at all when deciding about how much to declare. Out of curiosity and to round up the discussion, we also conducted a separate Ordinal Logit analysis in order to verify the findings of (Rubinstein, 2006b,a). Our subjects with a background in economics did not have a different level of tax morale. Moreover, we found that women are significantly more likely to have a higher level of tax morale, whereas when taxes are perceived as unfair, tax morale decreases which accords with Alm and Torgler (2006). The full regression table can be consulted in appendix 2.C.2.

Finally, the highly significant constant indicates some baseline level of tax compliance, which could be explained by the extensive existing literature on the role of social norms to tax compliance. In our case we did control for Tax Morale as elicited through a direct question in the questionnaire. However, an additional latent level of intrinsic morale, or fear of incurring in socially defective behavior could have been induced by our type of experimental framing.

Regression analysis of the periods 15-28 (phase 2)

When repeating the same analysis in the second phase (within-subject treatment variations, period 15-28) we find some interesting results in dynamic terms, as shown in table 2.4. As predicted by our non-parametric assessment, the sensitivity to audit rates increases significantly in the second phase. Moreover, as the experiment went on, the linear time trend disappears. The effect of a previous fine is stable over time and significant also in the second phase, and again only the male participants are responsible for the negative post-audit response.

However, the significance of our 2-stage variable completely vanishes - for both sexes. One explanation for this non-result in the second phase of our experiment could be that

Table 2.4: Random-effects Tobit Model (Second phase only)

	(1)	(2)	(3)	(4)
Income	0.110	0.110	0.101	0.109
	(0.89)	(0.89)	(0.82)	(0.88)
Caught evading taxes in (t-1)	-21.94**	-22.00**	-22.26**	-21.88*
	(-2.58)	(-2.58)	(-2.61)	(-2.57)
2-stage	-4.511	-4.188	-13.17	-1.976
	(-0.24)	(-0.23)	(-0.70)	(-0.11)
Econ Discipline	-2.463	-1.690	-16.25	-5.388
	(-0.11)	(-0.07)	(-0.72)	(-0.23)
Compound detection rate	47.41***	47.12***	46.10***	48.43***
	(4.19)	(4.16)	(4.02)	(4.29)
Time	-0.771	-0.771	-0.778	-0.771
	(-1.04)	(-1.04)	(-1.05)	(-1.04)
Risk propensity	-20.94*	-21.48*		-21.99*
	(-2.48)	(-2.49)		(-2.57)
Seconds to Compliance Decision	1.275***	1.277***	1.251***	1.278***
	(4.51)	(4.52)	(4.43)	(4.53)
Tax Morale		-1.575	-1.952	-2.847
		(-0.31)	(-0.38)	(-0.55)
Taxes too high			-13.35	-12.09
			(-1.35)	(-1.25)
Female			39.90	
			(1.94)	
Constant	49.30	56.62	16.05	84.88
	(1.53)	(1.42)	(0.39)	(1.86)
ρ	0.513	0.513	0.514	0.507
Observations	1170	1170	1170	1170
Log-likelihood	-3221.564	-3221.516	-3222.142	-3220.741
Wald χ^2	56.330	56.464	55.369	58.288

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

there is a sort of 'numeracy learning'. Even though we presented the subjects with both, probabilities and ratios in order to avoid the numeracy bias described by Reyna and Brainerd (2008) it could be the case that in the second phase participants improved their calculation skills which ultimately resulted in the Reduction of Compound Lotteries Axiom not being violated. Also, they could have simply understood our experiment as they went on to the second phase and were presented with a two-stage (one-stage) decision problem, but having faced a one-stage (two-stage) environment in the previous phase. Furthermore, our student subject pool could be less affected by social norms on tax compliance and sanctioning, thus exhibiting overall weaker effects, similarly to the findings of Choo et al. (2014). Finally, an alternative explanation for the difference between the two phases could lie in latent effects of non-linear probability weighting functionals evolving over time. Under this hypothesis, subjects do not suddenly switch from being a non-expected utility maximizer to a rational choice type, but they adjust their weighting functions dynamically. In the second phase they still apply weighting functions, but for reasons we still have to determine, they become linear making subjects appear as expected utility maximizers with the two-stage variable not being significant anymore.

Furthermore, with all subjects learning over time, we also see how the competitive advantage of an ex-ante quantitative literacy disappears with *Econ Discipline* being not relevant anymore. Also the constant term looses most of its significance, indicating that the intrinsically motivated compliance level, disappears in favor of mechanisms based on individual cognitive characteristics and risk preferences.

A Panel Hurdle Model with censoring and tremble

In this section we present a robustness analysis which represents a more elegant way to account for fully compliant subjects than the somewhat brute-force method we applied in the previous tobit analyses which consisted in excluding the fully compliant types ex ante from the analysis. Differently, in the panel double-hurdle model (see Moffatt (2015), Engel and Moffat (2012), Engel and Moffat (2014) and references therein) a fully compliant subject is classified as a "type". To be classified as a potential evader they have to pass a first hurdle before actually being included in the analysis. Our first hurdle is defined as follows[18]:

[18]see Moffatt (2015) for a more detailed general description of the model and its variations

$$d_i^* = (RiskPropensity)\alpha_1 + (Age)\alpha_2 + \epsilon_{1,i}$$

$$d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{if } d_i^* \leq 0 \end{cases}$$

$$\epsilon_{1,i} \sim N(0,1)$$

where $d_i^* = 1$ if the subject is found to have had at least one event of evasion in the 28 periods and equals zero otherwise.

The second hurdle consists in a censored regression model which is very similar to the tobit model and we use the evasion rate $(income - x)/income$ as the dependent variable, which is censored at 0 and 1 with subject-specific random effects captured by u_i.

$$Ev.rate_{it}^{**} = (Inc.)\beta_1 + (2 - stage)\beta_2 + +(Caught_{(t-1)})\beta_3 + (Compound)\beta_4 + (Time)\beta_5 + (RiskAtt.)\beta_6 + u_i + \epsilon_{2,i}$$

$$Ev.rate_{it}^{*} = \begin{cases} Ev.rate_{it}^{**} & \text{if } Ev.rate_{it}^{**} > 0 \\ 0 & \text{if } Ev.rate_{it}^{**} \leq 0 \\ 1 & \text{if } Ev.rate_{it}^{**} \geq 1 \end{cases}$$

$$\epsilon_{2,it} \sim N(0,\sigma^2); u_i \sim (0,\sigma_u^2)$$

The observed data is therefore defined as follows:

$$y_{it} = d_i y_{it}^*$$

Considering our specific analysis we wanted to define a *Fully-compliant-Type* with less restrictive rules than full compliance in each period, *i.e.* with a cumulative evasion rate of zero. To do so, we used a specification that allows for a tremble parameter ω. In other words, we allow for the possibility of small errors, or some evasion occurrences made by mistake. Figure 2.8 displays the posterior probabilities of being a fully-compliant type with the number of number of times they evaded throughout the 28 period on the x-axis. The tremble parameter ω which was activated on the right side of the figure is compared to the simple model fully-compliant types are classified as such when they never evaded. In our example we used a tremble parameter of 0.02 which is the probability of the participant experiencing a deviance from their natural type due to an error. The idea behind this choice is that of accounting for negligence during the reporting process. These cases are often treated differently in terms of sanctions (see for example Snow and Warren (2005) and Andreoni et al. (1998) bringing the case of the US where fraud has a substantially higher penalty rate compared to non-fraudulent or negligent cases of under-compliance.

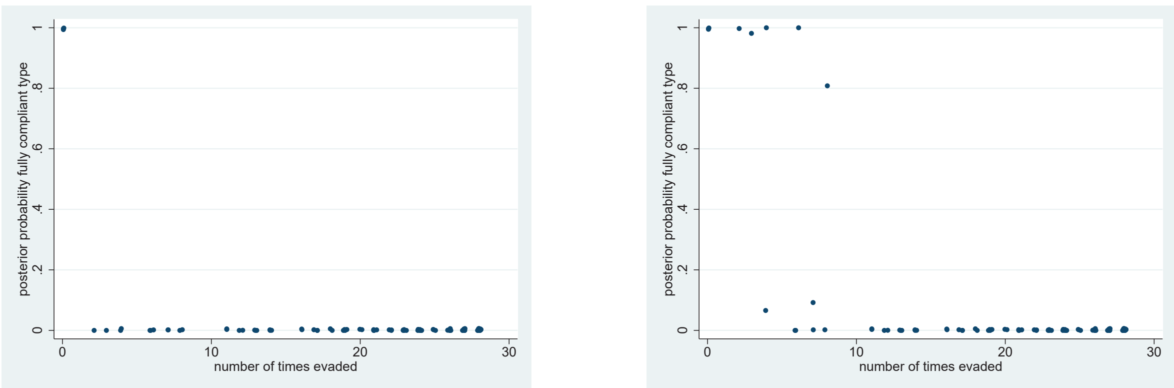


Figure 2.8: Posterior probability of the fully compliant type without tremble (left) and with tremble (right)

An extension of the hurdle model to a set-up that can accommodate two types, the *Total-evader-type* and the *Fully-compliant-Type* is a possibility we wish to explore in future studies whereas for the purpose of the current dataset one first hurdle is sufficiently representative. With the permission of the author, we use an adapted version of the codes for panel double hurdle models contained in Moffatt (2015), pp. 264-282[19]. Table 2.5 compares three different models: the classical panel tobit model (1), similarly to what we have already shown earlier and two panel double hurdle models in (2) and (3).

The first hurdle model includes a tremble parameter ω, whereas the second model considers only fully compliant subjects as *zero-types*. All three models are consistent in significance and sign of the effects, with the panel hurdle models showing some differences in coefficient values compared to the standard panel tobit model. Considering our data and the small number of full compliers, or "pathologically honest" types (Alm et al., 2015), this does not come as a surprise. The aim of this section was to test the panel double-hurdle model as it matches the storyline of our data in a more elegant way than the previous analysis. However, in order to exploit the full potential of this design, the experimental sample to be analyzed should ideally contain a higher number of observations and subjects.

The Bomb Crater Effect and Loss Repair

In the attempt to get a better insight into the interplay between the "Bomb Crater Effect" and the loss repair mechanism we conducted a third analysis. To isolate the two coefficients, we concentrate our third regression only on those subjects that were administered the two-stage treatment as shown in table 2.6.

We then disentangle the effects of having been detected in the preceding period and the reaction due to an audit only without having been detected, where the latter is done to account for effects that are due to the mere excitement experienced with an audit. For the two variables *Caught evading taxes in (t-1)* and *Not caught evading taxes in (t-1)* we consider only partial compliers. For this analysis we included all two-stage experiments

[19] STATA code available upon request

Table 2.5: Evasion Rate - Panel Tobit, Panel DH with Tremble and Panel DH without tremble

	(1) Panel Tobit	(2) Panel DH with tremble	(3) Panel DH No tremble
First Hurdle			
Risk attitude		0.709*	0.374
		(2.56)	(1.34)
Age		-0.00938	0.0420
		(-0.39)	(1.47)
Second Hurdle			
Income	0.00117***	0.00192***	0.00176**
	(3.42)	(3.41)	(3.27)
2-stage	-0.0879***	-0.144***	-0.136***
	(-5.10)	(-5.01)	(-4.88)
Caught evading taxes in (t-1)	0.149***	0.240***	0.245***
	(5.62)	(5.50)	(5.65)
Compound detection rate	-0.163***	-0.280***	-0.271***
	(-13.34)	(-13.54)	(-13.57)
Time	0.0110***	0.0186***	0.0179***
	(10.40)	(10.51)	(10.44)
Risk propensity	0.0891***	0.112***	0.114***
	(5.18)	(4.67)	(4.89)
Observations	2660	2660	2660
Log-likelihood	-1685.760	-2255.105	-2271.443
Wald χ^2	451.037	31.634	44.030
p > χ^2	0.000	0.000	0.000

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

into our regression, those from the first and those from the second phase for a total of 1260 observations, divided into six treatment groups.

When partial compliers were only audited, but not caught evading, the tax audit has not been successful, they did not have to pay a fine and therefore there was no motivation for loss repair. The negative reaction in this case is a pure "Bomb Crater Effect" (BCE) in reference to the first-stage audit probability only. The negative reaction after a detection and payment of a fine on the other hand cannot be easily split, as it includes a stacked BCE ascribable also to the second probability stage, as well as the urge for "loss repair".

Table 2.6: Random-effects Tobit Model (2-stage only)

	Declared Income		
	β	SE	p-value
Income	0.273***	(0.100)	0.006
Not caught evading taxes in (t-1)	-16.054***	(5.775)	0.005
Caught evading taxes in (t-1)	-26.686***	(7.610)	0.000
Econ Discipline	-37.093*	(20.609)	0.072
Compound detection rate	36.351***	(10.097)	0.000
Time	-1.138*	(0.606)	0.060
Risk propensity	-18.242**	(7.547)	0.016
Seconds to Compliance Decision	0.397**	(0.169)	0.019
Constant	97.738***	(27.689)	0.000
ρ	0.566		
Observations	1170		
Log-likelihood	-3398.864		
Wald χ^2	66.316		

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

It seems that the fact of having been caught is stronger than that of an audit only, but a Wald test performed to test the difference in coefficients fails to reject the null hypothesis of equality (Prob > chi2 = 0.1589). The implication of this result is that the negative post-audit effect is likely to be ascribable mainly to the BCE of the audit itself. The loss repair mechanism as a consequence of experienced fines seems to be negligible when looking at the aggregate data. The gender split in table 2.C.5 in the appendix shows that both sexes have a negative post-audit reaction and that the response is equivalent in size and significance. However, when looking at the post-fine behavior, the negative effect for women disappears whereas the male reaction doubles and becomes strongly significant. This indicates that a bomb crater effect and loss repair are likely to play a role in post-fine responses of males. Also, it seems that the experience of a fine in the previous period cancels out the bomb crater effect of the audit for women which translates into an unaffected level of post-fine compliance.

2.5 Concluding notes

With the experimental assessment described in this paper we aimed to test the reduction of compound lotteries axiom by using the classical models of Allingham and Sandmo (1972) and Yitzhaki (1974) of tax evasion. Agents are assumed to be rational expected utility maximizers, who take decisions based on the probability of getting caught in a random audit. Yitzhaki's representation of the tax compliance decision has been challenged numerous times and, similarly to lottery games in the standard expected utility framework, a number of alternative approaches have been proposed. Nonetheless, we needed to go back to the classic model in order to test one very specific axiom of expected utility theory in this context: the reduction of compound lotteries axiom. The inspirational trigger for our assessment lied in considering two separate probability types that enter an individual's decision problem. There is the audit rate on one hand and, conditional on having been audited, we allow for a detection rate. Traditional utility-maximizing models consider only one compound audit/detection rate. We wanted to know whether or not these two stages are reduced following the reduction of compound lotteries axiom and can thus be considered equivalents in theoretical modeling.

We conducted an exploratory tax compliance experiment in order to determine the level of individual compliance in the one-stage control treatment compared to the two-stage setting, where for the latter the overall probability was split into one probability of being subject to an audit and another probability of being actually detected.

In the first phase, our results are highly significant with participants declaring a higher amount in the two-stage experiment, which is confirmed by both, parametric and non-parametric analyses. More in detail, violations of the ROCL axiom are observed in the female subject pool only and become stronger as the joint probability level of being audited and detected increases. In the second phase the reduction axiom is not violated anymore and compliance in both treatment groups converges to one level. The pattern that emerges is of linear nature meaning that over time the only determinant of the compliance decision is the level of compound detection (10%, 20% or 30%).

Concerning post-audit and post-detection responses, we find that reactions to fines are strong and significant, but also that the sole experience of having been audited impacts negatively on compliance in the subsequent period. The separation between past 'audit only' versus past detection experiences allows us to show that there are two separate drivers at work, the Bomb Crater Effect and the loss-repair drive. Also, there are significant differences in gender with females having the same post-audit reactions as men but no significant post-fine responses. Males on the other hand have a highly significant and strongly negative post-fine reaction in compliance which could be an indication of an underlying loss-repair mechanism[20].

Outside the laboratory, however, the perceived detection rate of undeclared income is influenced by multiple factors such as the credibility of institutional efficiency, the level of self-evaluation and exogenous institutional determinants. In a laboratory experiment, where subjects are aware that a computer instead of a real auditor is taking these decisions there are only limited chances to capture the full impact of the detection rate (see also (Alm et al., 1992b)). We therefore suggest culturally nested field experiments as a method to assess how subjective expectations of detection rates affect the level of tax evasion.

Our results add another piece to the complex picture describing the decision-making process of paying taxes. More precisely, it provides support to the argument that the reduction of compound lotteries axiom is violated, which and makes the case for a revision of classical models describing tax compliance. Hence, formal models should consider the audit and the detection rate separately in order to be able to account for their different sensitivity.

[20]The effect, however, could also be a mixed one which includes a double-BCE effect (α-BCE and β-BCE)

We also presented a number of alternative utility specifications and approaches that can be extended to a two-stage setting when modeling tax compliance behavior without relying on the ROCL axiom. Among these approaches we find the substitution of certainty equivalents (Segal, 1992), Cumulative Prospect Theory with rank-dependent utility functions (Tversky and Kahneman, 1992) or the possibility of relaxing the assumption of time indifference by adding elements of suspense, as already suggested by Samuelson (1952).

Besides behavioral drivers that diverge from the expected utility framework, violations of the reduction axiom could also be simply due to the more complex situation of having to evaluate two probability stages which go beyond the numeracy bias we mentioned earlier. Keller (1985) finds for example that the reduction axiom holds more frequently when decision problems are presented with tubes and marbles. This could be a further interesting point to be tested in future research. If it is found that the mere increase in quantitative complexity is the responsible executioner of the reduction axiom, then it should be assessed how individuals evaluate such complex situations. Tax situations are indeed typically of complex nature with a high level of ambiguity attached to probabilities and with taxpayers that on average have not received intensive quantitative training before deciding on their compliance level.

Finally, we recommend the two probability stages not to be aggregated in a theoretical model, the use of dynamic instead of static approaches able to account for lagged reactions and the careful consideration of behavioral gender difference. As a policy implication, the bottom-line message of our experiment would be "If you audit, do it well". The negative effect after an unsuccessful audit should not be disregarded in the cost-benefit analysis of the tax authority. One possible solution to minimize any negative post-audit and post-sanction effects could be pre-announcing the audit. The latter was found to have a positive effect on tax compliance within the same time frame, as found by Slemrod et al. (2001) which might partly neutralize the decrease in compliance in the ensuing periods.

Appendix

2.A Example of instructions - one-stage/first phase

2.A.1 One-stage/first phase

Istruzioni

Benvenuto/a. Grazie per la partecipazione a questa sessione sperimentale. Seguendo attentamente le istruzioni potrai guadagnare una cifra in euro che ti sarà pagata in contanti alla fine della sessione.

A questo esperimento partecipano varie persone. Di ognuno degli altri partecipanti non conoscerai né le scelte, né i guadagni. Ciò significa che nessun'altro partecipante riceverà alcuna informazione sulle tue scelte e sui tuoi guadagni. I tuoi dati saranno trattati in forma completamente anonima. Durante la sessione non è consentito parlare o comunicare in alcun modo con gli altri partecipanti. Se hai una qualsiasi domanda alza la mano e uno degli assistenti verrà a risponderti. Le regole che seguono sono uguali per tutti i partecipanti.

Regole generali.

Le decisioni che dovrai prendere in questo esperimento si riferiscono al pagamento delle imposte. L'esperimento consiste in vari periodi. In ciascun periodo ti verrà assegnato dal computer in modo casuale un reddito che può variare **da 80 a 160 ECU** (*Experimental Currency Units*, la valuta del laboratorio). Le tue decisioni in ogni periodo consistono in quanto dichiarare del reddito che ti è stato assegnato. Puoi dichiarare un qualsiasi ammontare tra 0 e l'intera assegnazione del periodo.
Sul reddito dichiarato devi pagare delle imposte con **un'aliquota del 30%**. Dunque, le imposte che paghi ed il reddito netto del periodo sono pari a

Imposte=0,30*(reddito dichiarato)

Reddito netto = (reddito assegnato) – Imposte

Nella schermata della dichiarazione in basso a sinistra troverai un'icona con una calcolatrice a disposizione per eventuali calcoli. Dopo aver completato la dichiarazione, essa potrebbe essere estratta per un controllo a campione con una probabilità pari al **30% (3 su 10).** La probabilità di estrazione è determinata in modo casuale ed indipendente in ogni periodo. Se subisci un controllo, l'importo che eventualmente non hai dichiarato sarà scoperto. Se hai dichiarato meno del reddito assegnato dovrai pagare le imposte non pagate e una **sanzione aggiuntiva pari a due volte** le imposte non pagate. Facciamo alcuni esempi:

Esempio 1:

- Reddito assegnato: 100

- Reddito dichiarato: 60
  (a) Risultato in assenza di controlli: 100 - 0,30*60 = 100 - 18 = 82 ECU
  (b) Ammontare da pagare in caso di accertamento:
  0.30*40 +2*0,30*40= 12+24 = 36 ECU (Imposte sul reddito non dichiarato + Sanzione)
  (C) Risultato in caso di controllo: 82 – 36 = 46 ECU

Esempio 2:

- Reddito assegnato: 100
- Reddito dichiarato: 100
  Risultato finale: 100 - 0,30*100 = 70 ECU

Esempio 3:

- Reddito assegnato: 100
- Reddito dichiarato: 0
  (a) Risultato in assenza di controlli: 100 ECU
  (b) Ammontare aggiuntivo da pagare in caso di accertamento:
  0.30*100 + 2*0,30*100 = 30 + 60 = 90ECU (Imposte sul reddito non dichiarato + Sanzione)
  Risultato in caso di controllo: 100 – 90= 10 ECU

In ciascun periodo sarai informato del risultato del periodo.

Nel corso dell'esperimento la probabilità di accertamento potrebbe subire delle modificazioni. Nel caso questo succeda sarai informato di tutti i dettagli.

Dopo un certo numero di periodi sarai informato che l'esperimento è concluso. Il tuo pagamento finale è dato dai guadagni accumulati in tutti i periodi dell'esperimento e il tuo guadagno totale sarà convertito in Euro al tasso di 180 ECU = 1 €.

Prima del pagamento, ci sarà un breve questionario. Quindi gli sperimentatori verranno a verificare il tuo guadagno per la partecipazione all'esperimento che ti verrà pagato immediatamente.

2.A.2 Two-stage/second phase

Modifica della probabilità di accertamento.

Nei periodi che seguono tutte le regole generali rimangono le stesse ad eccezione delle probabilità di accertamento. In particolare, come nei periodi precedenti in ciascun periodo ti verrà assegnato dal computer in modo casuale un reddito che può variare **da 80 a 160 ECU** (*Experimental Currency Units*, la valuta del laboratorio). Le tue decisioni in ogni periodo consistono in quanto dichiarare del reddito che ti è stato assegnato. Puoi dichiarare un qualsiasi ammontare tra 0 e l'intera assegnazione del periodo. Sul reddito dichiarato devi pagare delle imposte con **un'aliquota del 30%**.

Le modificazioni riguardano il fatto che, dopo aver completato la dichiarazione, essa potrebbe essere estratta per un controllo con una probabilità pari al **50% (5 su 10).** Se subisci un controllo l'importo che eventualmente non hai dichiarato sarà scoperto con una probabilità del **40% (4 su 10).** Le probabilità di estrazione e dell'esito della verifica sono determinate in modo casuale ed indipendente in ogni periodo.

Come prima, quando viene scoperto che hai dichiarato meno del reddito assegnato dovrai pagare le imposte non pagate e una **sanzione aggiuntiva pari a due volte** le imposte non pagate.

2.B Individual Compliance Rates

2.C Additional Tables

2.C.1 Summary Statistics

Table 2.C.1: Summary Statistics (Panel Data)

Variable		Mean	Std. Dev.	Min	Max	Observations
Declared Income	overall	64.82	52.35	0.00	160.00	N = 2660
	between		29.83	5.00	131.43	n = 95
	within		43.12	-44.97	191.42	T = 28
Income	overall	120.65	23.34	80.00	160.00	N = 2660
	between		5.16	108.21	132.50	n = 95
	within		22.77	68.86	172.43	T = 28
Seconds (Decision)	overall	12.79	19.13	0.00	164.00	N = 2660
	between		9.92	2.79	59.07	n = 95
	within		16.39	-29.82	159.25	T = 28
Female	overall	0.53	0.50	0.00	1.00	N = 2660
	between		0.50	0.00	1.00	n = 95
Econ Discipline	overall	0.72	0.45	0.00	1.00	N = 2604
	between		0.45	0.00	1.00	n = 93
Risk Propensity	overall	2.89	1.21	1.00	6.00	N = 2660
	between		1.22	1.00	6.00	n = 95
Tax Morale	overall	3.96	1.87	0.00	6.00	N = 2660
	between		1.88	0.00	6.00	n = 95
Taxes too high	overall	1.81	1.03	0.00	6.00	N = 2660
	between		1.03	0.00	6.00	n = 95
Trust	overall	2.32	1.22	0.00	5.00	N = 2660
	between		1.22	0.00	5.00	n = 95

The panel data summary statistics were obtained by using the STATA® command *xtsum*. For the within-subjects component, the between-subjects mean $\bar{x}_i$ is subtracted from the value x_{it} and the global mean $\bar{x}$ is added back in order to get comparable results. Within minimum and maximum values are calculated as deviations from individual means (note: to obtain deviation values, the global averages have to be subtracted from the reported minimum and maximum values).

2.C.2 Determinants of tax morale - Ordered Log-odds

Table 2.C.2: Ordinal Logit Tax Morale (Ordered Log-Odds)

	Tax Morale		
	Coef.	SE	p-value
Female	0.807**	(0.387)	0.037
Econ Discipline	0.370	(0.431)	0.390
Taxes too high	-0.576***	(0.220)	0.009
Trust	-0.220	(0.159)	0.165
Observations	93		
Log-likelihood	-160.981		
LR χ^2	11.224		

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

(a)The Brant test is not significant, thus the Parallel Line Assumption is not violated

2.C.3 Effects over time of previous audits/detections

Table 2.C.3: Random-effects Tobit Model (First phase only) - lagged variables and gender effects

	Declared Income		
	β	SE	p-value
Income	0.183*	(0.101)	0.069
Caught evading taxes in (t-1)=1 × male	-32.968***	(9.995)	0.001
Caught evading taxes in (t-1)=1 × female	-13.661	(9.042)	0.131
Caught evading taxes in (t-2)=1 × male	-4.323	(9.448)	0.647
Caught evading taxes in (t-2)=1 × female	-11.955	(9.113)	0.190
2-stage=1 × male	21.884	(14.025)	0.119
2-stage=1 × female	54.218***	(15.870)	0.001
Econ Discipline	-29.068**	(14.700)	0.048
Compound detection rate	23.300***	(7.275)	0.001
Time	-2.288***	(0.650)	0.000
Risk propensity	-14.082***	(5.381)	0.009
Seconds to Compliance Decision	0.321**	(0.138)	0.020
Constant	96.321***	(23.067)	0.000
ρ	0.387		
Observations	1080		
Log-likelihood	-3448.912		
Wald χ^2	85.742		

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

Table 2.C.4: Random-effects Tobit Model (Second phase only) - lagged variables and gender effects

	Declared Income		
	β	SE	p-value
Income	0.101	(0.127)	0.428
Caught evading taxes in (t-1)=1 × male	-66.248***	(13.850)	0.000
Caught evading taxes in (t-1)=1 × female	-1.329	(11.601)	0.909
Caught evading taxes in (t-2)=1 × male	-18.455	(12.639)	0.144
Caught evading taxes in (t-2)=1 × female	7.406	(11.656)	0.525
2-stage=1 × male	-13.725	(24.719)	0.579
2-stage=1 × female	0.438	(20.778)	0.983
Econ Discipline	0.283	(23.467)	0.990
Compound detection rate	46.306***	(11.200)	0.000
Time	-1.118	(0.822)	0.174
Risk propensity	-18.220**	(8.407)	0.030
Seconds to Compliance Decision	1.416***	(0.308)	0.000
Constant	45.698	(32.750)	0.163
ρ	0.514		
Observations	1080		
Log-likelihood	-2945.299		
Wald χ^2	75.319		

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

2.C.4 Bomb Crater Effect - Gender Split

Table 2.C.5: Bomb Crater or Loss Repair - Gender Split

	Declared Income		
	β	SE	p-value
Income	0.284***	(0.100)	0.004
Not caught evading taxes in (t-1)=1 × male	-16.030**	(8.038)	0.046
Not caught evading taxes in (t-1)=1 × female	-16.246**	(7.697)	0.035
Caught evading taxes in (t-1)=1 × male	-38.400***	(10.915)	0.000
Caught evading taxes in (t-1)=1 × female	-16.150	(10.273)	0.116
Econ Discipline	-36.113*	(20.588)	0.079
Compound detection rate	36.671***	(10.082)	0.000
Time	-1.171*	(0.606)	0.053
Risk propensity	-18.111**	(7.537)	0.016
Seconds to Compliance Decision	0.382**	(0.169)	0.023
Constant	95.526***	(27.706)	0.001
ρ	0.566		
Observations	1170		
Log-likelihood	-3397.525		
Wald χ^2	68.998		

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

Chapter 3

Risk attitudes, investment behavior and linguistic variation

This chapter is drawn from a joint work with Matija Kovacic (Ca' Foscari University of Venice, Department of Economics) and Francesco Costantini (University of Udine, Department of Linguistics) which has been published as Kovacic et al. (2016)

Abstract: *This paper explores the relationship between linguistic variation and individual attitudes toward risk and uncertainty. Linguistic variation refers to differences in linguistic forms across languages. According to the linguistic relativity hypothesis, differences in grammatical structures and the vocabulary may affect how speakers of distinct languages perceive and think about the world. We develop a specific linguistic marker that classifies languages according to the number of non-indicative moods in* Irrealis *contexts in their respective grammars. These grammatical categories express situations involving uncertainty, and the frequency of their use may be closely related to the overall degree of uncertainty perceived by individuals. Using data from the Survey of Health, Aging and Retirement in Europe (SHARE) and the World Value Survey (WVS), we show that speakers of languages where non-indicative moods are used more intensively, are on average more risk averse. This evidence holds both across countries and within linguistically heterogeneous countries. The results are robust to the inclusion of an additional set of regressors and several fixed-effect controls for individual characteristics. Finally, we use our linguistic marker to instrument individual attitudes toward risk in the model for financial asset accumulation.*

Keywords: Language - Uncertainty - Risk Aversion - Time Preferences - Asset Accumulation - Instrumental Variables

3.1 Introduction

Consider a situation in which you have to take a decision about something that has an uncertain prospect, may it be related to sports, health or financial choices. It is clear that your personal characteristics and preferences have an influence on how you evaluate the potential outcomes, but would you argue that the language you speak reveals part of how you perceive risk?

This paper proposes an innovative approach to analyze the individual attitudes toward uncertainty and risky behavior based on the *Sapir-Whorf* hypothesis of linguistic relativity. Geeraerts and Cuyckens (2010) and Gentner and Goldin-Meadow (2003) describe the twofold meaning of linguistic relativity which, on one hand, refers to the differences in concepts conveyed by distinct languages and on the other hand denotes that the way concepts are semantically expressed has a direct influence on the cognitive domain of individuals. The basic principle of linguistic relativity hence stems from the idea that differences in grammatical structures and vocabulary may induce speakers of different languages to conceptualize and experience the world differently (Hill and Mannheim, 1992).

The research in support of this hypothesis has mainly concentrated on conceptual contents of languages. For instance, Majid et al. (2004) show that the individual perception of space is influenced to some extent by linguistic spatial frames which directly affect the categorization of space in the cognitive domain. On the other hand, Davies and Corbett (1997), Roberson et al. (1999), and Winawer et al. (2007) suggest that words for colors may influence color perception and the ability of individuals to distinguish between different types of the same basic color. In a recent paper on cross-country differences in gender political quota, Santacreu-Vasut et al. (2013) show that pervasiveness of gender distinctions in grammar is an important correlate for the individual perception of the general role of men and women in the society, which in turn directly influences the extent of regulation of gender political quota, even after controlling for economic development, religion and the political system.

Hence, if speakers of different languages vary in their *Weltanschauung*, some dimensions of linguistic structures may also shape individual preferences and their economic decision-making. The literature on the relationship between linguistic differences and economic behavior, however, is still very poor. The aim of this paper is to fill this gap and to shed some light on the effects of language on risky behavior and attitudes toward uncertainty. To the best of our knowledge, there is only one research article that deals explicitly with the features of linguistic differences and certain aspects of individual economic behavior. In a recent paper on the effect of language on economic behavior, Chen (2013) tests a linguistic-savings hypothesis: when people are required to speak in a distinct way about future events, they take fewer future-oriented actions. According to Chen, seeing the future as contiguous with the present encourages actions that are future-oriented because they are directly connected to the present. Conversely, by using specific verb forms that identify futurity, events are considered farther away and thus appear to be less relevant. The author adopts a future time criterion from typological linguistics discussed in Dahl (2000) and Thieroff (2000), which separates languages into two broad categories: weak and strong Future Time Reference (FTR henceforth) according to how they require speakers to mark the timing of events. Some languages require an explicit verb conjugation in order to distinguish between present and future events (strong FTR languages), while others allow their speakers to talk about the future by using the same verb forms as for present events (weak FTR languages). The author then examines how these differences correlate with future oriented behavior such as saving, smoking, physical activity, and wealth accumulation by retirement. The association between weak future time reference and future oriented behavior is strong: speakers of weak FTR languages save more, accumulate more wealth by retirement, smoke less frequently and are more physically active (and, hence, less obese).

The approach adopted in this paper is conceptually in line with Chen (2013) since it relies on a weak version of the linguistic relativity hypothesis. However, it departs from Chen (2013) for at least two reasons. First, we propose to consider the linguistic relativity hypothesis on the background of a different grammatical property and in a different economic context, namely *mood* and *uncertainty*. We develop a new linguistic mapping based on the number of grammatical categories (moods) concerned with the expression of uncertainty. We hypothesize that speakers of languages where these specific grammatical forms are used more frequently perceive the world as being more mutable and uncertain with respect to speakers of languages where these forms are less frequently used, or do not exist at all. Our mapping offers a rigorous, and to the best of our knowledge, the first linguistic mapping related to grammatical treatment of uncertainty. Second, we analyze the correlation between our linguistic markers and individual self-declared risk aversion for a large set of individuals from the Survey of Health, Aging and Retirement in Europe (SHARE) and World Value Survey (WVS). We show that a more intensive use of grammatical forms concerned with the expression of uncertainty is strongly correlated with individual risk attitudes, even after including a rich set of explanatory and control variables, and fixed-effects for individual characteristics. Finally, we estimate a reduced form equation of the probability of investing in risky assets. We use the linguistic marker to instrument the individual attitudes towards risk and quantify a direct effect of risk aversion on the probability of holding risky financial assets.

The paper is structured as follows. In the next section we introduce the issue linguistic relativity and mood, as well as a discussion of the typological distinction used in Chen (2013). In section 3.3 we exploit the relationship between our linguistic marker and individual attitudes toward risk, and estimate the direct separate of risk aversion and time preferences on the probability of holding risky financial assets. Section 3.4 concludes.

3.2 Linguistic Relativity and Economic Behavior

The idea that language categories can influence thought has come to be known as *Sapir-Whorf hypothesis* after Sapir (1921) and Whorf and Carroll (1964) and boasts a long history in the philosophy of language and linguistics which can be traced back at least to Humboldt's (1836) idea of *Innere Sprachform*. Following Geeraerts and Cuyckens (2010), the hypothesis of linguistic relativity encompasses two basic notions: the first being that languages are relative as they vary in their expression of concepts, and the second being that the semantic expression of concepts influences, at least to some extent, conceptualization at the cognitive level. Therefore, speakers of distinct languages may perceive reality differently. For instance, the division of the color spectrum varies between languages. Unlike English, Italian speakers distinguish between three kinds of blues ("blu", "azzurro" and "celeste") and Russian makes an obligatory distinction between lighter blues ("goluboy") and darker blues ("siniy") (Winawer et al. (2007)). According to the linguistic relativity hypothesis, different linguistic structures will make Russians and Italians more sensitive to color discrimination than English speakers.

The linguistic relativity hypothesis has generally been interpreted according to two versions. The "strong" one, also known as *linguistic determinism*, states that linguistic categories control general cognitive variables. This version of the hypothesis, however, has generally been refuted (Pinker, 1994). The "weak" version claims that linguistic categories have some effect on cognitive habits, particularly with respect to memory and categorization. The latter version of the Sapir-Whorf hypothesis was taken to be more feasible and has inspired research on topics such as color perception, shape classification, conditional reasoning, number, space, and time categorization.

If speakers of different languages tend to think and behave differently depending on the language they use, some dimensions of linguistic structures may also shape individuals preferences. Chen (2013) represents the first attempt to analyze the impact of language differences on the cognitive domain and consequently on several aspects of individual economic behavior. The empirical analysis in Chen (2013) uses a typological distinction discussed in Dahl (2000) and Thieroff (2000) whereby there are languages that employ a specific verb morphology for FTR, whereas other languages do not. By adopting the weak version of the Sapir-Whorf hypothesis, Chen (2013) hypothesized that this typological divide has an effect on how speakers conceive time. Specifically, speakers of languages that separate the future from the present tense ("strong FTR" languages) are more prone to dissociate the future from the present compared to speakers of languages that do not employ that specific verb morphology when referring to future events ("weak FTR" or "futureless" languages). As a consequence, this may induce people to perceive the future as being more distant and, as a consequence, to undertake fewer future-oriented actions such as saving, smoking, using condoms, accumulating wealth before retirement, and taking initiatives to enhance long-run health. The association between weak FTR and future oriented behavior in Chen (2013) is strong: speakers of weak FTR languages save more, accumulate more wealth by retirement, smoke less frequently and are more physically active.

This notwithstanding, Dahl (2000) and Thieroff (2000)'s classification appears to have some limitations. First, it sorts languages into one of the two categories - strong and weak FTR languages - within contexts involving prediction, such as weather forecasts - and *only* within these contexts[1]. However, as Chen himself points out, prediction-based contexts are not the only contexts involving FTR. Schedules, plans, ongoing processes having a natural terminus in the future also call for the use of FTR. It follows that there may

[1]The reason for this choice is due to the fact that if a language has an obligatory FTR marking, which is naturally enforced in prediction-based contexts.

be different criteria from the one chosen by Dahl and Thieroff to differentiate between languages based on their use of future tenses. If we consider schedules, for instance, English becomes a weak FTR language since it allows for the use of the present tense in reference to future situations planned according to a schedule, like in the sentence "The train leaves at five o'clock.". Chen eventually decided not to use scheduled and law-like events because he relies on the results of Dahl (1985) and Dahl (2000) which show that in many, if not most, languages they are not treated as having future time reference.

We consider Chen (2013)'s idea of linking language features to economic behavior through the linguistic relativity hypothesis appealing enough to propose a reconsideration based on a different grammatical property and in a different economic context, namely the property of mood and the perception of uncertainty. Following a weak version of the linguistic relativity hypothesis, we conjecture that individual levels of risk aversion are influenced by differences in the intensity of use of indicative versus non-indicative (*i.e., Irrealis*) moods as they assign a different degree of uncertainty to possible situations. In other words, when describing possible or hypothetical situations, the displacement of the actual from the alternative state of facts is perceived as larger when an *Irrealis* mood is used. According to this conjecture, in sentences (1) and (2), for example, the leaving event should be perceived as less uncertain by an English speaker than by an Italian speaker, even though they describe the same possible situation.

(1) *I think he has left.* (English)

(2) *Penso sia partito.* (Italian)
Think-1SG is-SUBJ left
"I think he has left"

The former expresses the leaving situation by resorting to the indicative mood, while the latter has to use a subjunctive (*Irrealis*) mood. In general, by using *Irrealis* more intensely, speakers move from the region of certainty to that of uncertainty, in other words, their latent area of the unknown is greater with respect to their peers who speak a less *Irrealis*-intensive language. They are expected to be more risk averse as the semantic salience of their region of uncertainty increases.

For this purpose we develop a specific linguistic marker defined on the number of non-indicative moods used in *Irrealis* contexts, *i.e.*, contexts that involve grammatical categories concerned with the expression of uncertainty and we relate it to the individual's perception of risk and risky behavior. In what follows we describe the definitions of displacement, modality and mood more in depth, providing also some applied examples and contexts that define our *Irrealis* indicator.

3.2.1 Displacement and Modality

By displacement semanticists mean the specific characteristic of human language whereby language expressions do not only refer to the here and now but are able to range over future, past, potential, possible and even impossible situations (Hockett (1960), Hockett and Altmann (1968)). In that sense, futurity is an instance of displacement within the temporal dimension. Another crucial dimension of displacement is modality, the grammatical category that indicates whether a sentence expresses a fact, a command, a condition, an opinion, a desire, etc. Consider for instance the following sentences where "KONJ" stands for the German *Konjunktiv* mode, "1SG" for first person singular, and "IMP" for imperative mood

By observing sentences (1), (3), and the embedded clause in (2), we notice that they do not concern actual facts, the truth or falsity of the expressions can be decided simply by considering whether the state of facts described in the sentences is true (or false). Sentence (1) does not assert that it is sunny and that the speaker is having a walk; sentence (2) does not assert that the meeting is finished. It may be finished, and the

(1)	*Wenn*	*es*	*sonnig*	*wäre,*	*ginge*	*ich*	*spazieren.*	(German)
	If	it	sunny	be-KONJ	go-KONJ	I	walk	
	"If it were sunny, I would go for a walk."							

(2)	*Penso*	*che*	*la*	*riunione*	*sia*	*finita.*	(Italian)
	Think-**1SG**	that	the	meeting	is	finished	
	"I think the meeting has finished."						

(3)	*Chodżmy*	*do*	*mnie*	*na*	*kawę.*	(Polish)[2]
	Go-**IMP**	to	me	for	coffee	
	"Let's go to my place for a coffee."					

speaker probably believes that it has, but one's belief may turn out to be wrong when actual states of facts are taken into consideration. Sentence (3) does not assert that the speaker is at home having a coffee with the hearer. Sentences (1) to (3) do not refer to actual facts, differently from sentences like "It is sunny today", "I am having a walk", "the meeting has finished", "I'm having a coffee at home with a friend". They refer to possible situations or "possible worlds" (Carnap, 1947), not to real ones. Possible worlds represent alternative state of facts, which cannot be asserted as of the world we actually live in (the "actual world"), and as such they involve the notion of uncertainty.

3.2.2 Mood and *Irrealis* context of use

Mood is the grammatical category concerned with the expression of situations involving the "world" parameter. What grammarians call *indicative*, for instance, is the mood generally used to assert that a proposition is true as of the actual world.[3] To express possible situations languages can use moods other than the indicative. In the German sentence (1) in the previous section verbs are in the so-called *Konjunktiv II*. The embedded clause in Italian in sentence (2) is in the *Subjunctive*. The Polish sentence (3) is in

[3] This does not exclude that the indicative may have modal functions, too.

the *Imperative*. In sentence (2), the English language uses an indicative while Italian uses a non-indicative mood (subjunctive). The difference between indicative and non-indicative moods lies in the fact that they assign a different degree of uncertainty to possible situations. Also Hungarians (see example (4)) use a specific verb morphology to express a hypothetical condition A French speaker, on the other hand, would use a verb in the indicative in an if-clause (see example (5) below), and so would a Romanian or a Dutch speaker (see examples (6) and (7))[4].

(4) (Hungarian)

Ha	*több*	*időm*	***lenne,***	*akkor*	*többet*	*olvasnék.*
If	more	time	be.COND,	then	more	read.COND

"If I had more time, then I would read more."

(5) (French)

S'	*il*	***avait***	***faim,***	*il*	*mangerait.*
If	he	have.IND.IMPF	hunger,	he	eat-COND

"If he were hungry, he would eat."

(6) (Romanian)

Dacă	*te*	***iubea,***	*ţi-o*	*spunea.*
If	you	love-IND.IMPF,	you-it	tell-IND.IMPF

"If he had loved you, he would have told you."

(7) (Dutch)

Als	*hij*	***gekomen***	***was,***	*hadden*	*wij*	*het*	*kunnen*	*doen.*
If	he	come	was.IND,	had	we	it	could	done.

"If he had come, we could have done it."

Some languages have a wide range of morphological moods, some, the most in fact, have a limited number of grammatical categories concerning mood, which are basically the indicative, the imperative and the subjunctive/conditional and others do not have any specific morphological markers for mood[5]. Most importantly, languages may vary as for the contexts of use of different moods. While in all languages the indicative is the

[4] "SUBJ" stands for subjunctive; "COND" for conditional; "IND.IMPF" for imperfect indicative.

[5] With the exclusion of nonfinite moods, like the infinitive or the gerund, most Romance languages have four moods according to traditional grammars: the indicative, the subjunctive, the conditional and the imperative. Most Slavic languages have three moods: the indicative, the conditional and the imperative. German has three moods, too: the indicative, the "Konjunktiv" and the imperative. Northern Germanic

mood used to assert a state of fact and imperative the one to command, the other moods (e.g. subjunctive, conditional, etc.) have different functions and may be used in contexts that vary from language to language. The contexts where *Irrealis* moods are used more consistently from a cross-linguistic viewpoint include the following:

- complements of modal predicates (e.g. to be possible, to be likely, to be necessary, to be probable);

  It is probable that these events were coincidences.

- complements of volitional predicates (e.g. to want, to wish, to desire);

  I wish I hadn't been late for school.

- complements of epistemic (non-factive) predicates (e.g. to think, to believe, to doubt);

  I think we should keep a diverse energy portfolio.

- complements of emotive factive predicates (e.g. to regret, to be happy, to be sad);

  I regret that this joke has garnered so much attention.

- complements of declarative predicates (e.g. to say, to tell, to announce);

  I said that one day in my career bad results will come.

- the protasis (the if-clause) in conditional sentences;

  If he had studied harder, he would have passed the exam.

- the apodosis (the main clause) in a conditional sentences.

languages have only two moods: the indicative and the imperative-subjunctive mood is also mentioned in some traditional grammars, but it has only residual uses and is no longer productive.

If he had studied harder, he would have passed the exam.

For the purpose of our index, we take the extent of use of different *Irrealis* moods in these syntactic contexts as a ratio of use of the different moods in a language in general. We assign a value of 1 to the occurrence of a non-indicative mood in a particular syntactic environment and 0 otherwise. By addition we obtain an indicator (IRR henceforth) of how frequently *Irrealis* forms are used in a language, so that languages can be ranked according to the intensity of use of *Irrealis* moods.[6]. Finally, languages that do not require *Irrealis* moods in any of the context above are called "moodless" languages.

Our linguistic mapping covers 39 languages as listed in Table 3.A.1 in the Appendix. Data on grammatical mood were mainly collected from Rothstein and Thieroff (2010) (RT henceforth) as it is the most comprehensive typological survey on grammatical mood in the languages of Europe (see Appendix for further details). Since not all the data we needed were included in RT, we also collected some additional primary data. For this purpose we worked out a questionnaire compiled by a number of linguists throughout Europe (see full version in the Appendix 3.B). We contacted linguistic experts who were asked to provide a translation of various sentences into their native language and to produce, for each sentence, explanations on which mood they were using in their versions (indicative versus other non-indicative moods to be described).

Table 3.A.1 shows that six languages are moodless, whereas three languages use *Irrealis* moods in all of the six context. The remaining 27 languages range from two to four contexts in which they employ *Irrealis* moods. Thus, significant variation of *Irrealis* moods across languages may represent a good platform for testing the linguistic relativity hypothesis in the context of several economic behaviors involving risk and uncertainty.

[6]Since there is no qualitative difference between contexts in defining IRR, the index sum was calculated using a uniform weighting function.

3.3 Linguistic Differences and Risk Attitudes

3.3.1 Determinants of Risk Aversion

The definition of risk aversion as we use it today was provided by Kenneth Arrow (1965) and John Pratt (1964) who developed a measure of local risk aversion using the von Neumann-Morgenstern utility specification. The resulting Arrow-Pratt measure of individual risk aversion is given by the curvature of an individual's expected utility function $u(w)$ as a function of wealth and defines whether he is risk-neutral, risk-loving or risk-averse. Theoretical, empirical and experimental research aimed at determining the relationship between wealth and risk aversion has given mixed results so far. It is not fully clear whether and to what extent economic agent exhibit increasing (IARA), decreasing (DARA) or constant absolute risk aversion (CARA). For example, Binswanger (1980) used a list of risk gambles with farmers in rural India and found that risk aversion increases with the increase in payoff. Pratt (1964) himself argues that "Utility functions for which r(x)[7] is *decreasing* are logical candidates to use when trying to describe the behavior of people who, one feels, might generally pay less for insurance against a given risk the greater their assets." (pp122-123). Likewise, it seems to be mostly contingent to situational factors whether the coefficient of relative risk aversion is increasing, decreasing or constant. However, as also mentioned in Xie (2000), the most common assumption in existing literature is that of Constant Relative Risk Aversion (CRRA) and as a consequence decreasing absolute risk aversion (DARA). Still, existing results are somewhat inconclusive hence the functional form of the utility function and, as a consequence, the specification of an agent's level of risk tolerance is largely left to the discretionary assumptions of the researcher.

[7] the ARA coefficient

A higher level of agreement has been reached regarding the determinants of risk aversion. Among the most influential explanatory variables we find level of education[8], type of employment[9], wealth and income[10] behavioral habits like smoking or drinking[11], health status[12], marital status[13], family size[14], immigrant status[15] and several other socio - demographic variables, such as age[16], gender [17], race [18], religion[19] and even the experience of macroeconomic shocks[20]. Sapienza et al. (2009) found that MBA students with a higher level of testosterone were more likely to make risky financial choices and engage in careers in the finance sector. Also Dohmen et al. (2005) shows that risk aversion is higher for women and older people. Note how, among all covariates we mentioned, only age, gender and race can be considered as exogenous[21]. With respect to the other correlates it is more difficult to determine the flow of causality.

Building upon existing literature on the role of risk aversion in predicting stock market participation (see Lee et al. (2015) and Kapteyn and Teppa (2011)), we attempt to resolve a crucial issue common to these kind of analyses. Kapteyn and Teppa (2011) find that - as opposed to theoretical predictions and simple intuition - risk aversion does have a significant, yet not very strong effect in explaining portfolio choice. They provide possible explanations related to the way the survey question was asked to the respondents. Yet, less risk averse persons should be more likely to invest in stocks which by definition are

[8]see for example Riley Jr. and Chow (1992), Bellante and Green (2004) and Dohmen et al. (2011)

[9]see for example Hartog et al. (2002)

[10]see for example Barsky et al. (1997) and Guiso and Paiella (2008)

[11]see for example Barsky et al. (1997) and Dohmen et al. (2011)

[12]see for example Hartog et al. (2002) and Bellante and Green (2004)

[13]see for example Sundén and Surette (1998), Halek and Eisenhauer (2001) and Bellante and Green (2004)

[14]see for example Siegel and Hoban (1991) and Lin (2009)

[15]see for example Halek and Eisenhauer (2001)

[16]see for example Halek and Eisenhauer (2001), Lin (2009) and Cohen and Einav (2007)

[17]see for example Jianakoplos and Bernasek (1998), Sundén and Surette (1998), Cohen and Einav (2007), Dohmen et al. (2011) and Hersch (1996)

[18]see for example Hersch (1996), Jianakoplos and Bernasek (1998) and Halek and Eisenhauer (2001)

[19]see for example Barsky et al. (1997) and Halek and Eisenhauer (2001)

[20]see for example Malmendier and Nagel (2011)

[21]However, Dohmen et al. (2011) points out that age might eventually be endogenous if more risk adverse persons live longer

riskier than other kinds of investments and may be subject to dramatic fluctuations due to speculation. Our explanation for the lack of magnitude of the risk aversion coefficient is different. Simple correlation does not say enough about the causal relationship between individuals' risk attitudes and investment behavior. The reduced form equation that relates risk aversion to asset accumulation is highly likely to suffer from reverse causality and omitted variables bias resulting in the endogeneity of the regressor. As discussed earlier, a person's income is a predictor of her level of risk aversion on one hand, as higher income is associated with a lower level of risk aversion, under the common assumption of decreasing absolute risk aversion. At the same time current and expected asset returns represent additional financial income. How can we know whether risk aversion determines stock market participation or stock market returns impact on the individual's risk attitude and to what extent? It becomes clear how asset accumulation and risk aversion may be simultaneously determined. Moreover, it is highly likely that there are several unobservable variables excluded from empirical models that jointly determine risk attitude and investment behavior. Both of the issues discussed will result in the variable of risk-aversion being correlated with the error term. We will try to disentangle the effect of risk aversion on stock-market participation by using an instrumental variable approach.

3.3.2 A simple model of portfolio choice

Linguistic relativity suggests that speakers of languages without mood distinctions should perceive the divide between actual and possible situations differently from speakers of languages that have specific markings for mood. Our main hypothesis stems from the idea that speakers of languages in which non-indicative moods are used more intensively to express potential situations, perceive the world as being more mutable and hence more uncertain. As a consequence, they are expected to be more risk averse than others, and to engage less in any kind of activity with an uncertain outcome.

To illustrate this mechanism, consider a simple economic system populated by two types of individuals, one speaking a low intensity *IRR* language, L, and another speaking a more intensive *IRR* language, H. Both are endowed with the same level of initial wealth, W_0, and decide on the amount A of their wealth to be invested in a risky asset with an uncertain rate of return $\tilde{r}$. The part of wealth that has not been invested in risky asset pays a risk-free interest r_F. The investment horizon is one period and the individuals solve the following maximization problem:

$$\max_{A_i} EU_i(\widetilde{W}_{t=1,i}) = \max EU_i[(1+r_F)(W_0 - A_i) + (1+\tilde{r})A_i], \quad i \in \{L, H\}; \tag{3.1}$$

where $\widetilde{W}_{t=1,i}$ is the individual $i's$ wealth at the end of the period, $U(\cdot)$ is a twice-differentiable strictly concave Bernoulli utility function, and EU follows the axioms of the Von Neumann - Morgenstern expected utility specification. Since $U' > 0$ and $U'' < 0$, both the L-type and the H-type are strictly risk averse. If we assume that their preferences are represented by

$$U_i(\widetilde{W}) = \frac{\widetilde{W}^{1-\gamma_i}}{1-\gamma_i}, \ \gamma_i > 0, \quad i \in \{L, H\};$$

with γ_i being the parameter of the individuals' constant relative risk aversion, and for simplicity set r_F equal to zero, the maximization problem in (1) yields the following first order condition:

$$EU_i^{'} = \frac{E\tilde{r}}{(W_0 + A_i E\tilde{r})^{\gamma_i}} = 0. \tag{3.2}$$

As long as $E\tilde{r} > 0$, both types of individuals choose to invest a positive amount of wealth in risky assets, *i.e.,* $A_i^* > 0$. Moreover, for any W_0, more risk averse individuals invest less with respect to less risk averse ones.

Following our initial intuition, individuals speaking a more intensive *IRR* language (H-type) should be more cautious and, hence, more risk averse with respect to L-type individuals, *ceteris paribus.* As a consequence, for any given level of wealth, H-type individuals invest less in risky assets with respect to L-type individuals, *i.e.,* $A_H^* < A_L^*$. At this point we can state the following testable hypotheses:

Hypothesis 1 ***Direct Effect of Irrealis on Risk Aversion***
Speakers of languages characterized by a more pronounced displacement into uncertainty (H-type) are more risk averse with respect to speakers of languages with a weaker displacement into uncertainty (L-type), ceteris paribus, i.e., $\gamma_H > \gamma_L$.

Hypothesis 2 ***Indirect Effect of Irrealis on Investment Behavior***
If $\gamma_H > \gamma_L$*, speakers of languages with a more pronounced displacement into uncertainty invest less in risky assets, ceteris paribus, i.e.,* $A_H^* < A_L^*$.

In other words, the intensity of displacement into uncertainty directly influences the individuals' attitudes toward risk, and indirectly their propensity to invest in risky assets. The purpose of the proposed mechanism is not to assert that the intensity of displacement into uncertainty represents the main driver of individuals' attitudes toward risk. What our intuition suggests is that this particular linguistic feature may bias the individuals' perception of uncertainty, and consequently strongly correlate with their degree of risk aversion.

Our empirical exercise proceeds in two steps. We first test a direct association between our linguistic marker and the individuals' self-declared risk aversion. We consider one particular aspect of preferences related to financial risk that the individuals are willing to take when making investments. As a robustness check, we also consider another aspect of individual risk preferences related to adventure risk taking. As a next step, we specify a two stage empirical model in which we use our linguistic marker to instrument risk aversion in the reduced form equation for the probability of holding risky assets. In addition, we disentangle the effects of risk and time preferences by using the FTR linguistic marker as a proxy for the individuals' subjective discount rate.

3.3.3 Data and Methods

Our empirical analysis is run on individuals in the Survey on Health, Aging and Retirement in Europe (SHARE henceforth), Waves 2, 4 and 5 - release 5.0.0.[22], and in the two last waves (Waves 5 and 6)[23] of the World Value Survey (WVS henceforth). The respondents in SHARE come from 18 European countries and Israel, speaking 18 different languages[24]. We excluded first-generation immigrants from our analysis[25]. In addition to the entire set of countries, we run separate regressions for individuals living in linguistically heterogeneous countries. There are in total 5 linguistically heterogeneous countries in SHARE speaking 10 different languages with a significant variation of IRR.[26]. As a robustness check, we consider the WVS for individuals speaking 22 different languages in 54 countries. The list of countries and languages in the WVS is reported in Appendix 3.C.

[22]The variable of financial risk preferences in SHARE is not present in Wave 1. Wave 3 is a retrospective survey with a different methodology. For those respondents that had been interviewed in two or more waves, we imputed data from the most recent wave.

[23]In the versions WV5_Data_stata_v_2015_04_18.dta and WV6_Data_stata_v_2016_01_01 (Stata DTA)

[24]The countries and languages in SHARE are Austria (German), Germany (German), Sweden (Swedish), The Netherlands (Dutch), Spain (Spanish and Catalan), Italy (Italian), France (French), Denmark (Danish), Greece (Greek), Switzerland (German, Italian and French), Belgium (French and Dutch), Israel (Russian, Arabic and Hebrew), Czech Republic (Czech), Poland (Polish), Luxembourg (French and German), Hungary (Hungarian), Portugal (Portuguese), Slovenia (Slovenian) and Estonia (Estonian and Russian)

[25]Robustness tests for the immigrant subsample are available upon request.

[26]The list of linguistically heterogeneous countries in SHARE includes: Spain, Belgium, Israel, Luxembourg, and Switzerland. Estonia was not included as there are too few observations for Russian. There are 10 different languages spoken in these countries (IRR in parentheses): Arabic (4), Catalan (3), Dutch (2), French (3), German (2), Hebrew (0), Italian (6), Russian (4) and Spanish (4).

The immigrant status in WVS was proxied by the country of birth of the respondent's parents: we only kept those subjects with both parents born in the country of interview.

There is a substantial difference in the language variable treatment between SHARE and WVS. While in WVS the individuals are asked to declare the language they normally speak at home, in SHARE those individuals living in countries with two or more official languages are given the possibility to choose whether to compile the questionnaire in one language or another. We assume that the language in which the questionnaire is compiled is also an individual's primary language. As for risk attitudes in SHARE, we consider only individuals who are responsible for financial matters in the household (heads of household). They were asked to answer a simple risk tolerance question:

> *When people invest their savings they can choose between assets that give low return with little risk to lose money, for instance a bank account or a safe bond, or assets with a high return but also a higher risk of losing, for instance stocks and shares. Which of the statements on the card comes closest to the amount of financial risk that you are willing to take when you save or make investments?*
>
> (1) Take substantial financial risk expecting to earn substantial returns
>
> (2) Take above average financial risks expecting to earn above average returns
>
> (3) Take average financial risk expecting to earn average returns
>
> (4) Not willing to take any financial risk.

Individuals who answered (1) and (2) are considered as risk lovers. The intermediate risk takers are those who answered (3) while all the individuals in (4) are considered as risk averse. In our sample, 75.51% of individual declare to be risk averse, 20.52% of individuals are ready to take average financial risks, and only 3.98% of individuals are willing to take above average or substantial financial risk.

Tables 3.1 and 3.2 show the distribution of individual attitudes toward risk for each value of IRR. Table 3.1 considers the entire range of IRR (from 0 to 6), while in Table 3.2 we classify IRR in 3 different categories: category 0 contains no IRR, category 1 refers to intermediate IRR usage (2 or 3 IRR) and category 2 represents strong and very strong IRR usage (4 or 6 IRR)[27].

Table 3.1: Risk Aversion by Irrealis (%)

	IRR Linguistic Marker					
	IRR=0	IRR=2	IRR=3	IRR=4	IRR=6	Total
Risk Lovers	38.40	23.57	16.22	15.69	6.13	100.00
Risk Neutral	23.02	32.04	20.47	18.09	6.39	100.00
Risk Averse	8.77	30.70	27.74	22.01	10.79	100.00
Total (%)	12.87	30.69	25.79	20.95	9.70	100.00
Observations	9,862	23,521	19,765	16,060	7,434	76,642

Source: Elaboration of SHARE data, Waves 2, 4 and 5.

Nearly 40% of risk prone individuals also speak a moodless language. More than half of the individuals with an intermediate level of risk aversion also classify as intermediate IRR users. Finally, those who declare to be highly risk averse are mostly either intermediate or high IRR users.

Figure 3.1 shows the distribution of individuals by IRR for the entire set of countries (left-hand side figure) and for a restricted set of linguistically heterogeneous countries

[27] None of the languages in our dataset has IRR=1 or IRR=5. However, generally these values are admissible.

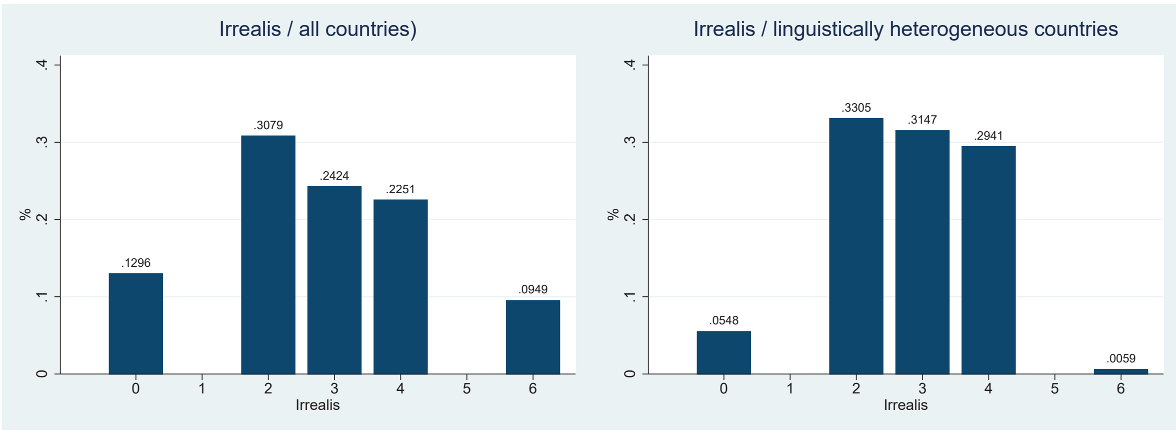
Table 3.2: Risk Aversion by Irrealis (categorized) (%)

	Categorized IRR			
	CatIRR0	CatIRR1	CatIRR2	Total
Risk Lovers	38.40	39.78	21.82	100.00
Risk Neutral	23.02	52.51	24.48	100.00
Risk Averse	8.77	58.43	32.80	100.00
Total (%)	13.65	55.97	30.38	100.00
Observations	9,862	43,286	23,494	76,642

Source: Elaboration of SHARE data, Waves 2, 4 and 5.

(right-hand side figure).

Figure 3.1: Distribution of Irrealis users in all countries and in linguistically heterogeneous countries.



Source: SHARE, Wave 2, 4 and 5. N. Observations: 89 165 (all countries); 21 182 (linguistically heterogeneous countries).

Roughly 13% of all respondents are moodless speakers, slightly more than 50% are intermediate IRR users, while over 30% are intensive and very intensive IRR users. Around two thirds of individuals in linguistically heterogeneous countries are intermediate Irrealis users, whereas 5% do not use IRR at all. Around 29% are intensive and very intensive IRR users.

Differently from SHARE, the question on risk aversion of the WVS refers to adventure risk taking.[28] The risk preference questions in SHARE and WVS, hence, refer to two different types of risk attitudes.[29] As a consequence, the distribution of respondents in WVS differ from the one in SHARE with roughly 17% of individuals being highly averse to any adventure and risk taking and 21% of respondents that classify as risk lovers. In addition, the information on stock holdings is available only in SHARE. Since the features of risk preferences elicited in SHARE fit particularly well with the nature of our research question, and WVS does not contain any information on individuals' asset holdings, we use SHARE as a primary source of data.

3.3.4 Irrealis and Risk Aversion: Empirical Strategy

Our first set of regressions examines the relationship between individual attitudes toward risk and the IRR linguistic marker associated to the individual's primary language (*Hypothesis 1*). The dependent variable RA_i is equal to 1 for an individual declaring to be averse to taking risks and 0 otherwise. The empirical problem consists of estimating the following probit model:

$$P(RA_i) = \Phi(r_i) \equiv \int_{-\infty}^{r_i} \phi(r_i) dr_i$$

where $\phi(r_i))$ is the standard normal density

$$\phi(r_i) = (2\pi)^{-1/2} exp(-r_i^2/2)$$

and

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i. \tag{3.3}$$

[28] Moreover, the information on risk attitude in WVS includes all the respondents and not only the heads of household.

[29] The risk preference question from WVS is reported in Appendix 3.C.

Our main variable of interest IRR_i denotes the number of non-indicative moods in *Irrealis* contexts in the individual $i's$ primary language. X_i is the vector of demographic and socio-economic characteristics of individual i, such as gender, marital status, family size, occupation, education and household's income level. Z_i contains controls for cognitive ability and literacy, level of trust and health conditions. Finally, CW_i is a country-wave fixed effect.

In addition to the probit analysis, we estimate our model with a set of fixed-effect controls for individual demographic and socio-economic characteristics. In such a way we compare individuals identical on these dimensions, but who differ in their IRR usage. These regressions are estimated using a fixed-effect (conditional) logistic model:

$$P(RA_i) = \frac{exp(r_i)}{1 + exp(r_i)}$$

where:

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \lambda FE_i + \eta_i. \tag{3.4}$$

FE_i is the set of individual specific fixed-effects, such as gender, education, age, income, marital status, and number of children[30].

Even though the literature on linguistic differences and individual attitudes toward risk is very scarce, there is a large body of evidence on the relationship between attitudes toward risk and several socio-demographic and behavioral characteristics, such as educational attainment, income and wealth endowments, occupational status, age, family size, cognitive and health conditions, and trust. Bellante and Green (2004), Dohmen et al.

[30]The STATA® command *clogit* was used for the empirical execution of the model. Fixed effects are obtained by applying a matched group mechanism.

(2011), Lin (2009) and Riley Jr. and Chow (1992) for instance show that a higher level of education increases risk tolerance. In addition, Dohmen et al. (2011) consider the family background and find that the fathers' education is negatively correlated with the individual's risk aversion and risky behavior. The relationship between risk aversion and income is generally negative.[31] However, Barsky et al. (1997) and Hartog et al. (2002) show that this relationship is not linear. Similarly, Guiso and Paiella (2008), Riley Jr. and Chow (1992), and Dohmen et al. (2011) find that the level of risk aversion decreases in wealth. Dohmen et al. (2011) and Cohen and Einav (2007) draw attention to a potential problem of endogeneity as a greater willingness to take risks could *ex ante* lead to higher levels of wealth.

Regarding the occupational status, self-employment correlates significantly with the level of individual risk attitude with entrepreneurs being significantly more risk tolerant than others (Hartog et al. (2002), Siegel and Hoban (1991)). On the other hand, Dohmen et al. (2011) and Hartog et al. (2002) show that an individual's unemployment status does not seem to be relevant to their attitude toward risk. Moreover, a higher level of risk aversion for married individuals is found in Cohen and Einav (2007) and Halek and Eisenhauer (2001), whereas Bellante and Green (2004) and Hartog et al. (2002) do not observe any significant effects in that sense.[32] The number of children is found to increase the level of risk aversion (Dohmen et al. (2011) and Lin (2009)), but family size has a negative effect making individuals more risk tolerant (Siegel and Hoban (1991) and Lin (2009)). Finally, women are found to be more risk averse than men (Barsky et al. (1997), Dohmen et al. (2011), Halek and Eisenhauer (2001), Hartog et al. (2002), Cohen and Einav (2007) and Jianakoplos and Bernasek (1998)). Other factors influencing individual risk attitudes include a poor health status and cognitive decline which are shown to play a

[31] Pratt (1964) argues in favor of decreasing absolute risk aversion, that is, as wealth and income increases, individuals should become more risk tolerant in absolute terms.

[32] One should consider, as pointed out by Halek and Eisenhauer (2001), that more risk averse individuals could also have an ex ante preference for marriage instead of the opposite flow of causality.

role in Bellante and Green (2004) and Bonsang and Dohmen (2015) respectively, whereas Hartog et al. (2002) does not find any significant correlation between poor health and risk tolerance.

Results

The empirical estimation of equation 3.3 is presented in Tables 3.E.1, 3.E.2, 3.E.3 and 3.E.6. In all regression models we calculate the robust standard errors clustered by country. Models 1-5 in Table 3.E.1 consider IRR as a limited discrete variable. Model 6 includes the categorized version of IRR. The coefficients associated to IRR are highly significant in all model specifications. The coefficient on CatIRR2 in Model 6 indicates that strong and very strong IRR speakers are on average 9% more risk averse with respect to speakers with no indicative moods in Irrealis contexts. The results confirm our initial intuition (*Hypothesis 1*: there is a strong association between IRR and the individual attitudes toward risk. Females are on average more risk averse than men, while higher levels of education are generally associated with lower risk aversion. In line with the existing literature we also find that individuals with higher income are on average less risk averse than poorer ones. Regarding the occupational status, employment correlates significantly with the level of individual risk attitude with unemployed individuals being significantly less risk tolerant than others. Also, there is evidence for married individuals being more risk averse. Restricting our sample to linguistically heterogeneous countries (Table 3.E.2) does not significantly alter the results: a more intensive use of IRR is associated with higher aversion to risk. Given that for linguistically heterogeneous countries, we only have Israel in the category of CatIRR0, we include CatIRR2 (IRR=4 and IRR=6) as a dummy variable for the categorical robustness test. Moreover, In order to control for possible interdependencies between languages belonging to the same linguistic family and/or subfamily, in Table 3.E.3 we control for linguistic families and run separate regressions for two main linguistic sub-families, namely Romance and Germanic. Coefficients in Models 1 and 2 indicate a strong and significant association between IRR and individual risk aversion even after controlling for main linguistic families (Indo-European, Semitic, and Uralic). The effect of IRR remains strong and highly significant even within each linguistic sub-family (Models 3-5).

In Tables 3.E.4 and 3.E.5 we estimate a conditional logit model with the set of additional fixed-effect controls (Equation (3.4)). All coefficients are reported as odds-ratios. The results show that even when comparing individuals that are identical on every dimension, speaking a higher-intensive IRR language is associated with significantly higher probability of being risk averse. For instance, the coefficient on IRR in Model 5 shows that among individuals of the same age, gender, income, educational attainment, marital status and number of children, a marginal increase in IRR translates into 16% higher probability of being risk averse. These effects do not change significantly when we restrict our sample to linguistically heterogeneous countries.

Finally, Table 3.E.6 considers individuals from the World Value Survey. Since the distribution of IRR in WVS differs from that in SHARE, we consider a binary coded IRR variable that equals 1 whenever IRR is greater or equal to 4 (66% of observations) and 0 otherwise. As before, risk aversion is a binary coded individual self-declared risk aversion variable equal to 1 if the individual is highly risk averse and 0 otherwise. In line with the results from SHARE, the association between IRR and risk aversion is positive and statistically significant. All the other coefficients have the expected sign, except for unemployment which is not significant. The association between IRR and individual risk aversion seems to be very robust.

3.3.5 Irrealis, Risk Aversion and Stock Ownership: Empirical strategy

Many economic decisions involve outcomes that are uncertain, delayed or both. As a consequence, individual risk attitudes and their preferences regarding current versus future consumption are of fundamental importance to economic modeling. The existing literature suggests an intuitive inverse relationship between risk aversion and risky asset accumulation with more risk averse individuals being less willing to invest for example in stocks, as addressed by the model we presented in section 3.3.

However, following the common assumption that the level of absolute risk aversion is sensitive to changes in wealth translates into different levels of risk aversion for different levels of initial wealth. Also, when considering the analysis from a dynamic viewpoint, individual returns from risky investments at time t add on to the level of personal wealth which again will impact negatively on the individual level of risk aversion at time $(t+1)$. A thorough analysis of the effect of risk aversion on investment in risky assets needs therefore to consider potential problems associated with the endogeneity of risk aversion.

Individual time preferences, on the other hand, represent the attitude toward present and future consumption and are as such a fundamental driver of intertemporal decision-making. In the case of investment in risky assets, future returns are however uncertain by definition. Disentangling these two determinants of economic choices, such as the investment in risky assets and savings behavior, is not a trivial task and has been taken on by numerous authors since the development of the theoretical frameworks and empirical assessments by Epstein and Zin (1989). One of the novelties of their model consisted in the development of a model flexible enough to allow for the separation between the intertemporal preferences ρ and the attitudes toward risk α. However, even after more than 30 years, the intertwined relationship between these two parameters remains an open question, as also outlined by Andreoni and Sprenger (2015). In two papers of 2012

(Andreoni and Sprenger (2012a) and Andreoni and Sprenger (2012b)) the same authors confirm that risk aversion and intertemporal preferences cannot be considered as perfect substitutes but can also not be completely separated. In a similar fashion, Andersen et al. (2008) explain that the assumption of risk neutrality for individuals that are instead risk averse, result in upward-biased discount rate estimates. They also point out that the parameter values for risk and time preferences must come from the same population.

In the next section we are going to analyze the impact of risk aversion on stock ownership by using the level of *Irrealis* use as an instrument in the attempt to overcome the above mentioned problems of endogeneity. Furthermore, we were interested in assessing the separate effects of time preferences and risk aversion. As suggested by Chen (2013), we use the FTR parameter as a marker for intertemporal choice preferences. Both, FTR and *Irrealis* have the advantage that they are exogenous and orthogonal, hence their empirical coexistence, the first as a marker and the second as an instrument for risk aversion, could provide insightful results regarding the separate effects of risk aversion and time preferences.

The empirical estimation of the causal relationship between risk aversion and asset accumulation may suffer both from reverse causality and an omitted variables problem. Since the returns to financial assets represent a certain form of income, and income and risk aversion are negatively correlated, assets accumulation and risk aversion may be simultaneously determined. Moreover, there are several unobservables excluded from the model that jointly determine the individual attitudes toward risk, making the self-declared risk aversion variable correlated with the error term. In order to make accurate predictions, we need reliable instruments for measuring individual risk preferences.

According to our hypotheses, linguistic differences directly influence the individual perception of risk and uncertainty (*Hypothesis 1*) and indirectly their investment decisions (*Hypothesis 2*). In other words, language (IRR) affects investment decisions through its direct impact on risk aversion. In light of the empirical evidence in Section 3.3.4 which strongly supports *Hypothesis 1*, the IRR linguistic marker may represent a suitable instrument for individual risk preferences in the reduced form equation for the propensity of holding risky assets.

The empirical problem consists of estimating the following causal relationship:

$$AS_i = \alpha + \beta RA_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i \tag{3.5}$$

where RA_i denotes individual i's risk aversion, X_i is the vector of demographic and socio-economic individual characteristics, such as gender, marital status, family size, occupation, education and household's income level, and Z_i contains controls for cognitive ability and literacy, level of trust and health conditions. Finally, CW_i denotes country-wave fixed effects.

In the first stage we estimate the effects of socio-economic characteristics, linguistic variation and location on individual self-declared risk aversion:

$$RA_i = \alpha + \pi_{i1} IRR_i + \pi_{i2} X_i + \pi_{i3} Z_i + \pi_{i4} CW_i + \zeta_i \tag{3.6}$$

where IRR_i denotes the number of non-indicative moods in Irrealis contexts in the individual $i's$ language. By plugging the first stage fitted values in the second stage equation we obtain the reduced form model for asset accumulation:

$$AS_i = \alpha + \beta \widehat{RA_i} + \gamma X_i + \theta Z_i + \rho CW_i + error_i \tag{3.7}$$

Economic theory suggests an inverse causal relationship between risk aversion and asset accumulation: everything else equal, higher risk aversion reduces the individual propensity to invest in risky assets. If the prediction of the theory is correct, the empirical validation of equation 3.7 should yield a negative coefficient for $\widehat{RA_i}$.

Results

The empirical estimation of Equation 3.6 is presented in Table 3.E.7. Only linguistically heterogeneous countries are considered. In that way we can compare individuals living in the same (or a very similar) environment, but speaking different languages[33]. For the two-stage empirical model in Equations 3.6 and 3.7 to work, the IRR linguistic marker must satisfy three basic requirements: a) it must be correlated with the endogenous variable (instrument relevance), b) uncorrelated with the error term (independence), and c) it should not have any direct impact on the probability of holding assets other than through its first stage impact on risk aversion (exclusion restriction). The first stage test statistics in Table 3.E.7 confirm the strength of our instrument. In all model specifications, the F-statistic is significantly higher than a commonly used threshold (10 or 16). Our instrument, hence, is highly correlated with the endogenous variable even after controlling for the effect of other regressors, and for the FTR linguistic marker in Models 5 and 6. Even though the exogeneity of the instrument cannot be directly tested, there is no reason to suspect that there is any reverse effect of the propensity to invest in risky assets on the instrument. Since we control for country fixed effects (which capture institutional and other country-specific heterogeneities), trust, education, income, occupational status and health conditions (which may be influenced to some extent by linguistic variation), the exclusion restriction should not be violated. In other words, we can rule out any direct

[33] As a robustness check we also run our regressions on the full set of countries. Estimation results are available upon request.

effects of linguistic variation on the propensity to invest in risky assets through omitted variables.

Table 3.E.8 shows the second stage estimates of Equation 3.7 from a recursive bivariate probit model (Model 1). The dependent variable (asset accumulation) equals 1 whenever individuals hold some money in stocks or shares (listed or unlisted on the stock market), and 0 otherwise. The information on individual asset holding comes from the Survey of Health, Aging and Retirement in Europe (SHARE).[34] Only marginal effects are reported. In all regression models we control for country and wave fixed effects, cognitive abilities and individual health conditions. The coefficients reported in the table were estimated with the non-categorized version of the instrument. To obtain a direct effect of individual time preferences on asset accumulation, we run separate regressions using the FTR linguistic marker (Chen (2013)) as a proxy for the individual subjective discount rate (Model 2). In order to estimate the separate effects of risk aversion and time preferences on the propensity to invest in risky assets, we reestimate a recursive bivariate probit model using the FTR parameter as a marker for intertemporal choice preferences and the IRR marker as an instrument for risk aversion (Model 3).

The instrumented risk aversion is highly significant. Without controlling for time preferences, for an individual with average characteristics of the population, being highly risk averse reduces the probability of holding risky assets by approximately 10%. All the other coefficients have the expected sign. Income, Education and Age are positively associated with assets accumulation. Married couples invest more, whereas having more children is negatively related to asset holding. Also, even after controlling for risk aversion, females invest less in risky assets. Trust has a positive impact in investment in risky assets, in line with the findings of This result is consistent with the results of Guiso et al. (2008) who find that a general lack of trust limits stock market participation. Also

[34]The WVS dataset does not contain any information on the respondents' asset holdings.

Baumgartner et al. (2008) find that, by administering the trust-enhancing neuropeptide oxitocyn, subjects invest more of their experimental money endowment. Finally, the estimated coefficients in the joint model show that individuals with a high subjective discount rate invest 5% less in risky assets with respect to low discounting individuals. The separate effects of risk aversion and time preferences are shown in Model 3. The effect of the individual discount rate is negative and highly significant while the coefficient on risk aversion reduces by 3 percentage points. The effect of risk aversion in our reduced form model is almost two times larger than the effect of the individual discount rate.

3.4 Conclusions

This paper proposes an innovative approach to analyzing individual attitudes toward uncertainty and asset accumulation based on the *Sapir-Whorf* hypothesis of linguistic relativity. We develop a specific linguistic marker defined on the basis of the number of non-indicative moods used in *Irrealis* contexts, which refer to situations involving events that have yet to occur (nonfactual or non-actual). The two most notable examples of non-indicative moods used to describe hypothetical situations are the subjunctive and the conditional mood. Our empirical exercise consists in testing the hypothesis that speakers of languages in which non-indicative moods are used more frequently perceive the world as being more mutable and uncertain with respect to speakers of languages where these forms are less common, or do not exist at all. By resorting to verb forms that attach a higher semantic salience to situations of uncertainty, we hypothesize that the speakers' *Weltanschauung* differs and ultimately translates into a higher level of risk aversion.

The association between our linguistic markers and risk aversion is robust to different model specifications. Individuals speaking languages where non-indicative moods are used more intensively have a marginal increase in the probability of being risk averse of

16%. Even when we compare individuals that are identical on every other dimension, such as gender, education, age, income, marital status, and number of children, a more intensive use of non-indicative moods is associated with significantly higher levels of risk aversion. Moreover, these effects are robust to the restriction of our sample to linguistically heterogeneous countries. The results indicate that there is a significant variation in risk attitudes both, across individuals living in the same country and speaking a different language, and across countries.

Using our linguistic marker as an instrument for the individuals' risk aversion we show that being highly risk averse reduces the probability of holding risky financial assets by 10%. In addition to risk preferences, we run separate regressions with the FTR linguistic marker (Chen (2013)) as a proxy for the individual subjective discount rate. We find that both measures are relevant determinants in the decision of investing in stocks. In line with our hypotheses, the level of risk aversion and the preference for current consumption have a separate and significant negative impact on risky asset holdings when analyzed jointly. This also conforms with the findings of Andreoni and Sprenger (2012a) and Andreoni and Sprenger (2012b) supporting the conjecture that risk and time preference cannot be considered as substitutes. In absolute terms, the impact of risk aversion on risky asset accumulation is higher than the subjective discount rate. The approach adopted in this paper is, to the best of our knowledge, the first non-experimental attempt to measure a direct and unbiased effect of risk aversion and individual time preferences on investment in risky financial assets. Since linguistic variation is seen as a trait of individual identity, and can hence be exploited as a source of identity and as a cultural marker, not only at the individual but also at the group level, the results obtained in this paper also shed light on the importance of non-economic factors in shaping individual risk and time preferences, and consequently their economic behavior.

Appendix

3.A Linguistic Mapping

Table 3.A.1: Linguistic Mapping

Language	Family	Sub-Family	#Moods	a	b	c	d	e	f	g	$\sum$ Non Ind.
Albanian	Indo-Euro	—	>2	1	1	0	0	0	0	1	3
Arabic	Semitic	—	2	1	1	1	1	0	0	0	4
Basque	Isolate	—	2	1	1	0	0	0	0	1	3
Belorussian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Bulgarian	Indo-Euro	Slavic	1	1	1	0	0	0	0	0	2
Catalan	Indo-Euro	Romance	1	1	1	0	0	0	1	0	3
Croatian	Indo-Euro	Slavic	1	0	0	0	0	0	1	1	2
Czech	Indo-Euro	Slavic	0	1	1	0	0	0	1	1	4
Danish	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Dutch	Indo-Euro	Germanic	0	0	0	0	0	0	1	1	2
English	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Estonian	Indo-Euro	Finno-Ugric	2	0	1	0	0	0	1	1	3
Finnish	Uralic	Finno-Ugric	2	0	0	0	0	0	1	1	2
French	Indo-Euro	Romance	1	1	1	0	1	0	0	0	3
German	Indo-Euro	Germanic	1	0	0	0	0	0	1	1	2
Greek	Indo-Euro	—	1	0	1	0	0	0	0	1	2
Hebrew	Semitic	—	0	0	0	0	0	0	0	0	0
Hungarian	Uralic	Finno-Ugric	2	1	1	0	0	0	1	1	4
Icelandic	Indo-Euro	Germanic	1	1	1	1	0	1	1	1	6
Irish	Indo-Euro	Celtic	2	1	1	0	0	0	1	1	4
Italian	Indo-Euro	Romance	2	1	1	1	1	0	1	1	6
Latvian	Indo-Euro	Baltic	1	1	1	0	0	0	1	1	4
Lithuanian	Indo-Euro	Baltic	1	1	1	0	0	0	1	1	4
Macedonian	Indo-Euro	Slavic	1	1	1	0	0	0	0	0	2
Maltese	Semitic	—	0	0	0	0	0	0	0	0	0
Norwegian	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Polish	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Portuguese	Indo-Euro	Romance	2	1	1	1	1	0	1	1	6
Romanian	Indo-Euro	Romance	1	1	1	0	0	0	1	1	4
Russian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Serbian	Indo-Euro	Slavic	1	0	0	0	0	0	1	1	2
Slovak	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Slovenian	Indo-Euro	Slavic	1	0	1	0	0	0	1	1	3

Language	Family	Sub-Family	# Mood	a	b	c	d	e	f	g	∑ Non Ind.
Spanish	Indo-Euro	Romance	1	1	1	0	1	0	1	0	4
Swedish	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Turkish	Ural-Altaic	Turkic	>2	1	1	1	0	0	1	0	4
Ukrainian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Welsh	Indo-Euro	Celtic	1	1	0	0	0	0	1	1	3

Notes: *Contexts: a = Modal; b = Desire; c = Attitude (non factive); d = Attitude (factive); e = Declarative; f = Protasis (counterfactual conditional); g = Apodosis (counterfactual conditional).*

List of languages surveyed in Rothstein and Thieroff's (2010): 36 Languages

1. seven Germanic languages (Icelandic, Norwegian, Swedish, Danish, English, Dutch and German);
2. six Romance languages (French, Portuguese, Spanish, Catalan, Italian, Rumanian);
3. three Celtic languages (Irish, Breton, Welsh);
4. ten Slavic languages (Russian, Polish, Czech, Slovak, Sorbian, Bosnian, Croatian, Serbian, Bulgarian and Macedonian);
5. two Baltic languages (Latvian and Lithuanian)
6. three other Indo-European languages (Greek, Albanian and Armenian);
7. three Finno-Ugric languages (Finnish, Estonian and Hungarian);
8. four other non-Indo-European languages (Turkish, Maltese, Georgian and Basque).

Regarding the number of finite moods in the languages of Europe, Thieroff (2000) outlines some typological generalizations. First, all the languages have a distinct imperative mood while only one language, Maltese, does not have any non-indicative non-imperative mood. Seven languages (Norwegian, Swedish, Danish, Dutch, English, Irish and Welsh), have one non-indicative non-imperative mood, the subjunctive, but are in the process of losing it. In these languages subjunctive has a very limited use, often restricted to formulaic, almost unproductive forms. As a consequence, the use of indicative in these languages has spread in semantic domains where in previous stages the subjunctive was used. We consider those languages as "moodless". Most languages spoken in Europe have one non-indicative non-imperative mood, the subjunctive or the conditional. This group includes languages such as Breton, Bosnian, Bulgarian, Catalan, Croatian, Czech, French, Georgian, German, Greek, Icelandic, Italian, Latvian, Lithuanian, Macedonian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Sorbian, Spanish. Belarusian, Slovenian, and Ukrainian are not surveyed in RT, but can also be added to this group of languages. Notice that Thieroff follows the view that Romance conditional mood is not an independent mood (Iatridou (2000), Laca (2010)) and classifies it as part of the indicative paradigm. However, opinions vary on this point. Giorgi (2009), for instance, claims that the conditional is a mood of its own. If we assume conditional mood is an independent mood, Romance languages may be classified as two-mood languages. Four languages in RT have two non-indicative non-imperative moods: Armenian (subjunctive and debitive), Estonian (conditional and jussive), Finnish (conditional and potential) and Hungarian (subjunctive and conditional). Finally, two languages are traditionally considered as having more than two non-indicative non-imperative moods: Albanian (Breu, 2010) and Turkish (Optative, Irrealis, Necessitative, Conditional and Possibilitative).

Semantic import of subjunctive and the conditional moods: Examples

Italian subjunctive, for instance, is inflectional, as it adds a bound morpheme to the verb:

(1) *io vengo* — *(che) io venga*
I come-IND.1SG — (that) I come - SUBJ.1SG

German "Konjunktiv II", on the other hand, is an example of periphrastic mood (at least in weak verbs) formed by an auxiliary and the infinitive:

(2) ich würde kaufen
I **AUX** buy
"I would buy"

Conditional mood in Greek is periphrastic as well, but it resorts to an invariable modal particle (*tha*) followed by the verb in the imperfect or the pluperfect tense (Holton et al. 2004: 123, 151).

(3) tha éhana
PRT lose-IMPF.1SG
"I would lose".

For instance, in counter-factual conditional sentences, (e.g. "If it had not rained, I would have gone for a walk") German has the Konjunktiv II in both the if - clause and in the main clause.

(4) Wenn es nicht geregnet hätte, ginge ich spazieren.
If it not rained AUX.KONJ, go.KONJ I walking
"If it had not rained, I would have gone for a walk."

Italian has the subjunctive in the if-clause only, while the conditional must be used in the main clause.

(5) Se non avesse piovuto, avrei fatto una passeggiata.
If not AUX.SUBJ rained, AUX.KONJ made a walk
"If it had not rained, I would have gone for a walk."

French has the imperfect in the if-clause and the conditional in the main clause.

(6) Si il n'avait pas plu, j'aurais fait une promenade.
If it not AUX.IMP N. rained, I AUX.COND made a walk
"If it had not rained, I would have gone for a walk."

In argument clauses of desire verbs, for instance, some languages (like Italian, see (10)) use the subjunctive, while others (like German, see (11)) use the indicative.

(7) Spero che tu stia bene.
Hope-1SG that you be-SUBJ well.
"I hope you are fine".

(8) Ich hoffe dass es dir gut geht.
Hope-1SG that you be-SUBJ well.
"I hope you are fine".

In argument clauses of declarative verbs, some languages use obligatorily the indicative, while others can optionally use the subjunctive.

(9) Mi ha detto che sta bene.
Me has told that stays well
"He told me he's fine."

(10) Er sagte mir, es gehe ihm gut.
He told me, it goes him well
"He told me he's fine."

3.B Questionnaire

Questionnaire language mood

Q1 Thank you for your cooperation. We really appreciate you helping us with our survey! We are trying to discover what verb moods are used in the languages of Europe and how they are used.

Q2 First, would you kindly provide us with your contact details in order for us to eventually be able to follow up with you?

Q3 Contact Details
Name:
Surname:
(Academic title:)
Email:
City:
Country:

Q4 Personal Details (confidential - for statistical use only)
Occupation/Position:
Sector/Affiliation:
Mother tongue:
Please describe the level and nature of your linguistic training:

Q6 The questions start on the next screen. We only ask you to translate the sentences and tell us which mood you would use in your language to translate the verb in bold: is the verb in bold in the indicative or in a different mood than indicative (e.g. subjunctive, conditional, hypothetical, optative, etc.)?

Q7 Leo says that Bill has won the race. Translation:

Q8 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q9 Bill has won the race. Translation:

Q10 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q11 It is likely that Bill will win the race. Translation:

Q12 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q13 Leo regrets that Bill won't win the race. Translation:

Q14 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q15 If Bill won(1) the race, Leo would be(2) happy. Translation:

Q16 Mood (1):
❍ Indicative
❍ Other mood (please describe): ____________________

Q17 Mood (2):
❍ Indicative
❍ Other mood (please describe): ____________________

Q18 Schools teach a foreign language. Translation:

Q19 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q20 Leo thinks that Bill has won the race. Translation:

Q21 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q22 Leo regrets that Bill hasn't won the race. Translation:

Q23 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q24 Leo wants Bill to win the race. Translation:

Q25 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q26 Leo hopes that Bill will win the race. Translation:

Q27 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q28 Leo says that Bill will win the race. Translation:

Q29 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q30 It is compulsory that schools teach a foreign language. Translation:

Q31 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q32 Leo hopes that Bill has won the race. Translation:

Q33 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q34 It is possible that Bill has won the race. Translation:

Q35 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q36 Leo thinks that Bill will win the race. Translation:

Q37 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q38 Bill will win the race. Translation:

Q39 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q40 It is necessary that schools teach a foreign language. Translation:

Q41 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q42 It is possible that Bill will win the race. Translation:

Q43 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q44 It is likely that Bill has won the race. Translation:

Q45 Mood:
❍ Indicative
❍ Other mood (please describe): ____________________

Q46 If Bill had won(1) the race, Leo would have been(2) happy. Translation:

Q47 Mood (1):
❍ Indicative
❍ Other mood (please describe): ____________________

Q48 Mood (2):
❍ Indicative
❍ Other mood (please describe): ____________________

3.C Additional Information: WVS

List of countries - WVS: Andorra, Argentina, Australia, Bahrain, Brazil, Bulgaria, Belarus, Canada, Chile, Colombia, Cyprus, Ecuador, Estonia, Finland, France, Germany, Great Britain, Guatemala, Hungary, Iraq, Italy, Kazakhstan, Jordan, Kuwait, Kyrgyzstan, Lebanon, Libya, Mexico, Moldova, Morocco, Netherlands, New Zealand, Nigeria, Norway, Palestine, Peru, Poland, Qatar, Romania, Russia, Singapore, South Africa, Spain, Sweden, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Egypt, United States, Burkina Faso, Uruguay, Yemen, (former) Serbia and Montenegro.

Counties with with less than 100 observations excluded from the analysis (Algeria, Azerbaijan, Armenia, Ethiopia, Georgia, Ghana, Hong-Kong, India, Iran, Malaysia, Mali, Rwanda, Slovenia, Zimbabwe, Uzbekistan and Zambia).

List of languages - WVS: Arabic, Bulgarian, Catalan, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Swedish, Turkish, Ukrainian.

Languages with less than 100 speakers excluded from the analysis (Albanian, Basque, Belorussian, Croatian, Macedonian, Maltese and Slovak).

Risk preference question from WVS:

> *Now I will briefly describe some people. (...) would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you?*
>
> Description: **Adventure and taking risks are important to this person; to have an exciting life.**
>
> (1) Very much like me;
>
> (2) Like me;
>
> (3) Somewhat like me;
>
> (4) A little like me;
>
> (5) Not like me;
>
> (6) Not at all like me.

3.D Summary Statistics

Table 3.D.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Risk aversion	3.705	0.572	1	4	76642
Strong_RA	0.755	0.43	0	1	76642
No Irrealis Moods	0.13	0.336	0	1	89165
2 Irrealis Moods	0.308	0.462	0	1	89165
3 Irrealis Moods	0.242	0.429	0	1	89165
4 Irrealis Moods	0.225	0.418	0	1	89165
6 Irrealis Moods	0.095	0.293	0	1	89165
No Irrealis Moods	0.13	0.336	0	1	89165
2 or 3 Irrealis Moods	0.55	0.497	0	1	89165
4 or 6 Irrealis Moods	0.32	0.466	0	1	89165
Strong_IRR	0.562	0.496	0	1	89165
Income	4.533	2.866	0	9	89165
Owner	0.749	0.434	0	1	89165
EduCat	0.772	0.763	0	2	89165
Trust People	5.748	2.402	0	10	80862
Married	0.714	0.452	0	1	89165
HH Size	2.163	1.376	0	17	89165
AgeCat	61.774	10.78	40	90	89151
Female	0.555	0.497	0	1	89165
Retired	0.556	0.497	0	1	89165
Employed	0.268	0.443	0	1	89165
Unemployed	0.028	0.165	0	1	89165
Disabled	0.036	0.186	0	1	89165
Homemaker	0.09	0.286	0	1	89165
adl	0.296	0.979	0	6	88724
iadl	0.472	1.282	0	7	88724
reading	3.662	1.133	1	5	89165
writing	3.535	1.171	1	5	89165
Number of chronic deseases	1.751	1.58	0	14	89165

3.E Regression Tables

Table 3.E.1: *Probit Model (SHARE): Risk Aversion, Marginal Effects*

Risk Aversion (d)	RA 1	RA 2	RA 3	RA 4	RA 5	RA 6
IRR	0.034***	0.025***	0.022**	0.021***	0.018**	
	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)	
Cat=1						0.054***
						(0.013)
Cat=2						0.092***
						(0.006)
Age	0.007***	0.004***	0.004***	0.004***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.109***	0.088***	0.092***	0.090***	0.092***	0.092***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Low Edu.		0.075***	0.071***	0.069***	0.062***	0.062***
		(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
High Edu.		-0.098***	-0.091***	-0.089***	-0.080***	-0.079***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Income		-0.013***	-0.014***	-0.013***	-0.012***	-0.012***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Owner		-0.037***	-0.037***	-0.035***	-0.033***	-0.033***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Married			0.016**	0.014**	0.015**	0.015**
			(0.007)	(0.007)	(0.007)	(0.007)
Num. Children			0.002	0.002	0.002	0.002
			(0.002)	(0.003)	(0.003)	(0.003)
Trust People			-0.013***	-0.012***	-0.012***	-0.012***
			(0.001)	(0.001)	(0.001)	(0.001)
Retired				0.035***	0.034***	0.034***
				(0.006)	(0.006)	(0.006)
Unemployed				0.037***	0.034***	0.034***
				(0.008)	(0.008)	(0.008)
Disabled				0.068***	0.051***	0.051***
				(0.010)	(0.009)	(0.009)
Homemaker				0.031***	0.027***	0.026***
				(0.008)	(0.008)	(0.008)
Country, Wave dummy	Yes	Yes	Yes	Yes	Yes	Yes
Cognitive, Health	No	No	No	No	Yes	Yes
N. Observations	76633	76633	75842	74446	74426	74426
N. Countries	17	19	19	19	19	19

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.2: *Probit Model (SHARE): Risk Aversion, Marginal Effects. Linguistically Heterogeneous Countries*

Risk Aversion (d)	RA 1	RA 2	RA 3	RA 4	RA 5	RA 6
IRR	0.032***	0.024***	0.023***	0.023***	0.021***	
	(0.011)	(0.006)	(0.008)	(0.007)	(0.007)	
Cat=2						0.061***
						(0.022)
Age	0.005***	0.003***	0.003***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Female	0.100***	0.081***	0.083***	0.085***	0.086***	0.087***
	(0.003)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)
Low Edu.		0.075***	0.075***	0.075***	0.066***	0.066***
		(0.011)	(0.010)	(0.009)	(0.010)	(0.010)
High Edu.		-0.101***	-0.095***	-0.093***	-0.081***	-0.077***
		(0.015)	(0.015)	(0.014)	(0.014)	(0.015)
Income		-0.011***	-0.011***	-0.010***	-0.010***	-0.010***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Owner		-0.062***	-0.062***	-0.058***	-0.057***	-0.058***
		(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Married			0.013	0.011	0.012	0.012
			(0.008)	(0.009)	(0.009)	(0.010)
Num. Children			-0.006*	-0.006**	-0.007**	-0.007**
			(0.003)	(0.003)	(0.003)	(0.003)
Trust People			-0.008***	-0.008***	-0.008***	-0.008***
			(0.001)	(0.001)	(0.001)	(0.001)
Retired				0.028***	0.028***	0.028***
				(0.011)	(0.011)	(0.010)
Unemployed				0.032**	0.031*	0.031**
				(0.016)	(0.016)	(0.015)
Disabled				0.060***	0.048***	0.047***
				(0.014)	(0.010)	(0.010)
Homemaker				0.016	0.013	0.010
				(0.011)	(0.012)	(0.012)
Country, Wave dummy	Yes	Yes	Yes	Yes	Yes	Yes
Cognitive, Diseases	No	No	No	No	Yes	Yes
N. Observations	18536	18536	18281	17854	17851	17851
N. Countries	5	5	5	5	5	5

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods) and CatIRR1 (IRR=2 and IRR=3), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.3: *Probit Model (SHARE): Risk Aversion, Marginal Effects. Controls for Linguistic Families and Separated Regression for Linguistic Sub-families.*

Risk Aversion (d)	RA1	RA2	RA3	RA4	RA5
IRR	0.019**		0.033***	0.113***	0.022***
	(0.008)		(0.003)	(0.002)	(0.003)
Cat=1		0.047***			
		(0.015)			
Cat=2		0.091***			
		(0.007)			
Age	0.003***	0.003***	0.001***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Female	0.092***	0.092***	0.066***	0.131***	0.066***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Low Edu.	0.062***	0.062***	0.051***	0.071***	0.064***
	(0.008)	(0.008)	(0.011)	(0.010)	(0.013)
High Edu.	-0.080***	-0.079***	-0.067***	-0.084***	-0.088***
	(0.008)	(0.008)	(0.008)	(0.013)	(0.013)
Income	-0.012***	-0.012***	-0.011***	-0.015***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Owner	-0.033***	-0.033***	-0.023**	-0.044***	-0.024
	(0.008)	(0.008)	(0.010)	(0.013)	(0.016)
Married	0.015**	0.015**	0.004	0.024**	0.008
	(0.007)	(0.007)	(0.005)	(0.012)	(0.014)
Num. Children	0.002	0.002	0.002	0.007**	-0.006
	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)
Trust People	-0.012***	-0.012***	-0.008***	-0.014***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Unemployed	0.035***	0.035***	0.022	0.050***	0.035**
	(0.008)	(0.008)	(0.014)	(0.013)	(0.014)
Disabled	0.050***	0.050***	0.053***	0.044**	0.054***
	(0.009)	(0.009)	(0.012)	(0.020)	(0.007)
Homemaker	0.027***	0.026***	0.028***	0.024*	0.022
	(0.008)	(0.008)	(0.008)	(0.013)	(0.017)
Country, Wave dummy	Yes	Yes	Yes	Yes	Yes
Cognitive, Diseases	Yes	Yes	Yes	Yes	Yes
Linguistic family	Yes	Yes	Only	Only	Only
and sub-family			Romance	Germanic	Other
N. Observations	74426	74426	22907	29527	21992
N. Countries	19	19	7	8	7

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.4: *Fixed effects: Conditional Logit model (SHARE). Odds-Ratios. All Countries.*

Risk Aversion (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	1.170***	1.156***	1.142**	1.166**	1.157**
	0.057	0.057	0.068	0.079	0.083
Retired		1.263***	1.247***	1.173***	1.146***
		0.051	0.052	0.049	0.048
Unemployed		1.383***	1.351***	1.358***	1.340***
		0.066	0.068	0.115	0.113
Disabled		1.645***	1.572***	1.490***	1.327***
		0.108	0.105	0.112	0.101
Homemaker		1.219***	1.216***	1.150*	1.129
		0.071	0.074	0.092	0.090
Owner		0.818***	0.834***	0.850**	0.858**
		0.041	0.042	0.054	0.054
Trust People			0.920***	0.918***	0.921***
			0.005	0.006	0.006
Cognitive, Diseases	No	No	No	No	Yes
Fixed Effects:					
Sex x Age	Yes	Yes	Yes	Yes	Yes
Country x Wave	No	Yes	Yes	Yes	Yes
Income x Education	No	Yes	Yes	Yes	Yes
MarStatus x Num.Child	No	No	No	Yes	Yes
N. Observations	57162	56105	55656	30045	30038
N. Countries	19	19	19	19	19

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.5: *Fixed effects: Conditional Logit model (SHARE). Odds-Ratios. Linguistically Heterogeneous Countries.*

Risk Aversion (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	1.174***	1.156***	1.145**	1.169**	1.160*
	0.062	0.064	0.075	0.084	0.089
Retired		1.372***	1.346***	1.217*	1.190*
		0.088	0.086	0.129	0.125
Unemployed		1.386**	1.371*	1.530*	1.517*
		0.223	0.223	0.354	0.366
Disabled		1.597***	1.514***	1.589***	1.410***
		0.154	0.160	0.187	0.143
Homemaker		1.106	1.104	1.027	1.011
		0.102	0.102	0.170	0.166
Owner		0.681***	0.692***	0.668***	0.681***
		0.050	0.058	0.081	0.090
Trust People			0.930***	0.937***	0.940***
			0.009	0.012	0.012
Cognitive, Diseases	No	No	No	No	Yes
Fixed Effects:					
Sex x Age	Yes	Yes	Yes	Yes	Yes
Country x Wave	No	Yes	Yes	Yes	Yes
Income x Education	No	Yes	Yes	Yes	Yes
MarStatus x Num.Child	No	No	No	Yes	Yes
N. Observations	12812	12528	12404	5819	5819
N. Countries	5	5	5	5	5

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.6: *Probit Model (WVS): Risk Aversion, Marginal Effects*

Risk Aversion (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	0.024***	0.022***	0.023***	0.022***	0.022***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Age	0.004***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female (d)	0.080***	0.077***	0.073***	0.074***	0.073***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Low Education (d)		0.033***	0.018**	0.016**	0.015**
		(0.008)	(0.008)	(0.008)	(0.008)
High Education (d)		-0.023***	-0.009	-0.007	-0.005
		(0.005)	(0.004)	(0.004)	(0.004)
Income			-0.010***	-0.010***	-0.010***
			(0.001)	(0.001)	(0.001)
Married (d)				0.024***	0.022***
				(0.004)	(0.004)
Num. Children				0.003**	0.003***
				(0.001)	(0.001)
Retired (d)					0.034***
					(0.010)
Unemployed (d)					0.010
					(0.011)
Homemaker (d)					0.008
					(0.005)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes
Control: Cognitive	NA	NA	NA	NA	NA
Control: Health	No	No	Yes	Yes	Yes
N. Observations	64783	64301	59133	59025	58118
N. Countries	47	47	47	47	46

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit. Robust standard errors in parentheses. Reference categories variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or self-employed.*
***** $p < 0.1$, ****** $p < 0.05$, ******* $p < 0.01$.

Table 3.E.7: *IV Stocks (SHARE): First Stage Estimation and Test Statistics, Linguistically Heterogeneous Countries*

Risk Aversion	FS1_ML	FS2_ML	FS3_ML	FS4_ML	FS5_ML	FS6_ML
IRR	0.030***	0.028***			0.038***	0.036***
	(0.005)	(0.005)			(0.006)	(0.006)
Cat=2			0.057***	0.053***		
			(0.010)	(0.010)		
FTR					-0.028**	-0.027**
					(0.012)	(0.012)
Age	0.002***	0.001***	0.002***	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.108***	0.109***	0.109***	0.110***	0.108***	0.109***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Low Education	0.073***	0.064***	0.073***	0.063***	0.073***	0.064***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
High Education	-0.115***	-0.107***	-0.111***	-0.103***	-0.113***	-0.105***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Owner	-0.072***	-0.071***	-0.075***	-0.074***	-0.073***	-0.072***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Income	-0.011***	-0.010***	-0.012***	-0.011***	-0.011***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married	0.012	0.013	0.012	0.014	0.011	0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Num. Children	-0.007***	-0.008***	-0.007***	-0.007***	-0.007***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Retired	0.038***	0.038***	0.039***	0.039***	0.037***	0.037***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Unemployed	0.025	0.026	0.026	0.026	0.025	0.026
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Disabled	0.064***	0.054***	0.065***	0.054***	0.064***	0.054***
	(0.017)	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)
Homemaker	0.004	0.001	0.000	-0.003	0.001	-0.002
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Trust People	-0.007***	-0.007***	-0.008***	-0.007***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.739***	0.826***	0.808***	0.893***	0.739***	0.825***
	(0.038)	(0.042)	(0.033)	(0.037)	(0.038)	(0.042)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	No	Yes	No	Yes	No	Yes
N. Observations	12022	12020	12022	12020	12022	12020
N. Countries	5	5	5	5	5	5
Strong Instrument	**36.07**	**31.07**	**31.21**	**27.22**	**40.83**	**35.92**

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is ivreg2 (only the first stage estimates reported). Robust standard errors in parentheses. Reference categories for dichotomous variables: CatIRR0 (no IRR Moods) and CatIRR1 (IRR=2 and IRR=3), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed, Weak FTR (FTR = 0).*

Table 3.E.8: *IV Stocks (SHARE): Bivariate Probit, Marginal effects; Linguistically Heterogeneous Countries*

Pr(Risky Assets)	**MEff_IRR** 2nd Stage RA	**MEff_FTR** Direct MEff	**MEff_FTR** 2nd Stage RA
Risk Aversion (d)	-0.100***		-0.070***
	(0.027)		(0.025)
Strong FTR (d)		-0.052***	-0.037***
		(0.003)	(0.005)
Owner (d)	0.021***	0.060***	0.019***
	(0.006)	(0.003)	(0.005)
Income	0.006***	0.012***	0.005***
	(0.001)	(0.002)	(0.001)
Low Education (d)	-0.009	-0.038***	-0.010*
	(0.006)	(0.004)	(0.005)
High Education (d)	0.005	0.055***	0.011**
	(0.006)	(0.006)	(0.006)
Married (d)	0.010*	0.007	0.007
	(0.005)	(0.011)	(0.005)
Num. Children	-0.009***	-0.010***	-0.009***
	(0.001)	(0.003)	(0.001)
Age	0.001***	0.001*	0.001***
	(0.000)	(0.000)	(0.000)
Female (d)	0.006	-0.040***	0.005
	(0.005)	(0.004)	(0.004)
Retired (d)	0.012**	0.008	0.011*
	(0.006)	(0.014)	(0.006)
Unemployed (d)	-0.004	-0.013	-0.003
	(0.012)	(0.017)	(0.011)
Disabled (d)	-0.025**	-0.055***	-0.016
	(0.010)	(0.012)	(0.011)
Homemaker (d)	0.019**	0.023***	0.016*
	(0.009)	(0.006)	(0.009)
Trust People	0.001	0.004**	0.000
	(0.001)	(0.002)	(0.001)
Control: Country/Wave	Yes	Yes	Yes
Control: Cognitive, Health	Yes	Yes	Yes
N. Observations	12022	12022	12020
N. Countries	5	5	5

Notes: *The dependent variable is "Has Stocks (d)". The method of estimation is Recursive Bivariate Probit (only second stage reported). Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed, Weak FTR (FTR = 0).*

Bibliography

Abbink, K. and Hennig-Schmidt, H. (2006). Neutral versus loaded instructions in a bribery experiment. *Experimental Economics*, 9(2):103–121.

Allingham, M. G. and Sandmo, A. (1972). Income Tax Evasion: A Theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alm, J., Bloomquist, K. M., and McKee, M. (2010). On the external validity of tax compliance experiments. *... Meeting of the National Tax ...*.

Alm, J., Bloomquist, K. M., and Mckee, M. (2015). On the external validity of laboratory tax compliance experiments. *Economic Inquiry*, 53(2):1170–1186.

Alm, J., Deskins, J., and McKee, M. (2008). Do Individuals Comply on Income Not Reported by Their Employer? *Public Finance Review*, 37:120–141.

Alm, J., Jackson, B., and McKee, M. (1992a). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 45:107–114.

Alm, J., McClelland, G. H., and Schulze, W. D. (1992b). Why do people pay taxes? *Journal of Public Economics*, 48:21–38.

Alm, J. and McKee, M. (2006). Audit certainty, audit productivity, and taxpayer compliance. *National Tax Journal*, (March).

Alm, J., Sanchez, I., and de Juan, A. (1995). Economic and Noneconomic Factors in Tax Compliance. *Kyklos*, 48:3–18.

Alm, J. and Torgler, B. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology*, 27(2):224–246.

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

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2):818–860.

Andreoni, J. and Sprenger, C. (2012a). Estimating Time Prefernces from Convex Budgets. *American Economic Review*, 102(7):3333–3356.

Andreoni, J. and Sprenger, C. (2012b). Risk Preferences Are Not Time Preferences. *American Economic Review*, 102(7):3357–3376.

Andreoni, J. and Sprenger, C. (2015). Risk Preferences Are Not Time Preferences: Reply. *American Economic Review*, 105(7):2287–2293.

Arrow, K. (1965). *Aspects of the theory of risk bearing*. Yrjo Jahnsson Saatio, Helsinki.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 9(3):396–406.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2):537–579.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans. *Neuron*, 58(4):639–650.

Bazart, C. and Bonein, A. (2014). Reciprocal relationships in tax compliance decisions. *Journal of Economic Psychology*, 40:83–102.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2):169–217.

Bellante, D. and Green, C. A. (2004). Relative risk aversion among the elderly. *Review of Financial Economics*, 13(3):269–281.

Bernasconi, M. (1992). Different frames for the independence axiom: An experimental investigation in individual decision making under risk. *Journal of Risk and Uncertainty*, 5(2):159–174.

Bernasconi, M. (1998). Tax evasion and orders of risk aversion. *Journal of Public Economics*, 67(1):123–134.

Bernasconi, M., Corazzini, L., and Seri, R. (2014). Reference dependent preferences, hedonic adaptation and tax evasion: Does the tax burden matter? *Journal of Economic Psychology*, 40(0):103–118.

Bernasconi, M. and Loomes, G. (1992). Failures of the reduction principle in an Ellsberg-type problem. *Theory and Decision*, 32(1):77–100.

Bernhofer, J. (2016). Experimental Economies and Tax Evasion: The Order Beyond the Market. In Cecconi, F., editor, *New Frontiers in the Study of Social Phenomena - Cognition, Complexity, Adaptation*, chapter 5, pages 81–99. Springer International Publishing, 1 edition.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India.

Bloomquist, K. M. (2004). Modeling taxpayers' response to compliance improvement alternatives. In *Annual conference of the North American association for computational social and organizational sciences, Pittsburgh, PA.*

Bloomquist, K. M. (2006). A Comparison of Agent-Based Models of Income Tax Evasion. *Social Science Computer Review*, 24:411–425.

Bloomquist, K. M. (2011). Tax Compliance as an Evolutionary Coordination Game: An Agent-Based Approach. *Public Finance Review*, 39(1):25–49.

Bonsang, E. and Dohmen, T. (2015). Risk attitude and cognitive aging. *Journal of Economic Behavior & Organization*, 112:112–126.

Breu, W. (2010). Mood in Albanian. *Mood in the Languages of Europe*, pages 447–472.

Buchanan, J. M. (1987). The Constitution of Economic policy. *The American Economic Review*, 77(3):243–250.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Number v. 3 in Meaning and Necessity: A Study in Semantics and Modal Logic. University of Chicago Press.

Chen, M. K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. *American Economic Review*, 103(2):690–731.

Choo, C. Y. L., Fonseca, M. A., and Myles, G. D. (2014). Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment. *Journal of Economic Behavior and Organization*, 124:102–114.

Choo, L., Fonseca, M. A., and Myles, G. (2013). Lab Experiment to Investigate Tax Compliance : Audit Strategies and Messaging. Technical Report 308, University of Exeter.

Clotfelter, C. T. (1983). Tax evasion and tax rates: An analysis of individual returns. *The review of economics and statistics*, 65(3):363–373.

Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, 34(2):669–700.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). Dynamic choice and the common ratio effect: An experimental investigation. *The Economic Journal*, 108:1362–1380.

Dahl, Ö. (1985). Tense and aspect systems.

Dahl, Ö. (2000). The grammar of future time reference in European languages. In Dahl, Ö., editor, *Tense and Aspect in the Languages of Europe*, pages 309–328. Walter de Gruyter, Berlin.

Danthine, J.-P. and Donaldson, J. B. (2005). *Intermediate Financial Theory*. Academic Press.

Davies, I. R. L. and Corbett, G. G. (1997). A cross-cultural study of colour grouping: evidence for weak linguistic relativity. *British journal of psychology*, 88:493–517.

Davis, J. S., Hecht, G., and Perkins, J. D. (2003). Social Behaviors, Enforcement, and Compliance Dynamics. *The Accounting Review*, 78(1):39–69.

Dhami, S. and Al-Nowaihi, A. (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior & Organization*, 64(1):171–192.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2005). Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. *IZA Discussion Paper*, (1730):1–56.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Duffy, J. (2006). Agent-based models and human subject experiments. In *Handbook of computational economics*, volume 2, chapter 19.

Elffers, H., Weigel, R. H., and Hessing, D. J. (1987). The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology*, 8(3):311–337.

Elster, J. (1989). Social Norms and Economic Theory. *The Journal of Economic Perspectives*, 3(4):99–117.

Engel, C. and Moffat, P. G. (2012). Estimation of the House Money Effect Using Hurdle Models.

Engel, C. and Moffat, P. G. (2014). dhreg, xtdhreg, and bootdhreg: Commands to implement double-hurdle regression. *The Stata Journal*, 14(4):778–797.

Epstein, L. G. and Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica*, 57(4):937–969.

Feinstein, J. S. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. *The RAND Journal of Economics*, 22(1):14–35.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Frey, B. S. and Feld, L. P. (2002). Deterrence and morale in taxation: An empirical analysis.

Frey, B. S. and Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5):589–611.

Friedland, N. (1982). A Note on Tax Evasion as a Function of the Quality of Information about the Magnitude and Credibility of Threatened Fines: Some Preliminary Research. *Journal of Applied Social Psychology*, 12(1):54–59.

Geeraerts, D. and Cuyckens, H. (2010). *The Oxford handbook of cognitive linguistics*. Oxford University Press.

Gentner, D. and Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought*. Bradford Books. MIT Press.

Giorgi, A. (2009). Toward a syntax of the subjunctive mood. *Lingua*, 119(12):1837–1858.

Graetz, M. J. and Wilde, L. L. (1985). The Economics of Tax Compliance: Facts and Fantasy. *National Tax Journal*, 38:355–363.

Guala, F. and Mittone, L. (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology*, 12(March):495–515.

Guiso, L. and Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic Association*, 6(6):1109–1150.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Trusting the Stock Market. *The Journal of Finance*, 63(6):v.

Halek, M. and Eisenhauer, J. G. (2001). Demography of Risk Aversion. *The Journal of Risk and Insurance*, 68(1):1–24.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2015). Reduction of compound lotteries with objective probabilities: Theory and evidence. *Journal of Economic Behavior and Organization*, 119:32–55.

Harrison, G. W. and Swarthout, J. T. (2014). Experimental payment protocols and the Bipolar Behaviorist. *Theory and Decision*, 77(3):423–438.

Hartog, J., Ferrer-i Carbonell, A., and Jonker, N. (2002). Linking Measured Risk Aversion to Individual Characteristics. *Kyklos*, 55(1):3–26.

Hersch, J. (1996). Smoking, seat belts, and other risky consumer decisions: differences by gender and race. *Managerial and Decision Economics*, 17(5):471–481.

Hill, J. H. and Mannheim, B. (1992). Language and World View. *Annual Review of Anthropology*, 21(1):381–406.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203:88–96.

Hockett, C. F. and Altmann, S. A. (1968). A note on design features. *Animal communication: Techniques of study and results of research. Indiana University Press.*, 203:61–72.

Holt, C. A. (1986). Preference Reversals and the Independence Axiom. *The American Economic Review*, 76(3):508–151.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2):231–270.

Jianakoplos, N. and Bernasek, A. (1998). Are Women More Risk Averse. *Economic Inquiry*, 10(4):307–324.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):pp. 263–292.

Kahneman, D. and Tversky, A. (2000). *Choices, values, and frames*. Cambridge University Press.

Kapteyn, A. and Teppa, F. (2011). Subjective measures of risk aversion, fixed costs, and portfolio choice. *Journal of Economic Psychology*, 32(4):564–580.

Kastlunger, B., Dressler, S. G., Kirchler, E., Mittone, L., and Voracek, M. (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D:4D). *Journal of Economic Psychology*, 31(4):542–552.

Kastlunger, B., Kirchler, E., Mittone, L., and Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology*, 30(3):405–418.

Keller, L. R. (1985). Testing of the 'reduction of compound alternatives' principle. *Omega*, 13(4):349–358.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. S., and Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692.

Kovacic, M., Bernhofer, J., and Costantini, F. (2016). Risk Attitudes, Investment Behavior and Linguistic Variation.

Laca, B. (2010). Mood in Spanish. In Thieroff, R. and Rothstein, B., editors, *Mood in the languages of Europe*, Studies in Language Companion Series, pages 198–220. John Benjamins.

Laury, S. K. and Taylor, L. O. (2008). Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? *Journal of Economic Behavior and Organization*, 65(1):9–29.

Lee, B., Rosenthal, L., Veld, C., and Veld-Merkoulova, Y. (2015). Stock market expectations and risk aversion of individual investors. *International Review of Financial Analysis*, 40:122–131.

Lin, F.-T. (2009). Does the risk aversion vary with different background risk of households? *International Research Jounal of Finance and Economics*, 34(34):69 – 82.

Loomes, G. and Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92(368):805.

Machina, M. J. (1989). Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty. *Journal of Economic Literature*, 27(4):1622–1668.

Maciejovsky, B., Kirchler, E., and Schwarzenberger, H. (2007). Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, 28(6):678–691.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3):108–114.

Malmendier, U. and Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *The Quarterly Journal of Economics*, 126(1):373–416.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics*, 35(5):813–835.

Mittone, L. and Patelli, P. (2000). Imitative Behaviour in Tax Evasion. In Luna, F. and Stefansson, B., editors, *Economic Simulations in Swarm: Agent-Based Modelling and Object Oriented Programming*, volume 14 of *Advances in Computational Economics*, pages 133–158. Springer US.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillan.

Myles, G. and Naylor, R. a. (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12:49–66.

Nebout, A. and Dubois, D. (2014). When Allais meets Ulysses: Dynamic axioms and the common ratio effect. *Journal of Risk and Uncertainty*, pages 1–31.

Pellizzari, P. and Rizzi, D. (2014). Citizenship and power in an agent-based model of tax compliance with public expenditure. *Journal of Economic Psychology*, 40:35–48.

Pinker, S. (1994). The Language Instinct: The New Science of Language and Mind. *Penguin Books*, page 483.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32(1):122–136.

Quiggin, J. (1982). A theory of anticipated utility.

Reuben, E. and Suetens, S. (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics*, 15(1):24–43.

Reyna, V. F. and Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18(1):89–107.

Riley Jr., W. B. and Chow, K. V. (1992). Asset Allocation and Individual Risk Aversion. *Financial Analysts Journal*, 48(6):32–37.

Roberson, D., Davidoff, J., and Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, 71(1):1–42.

Roth, A. E. (1987). *Laboratory experimentation in economics.* Cambridge University Press.

Rothstein, B. and Thieroff, R. (2010). *Mood in the Languages of Europe.* Studies in language companion series. John Benjamins.

Rubinstein, A. (2006a). A sceptic's comment on the study of economics. *Economic Journal*, 116(510):1–9.

Rubinstein, A. (2006b). Dilemmas of an economic theorist. *Econometrica*, 74(4):865–883.

Samuelson, P. A. (1952). Probability, utility, and the independence axiom. *Econometrica: Journal of the Econometric Society*, 20(4):670–678.

Santacreu-Vasut, E., Shoham, A., and Gay, V. (2013). Do female/male distinctions in language matter? Evidence from gender political quotas. *Applied Economics Letters*, (September 2012):37–41.

Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36):15268–15273.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech.* Harvest books. Harcourt, Brace.

Segal, U. (1987). The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach. *International Economic Review*, 28(1):175–202.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica: Journal of the Econometric Society*, 58(2):349–377.

Segal, U. (1992). The independence axiom versus the reduction axiom: Must we have both? In *Utility Theories: Measurements and Applications*, pages 165–183. Springer.

Selten, R. (2001). What Is Bounded Rationality ? *Bounded rationality The adaptive toolbox*, pages 1–25.

Selten, R. and Stoecker, R. (1986). End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach. *Journal of Economic Behavior and Organization*, 7(1):47–70.

Siegel, F. and Hoban, J. (1991). Measuring risk aversion: Allocation, leverage, and accumulation. *Journal of Financial Research*, 14(1):27–35.

Simon, H. (1972). Theories of Bounded Rationality. In *Decision and Organization*, chapter 8, pages 161–176.

Simon, H. A. (1982). *Models of bounded rationality*, volume 2.

Slemrod, J., Blumenthal, M., and Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3):455–483.

Slemrod, J. and Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In *Handbook of Public Economics*, volume 3, pages 1423–1470.

Snow, A. and Warren, R. S. (2005). Tax evasion under random audits with uncertain detection. *Economics Letters*, 88(1):97–100.

Starmer, C. and Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review*, 81(4):971–978.

Sundén, A. E. and Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *American Economic Review*, 88(2):207–211.

Swan, O. E. (2002). *A Grammar of Contemporary Polish*. Slavica.

Thieroff, R. (2000). On the Areal Distribution of Tense-Aspect Categories in Europe. In Dahl, Ö., editor, *Tense and Aspect in the Languages of Europe*, pages 265–308. Walter de Gruyter, Berlin.

Torgler, B. (2002). Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. *Journal of Economic Surveys*, 16(5):657–683.

Torgler, B. (2003). Tax morale and institutions.

Torgler, B. and Schneider, F. (2007). Shadow economy, tax morale, governance and institutional quality: a panel analysis.

Torgler, B. and Valev, N. T. (2010). Gender and Public Attitudes Toward Corruption and Tax Evasion. *Contemporary Economic Policy*, 28(4):554–568.

Tourangeau, R. and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5):859–883.

Tversky, A. and Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5:297–323.

Vogel, J. (1974). Taxation and public opinion in Sweden: An interpretation of recent survey data. *National Tax Journal*, 27(4):499–514.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*, volume 60. Princeton university press Princeton.

Voors, M., Turley, T., Kontoleon, A., Bulte, E., and List, J. A. (2012). Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone. *Economics Letters*, 114(3):308–311.

Whorf, B. L. and Carroll, J. B. (1964). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT language antropology. M.I.T. Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):7780–7785.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach*. South-Western CENGAGE Learning, 4th editio edition.

Xie, D. (2000). Power risk aversion utility functions. *Annals of Economics and Finance*, 1:265–282.

Yitzhaki, S. (1974). A note on Income Tax Evasion: A theoretical analysis. *Journal of Public Economics*, 3:201–202.

Estratto per riassunto della tesi di dottorato

Dottorando: JULIANA BERNHOFER
Matricola: 804668
Dottorato: Economia
Ciclo: 27

Titolo della tesi: Essays on tax compliance, economic behavior and linguistic relativity

Abstract: This dissertation presents three essays on tax compliance, economic behavior and linguistic relativity. In Chapter 1, I explore the existing literature of behavioral and experimental economics, as well as multi-agent simulation models applied to tax compliance. The analysis begins with the presentation of the classical models of expected utility theory and the discussion of the limitations arising with the assumption of full rationality. During the last decades, the inclusion of institutions, social norms and psychological factors in laboratory experiments have revealed numerous interesting insights to policy-makers. However, given the complexity and the size of the research, I propose the use of agent-based simulation models as flexible tools when accounting for existing findings and to explore new research horizons. Some examples of computer-simulated agent-based models applied to the tax evasion problem are presented.

Chapter 2 tests the *Reduction of Compound Lotteries Axiom* in a tax evasion experiment. When the probability of audit and the probability of detection are considered separately, we find violations of the reduction axiom and thus of the assumption that tax payers are expected utility maximizers. Our results call for a reconsideration of existing theoretical models on individual tax compliance.

Finally, Chapter 3 is an interdisciplinary study of the role of language to economic behavior. In particular, we test if a more intense use of the *Irrealis* mood, i.e., the verb form that expresses potential situations, in an individual's spoken language impacts on their attitude towards risk by giving linguistic salience to the concept of uncertainty. We find that the use of *Irrealis* mood is positively and significantly related to the level of risk aversion. We use our findings to assess the role of (instrumented) risk aversion in investment decisions.

Estratto: Questa tesi presenta tre saggi su adempimenti fiscali, comportamento economico e relatività linguistica. Nel primo capitolo, esploro la letteratura esistente di economia comportamentale e sperimentale e di modelli di simulazione ad agenti multipli applicata all'evasione fiscale. L'analisi inizia con la presentazione dei modelli classici della teoria dell'utilità attesa e la discussione delle limitazioni derivanti dall'assunzione razionalità economica. Nel corso degli ultimi decenni, l'inclusione del ruolo di istituzioni, norme sociali e fattori psicologici in esperimenti di laboratorio hanno rivelato numerosi spunti interessanti per i *policy maker*. Tuttavia, data la complessità e la dimensione della ricerca, propongo l'utilizzo di modelli di simulazione basati su agenti come strumenti flessibili per

la considerazione delle scoperte esistenti e per esplorare nuovi orizzonti di ricerca. Alcuni esempi di modelli ad agenti applicati al problema dell'evasione fiscale sono presentati.

Il secondo capitolo testa l'*Assioma della riduzione delle lotterie composte* in un esperimento di evasione fiscale. Quando la probabilità di audit e la probabilità di essere scoperti sono considerate separatamente, osserviamo una violazione dell'assioma di riduzione e quindi dell'ipotesi che i contribuenti sono massimizza tori di utilità attesa. I nostri risultati richiedono una riconsiderazione dei modelli teorici esistenti sull'adempimento fiscale individuale.

Infine, il terzo capitolo è uno studio interdisciplinare sul ruolo del linguaggio nel comportamento economico. In particolare, si analizza se un uso più intenso della forma *Irrealis*, ossia la forma verbale che esprime situazioni possibili, nella lingua parlata abbia un impatto sull'atteggiamento individuale nei confronti del rischio, in quanto rende il concetto di incertezza linguisticamente evidente. Troviamo che l'uso della forma *Irrealis* è positivamente e significativamente correlata al livello di avversione al rischio. Usiamo i nostri risultati per valutare il ruolo dell'avversione al rischio (strumentata) nelle decisioni di investimento.